

Exploring the Dynamism between Propositional Complexity and Error Rate: A Case Study

Jordan Van Horn¹

Teachers College, Columbia University

ABSTRACT

This case study analyzed data from an asynchronous written mediated exchange between a native speaker and nonnative speaker of English. Three methods of analysis- error analysis, propositional complexity (idea units), and qualitative analysis were used to analyze learner performance over time. The results from these methods were visualized and interpreted according from Complex Dynamic Systems Theory (CDST) perspective, which allows longitudinal language production data to be interpreted from multiple perspectives so that change can be understood at both the micro and macro levels. Results from the analysis revealed patterns of linguistic change both within and between participants as they interacted over time. Several instances of continuity in language production were observed between the participants, including a drop in error rate that coincided with an increase in minor idea units during one turn of the exchange. Results from a qualitative analysis offer several explanations for this finding and are related to the implications this holds for CDST.

INTRODUCTION

As the field of second language acquisition continues to grow and develop, new methods for analyzing learner language have emerged, each building upon the limitations of those that were previously developed. Using these new methods to analyze learner data, may provide a researcher with a richer and more diverse view of a learner's interlanguage that can offer a better solution to some of the limitations that arise as a result of employing one method to analyze performance data. In this vein perhaps, Complex Dynamic Systems Theory (CDST) might offer an approach that can unify traditionally distinct methods of analysis to provide a means of analyzing learner data. CDST can be applied in conjunction with traditional methods of analyzing learner data, provided that the sample of learner data is longitudinal in nature. Rather than treating language development as a linear process, CDST posits that learners follow a complex and variable path as they acquire an L2. Although in its infancy, CDST affords SLA theorists an opportunity to look at learner data from multiple perspectives over time to study linguistic change from both a micro and macro perspective.

¹ Jordan Van Horn is a graduate of the M.A. Applied Linguistics program at Teachers College, Columbia University. She is the USCIS Program Coordinator at the Center for Immigrant Education and Training and works as an adjunct instructor at both The English Language Center (TELC) and the Center for Immigrant Education and Training (CIET) at LaGuardia Community College, City University of New York. Correspondence should be sent to: jav2164@tc.columbia.edu.

This paper will use a CDST perspective to examine a dataset using three methods of data analysis. It will offer a description, analysis, and interpretation of how these methods can inform what we know about learner interlanguage and about language development. In this case study, the nature of the kinds of errors produced by the learners in the dataset, the relationship between major and minor idea units and error rate, and evidence of patterns in the kinds of errors that arise in major and minor ideas will be explored in depth.

LITERATURE REVIEW

Although CDST has its origins in the field of mathematics, its application has been extended beyond this field to other disciplines, including Second Language Acquisition (SLA). CDST operates on the principles of complex systems, which posit that these systems exhibit several core characteristics, including, sensitive dependence on initial systems, complete interconnectedness, the emergence and development of attractor states, and variation which can be seen both between and within individuals over time (De Bot, Lowie, Verspoor 2007). CDST approaches in SLA view language as exhibiting the characteristics of complex systems and language development as one that involves natural flux and variation. This variation was exemplified in a study by Verspoor, Lowie, and van Dijk (2008), who examined patterns of intra-individual variability in a case study of an advanced learner of English. In this study, the researchers examined 18 academic writing samples written over the course of three years and analyzed errors in the development of vocabulary use and sentence complexity. The goal of the study was to observe the development of an academic style of writing over time. An analysis of the data found nonlinear patterns of change in several domains of vocabulary use and sentence complexity. Over the course of three years, the researchers found that these changes oscillated between both progress and regress. Within this environment of change, Verspoor et. al were able to show evidence of interactions between different subsystems of language, including interference as well as systems which exhibited complementary growth.

Linguistic accuracy and complexity have each been studied in a number of different ways in CDST studies. Defining and measuring complexity is a challenging endeavor because the construct consists of multiple components and dimensions (Norris & Ortega, 2009). Larsen-Freeman (2006), for example, used idea units to measure complexity in her study which examined the complexity, accuracy, and fluency of oral and written samples of five Chinese learners of English. This was an exploratory study that was conducted using a CDST framework and featured a time-series design which collected oral and written data over the course of six months. The task consisted of a prompt, which asked the learners to describe an important event in their lives, and three days after completing a written version of this narrative, participants were instructed to share an oral version of what they wrote. In accordance with CDST principles, Larsen-Freeman analyzed the data from a micro and macro perspective. At the macro-level, quantitative measures were used to observe the arrangement and re-organization of the system over time. At the micro-level, qualitative measures were employed to better understand the changes in language usage over time. Intra-individual variability was examined in order to identify attractor states within individual performance. The quantitative analyses revealed that as a whole, all five learners made progress in each of the measured variables. However when examined more closely these learners appeared to be following differential paths of development. Larsen-Freeman interpreted these differential paths of use as potential precursors

to a re-organization in a learner's linguistic system. Additionally, the researcher called for more micro-developmental studies to better examine how learners abandon old forms in favor of new varieties of a similar form.

Spoelman and Verspoor (2010) also examined complexity and accuracy. They conducted a longitudinal case study to examine changes in accuracy and complexity in a Dutch L1 speaker acquiring Finnish as a beginner. The study focused on the Finnish L2 learner's intra-individual variability in accuracy rates and measures of complexity from a dataset of written texts. The purpose of this study was to examine the role of intra-individual variability in the second language development of a beginner, by analyzing learner data from a language other than English. The case study involved the collection and analysis of 54 writing samples that were analyzed according to several measures of accuracy and complexity. After collecting and analyzing the data, the authors employed several CDST measures in order to study the development of the learner system. These measures included min-max graphs, regression analyses, and Monte Carlo analyses. This study revealed several important findings, namely that variability in the learner's language could provide insight into a second language learner's development. Additionally, the data revealed that interaction among several measures of complexity indicate there can be differential patterns of performance, and that a competition of resources can occur within several subsystems of language. Furthermore, an interpretation of this dataset reveals that patterns of interaction (such as those between accuracy and complexity) can only be observed from a macro developmental perspective.

Finally, Polat and Kim (2013) used a CDST approach in a longitudinal case study that examined the development of complexity and accuracy for an untutored learner of English. This study posed the research question, "How does complexity and accuracy develop in an untutored learner of English?" and "How can this data be interpreted according to a CDST perspective?". This study was unique in that it was the first to use CDST to examine the data of an untutored student and its implications in the field of SLA. In order to answer the research questions, the researchers interviewed an advanced (L1 Turkish) learner of English once every two weeks for thirty minutes over a one year period. This research design enabled the researchers to obtain the microdevelopmental data necessary in the investigation of their research questions. In addition to learner data, three interviews with three different native speakers of English were conducted in order to obtain a comparison on which to base the development of measures of complexity and accuracy. Data from these English native speaker interviews were then transcribed and analyzed according to a number of measures for syntactic complexity, lexical diversity, and accuracy. When compared to native speaker performance, the untutored learner of English performed very differently according to the three measures of accuracy and complexity. Additionally, the researchers found that within a specific measure (syntactic complexity), the learner's performance was very different based on the three sub-measures that were theorized to comprise this measure. The authors interpreted these findings as evidence for the view that each sub-component of a larger measure (like complexity or accuracy) may be indicative of differential patterns of development within and between learners, and that each subcomponent, in turn, should be measured in context to better understand language development overall. Thus, it is clear from previous research in the field that linguistic complexity and accuracy can be examined from a CDST perspective and that using this perspective in SLA research has the potential to yield important insights about learner interlanguage.

METHODOLOGY

Data and Participants

This case study was based on a dataset of email exchanges between two college-aged females, who were fulfilling the requirements for an email exchange project for their Intercultural Communication class in the spring of 2011. One participant (Michelle) is a 22 year-old bilingual (Spanish-English) student living in South Texas. The second participant, Belinda, is a 21-year-old second language learner of English from the Qingdao, Shandong Province, in China. The participants were paired up and instructed to exchange emails for a period of 5-7 weeks. The goal of the project was to meet and interact with an interlocutor from a different culture. There were no explicit directions or restrictions on the topics that they could discuss.

Data were examined using error analysis, complexity measures, and qualitative methods. A description of each method of analysis follows.

Methods of Analysis

Error Analysis

An error analysis was conducted by the researcher after a careful review and coding of errors found in the email exchange. The aim of this error analysis was to examine the types of errors and whether they showed variability between and within participants over time. Following Corder's (1974) steps in conducting an error analysis, error types were identified and coded into categories that described the nature of the error that was made. Because error analysis relies on the judgment of the researcher in making the coding decisions, it was necessary to employ the assistance of a second rater in evaluating these errors. The second rater was a native English speaker and graduate student in Applied Linguistics at Teachers College, Columbia University, who had worked as an ESL teacher for six years. The raters both examined Time 1, 2, 3 (determined by each partner's turn in the exchange) in the dataset and found discrepancies on only five errors. Because of their close congruency in identifying and coding errors in the dataset, the method of identifying and tagging errors was deemed a reliable system of analysis. The linguistic categories that accounted for the errors in this dataset were generated using an iterative bottom-up approach, whereby an error was given a descriptive tag when it was encountered in the data set. For this study, the linguistic categories of errors included: punctuation, adverb, prepositions, possession, vocabulary, plurality, run on sentences, tense, pronoun, article, and word order.

This method of tallying was chosen in order to see the kinds of errors made by each participant and to determine if there were any patterns in the categories of these errors over time. After tallying errors within each participant, an analysis was conducted to compare error types between participants. An error frequency graph was generated to visualize the dynamic patterns of errors made by the participants over time.

Propositional Complexity Measure

A second measure was performed to assess the propositional completeness, which Ellis and Barkhuizen (2005) define as a measure of complexity as well as taken as a measure which can provide evidence of a learner who has engaged in “conceptualization” (Levitt, 1989). The aim of this measure was to explore the nature of propositional complexity in each of the participants over time. In this measure, propositional complexity was measured by the number of major and minor idea units in a given text. Ellis and Barkhuizen’s definition of idea unit was used as a guide to code propositional complexity. In their definition, an idea unit is defined as “a message segment consisting of a topic and a comment that is separated from contiguous units syntactically and/or intonationally” (p.152). Idea units were further differentiated on the basis of “major” and “minor” idea units, with major idea units consisting of those which “convey the essential content of the message”, and minor idea units consisting of “those which relate to details that embellish the message but are not essential” (p. 152). Major and minor idea units were identified and coded according to Ellis and Barkhuizen’s definition based on the discretion of the researcher. An example of this kind of coding is given below:

Excerpt from Time 2: Belinda to Michelle:

“I like the name of your University, it’s pretty, mine is too formal”

Major idea units (1): “I like the name of your university”

Minor idea units (2): “it’s pretty”, “mine is too formal”

As this method of analysis relies heavily on the judgment of the researcher, a second rater (the same rater used for the error analysis) assessed the coding decisions in order to obtain a more objective measure. Using the first thirty lines of the dataset, the raters found discrepancies in only four instances in their coding of major and minor idea units. Similar to the procedure that was completed for error analysis, the strong agreement between the raters in coding the idea units was interpreted as evidence of a reliable system of coding. After establishing the reliability of the coding system, each of the participants’ emails was broken into units of analysis based on idea units. After this, major idea and minor idea units were identified and coded within each idea unit. To prevent each type of idea unit from being identified as a confounding variable, each number of total major and minor idea units in a text was divided by the total number of units of analysis in the text. The resulting number provided a measurement that was unadulterated by differences in the length of the text. After calculating the adjusted measurements for major and minor idea units, these measures were examined within participants as well as between participants over time. A line graph was generated to visualize the frequency of major and minor idea units over time to compare patterns of frequency both within and between participants.

Qualitative Analysis

A visualization of the data revealed patterns in the frequency of errors as well as the number of idea units generated by the participants in this study. After studying the patterns in the data, the questions that guided the nature of the qualitative analysis were “what is the relationship between major and minor idea units and error rate?” and “are there any patterns in the kinds of errors that arise in major and minor ideas?” To answer these questions, the text was

revisited in order to more closely examine several features related to the error, including the context of the error, the type or classification of the error, and the positioning of the error in relation to major and minor ideas in a written unit.

RESULTS

Error Analysis

After coding the errors, each error was tallied according to error type category and the time (Tables 1 and 2) for both participants.

TABLE 1
Error categories for Michelle

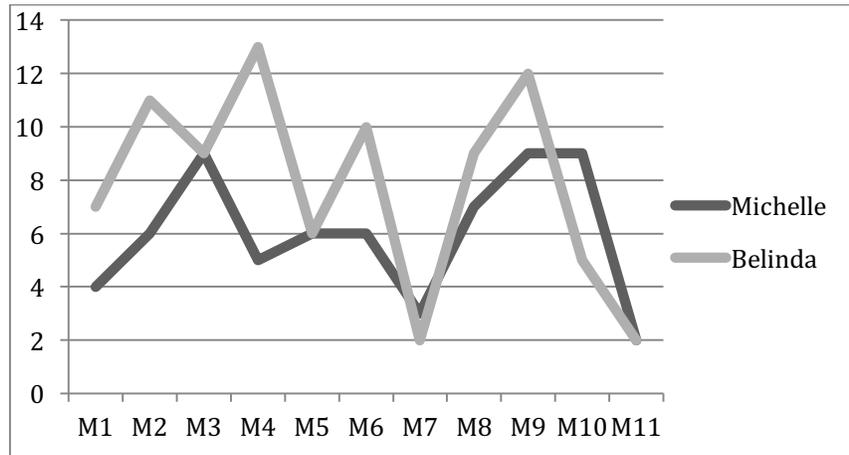
| | Punctuation | Adverb | Prepositions | Possession | Vocabulary | Plurality | Run on sentence | Tense | Pronoun | Article | Word order | Sum |
|-----|-------------|--------|--------------|------------|------------|-----------|-----------------|-------|---------|---------|------------|------------------|
| M1 | 3 | 1 | | | | | | | | | | 4 |
| M2 | 6 | | | | | | | | | | | 6 |
| M3 | 4 | | | | | | 3 | 1 | 1 | | | 9 |
| M4 | 2 | | 2 | | | | 1 | | | | | 5 |
| M5 | 3 | | 1 | | 1 | 1 | | | | | | 6 |
| M6 | 3 | | | 1 | | | 2 | | | | | 6 |
| M7 | 3 | | | | | | | | | | | 3 |
| M8 | 1 | | | | 1 | | 3 | | | | 2 | 7 |
| M9 | 7 | | 1 | | | | 1 | | | | | 9 |
| M10 | 8 | | | | 1 | | | | | | | 9 |
| M11 | 2 | | | | | | | | | | | 2 |
| | | | | | | | | | | | | Total: 51 |

TABLE 2
Error categories for Belinda

| | Punctuation | Adverb | Prepositions | Possession | Vocabulary | Plurality | Run on sentence | Tense | Pronoun | Article | Word order | Sum |
|-----|-------------|--------|--------------|------------|------------|-----------|-----------------|-------|---------|---------|------------|-------------------|
| B1 | 4 | | 3 | | | | | | | | | 7 |
| B2 | 2 | 2 | 2 | 1 | 3 | 1 | | | | | | 11 |
| B3 | 1 | 1 | 1 | | | 2 | 2 | 2 | | | | 9 |
| B4 | | 2 | 1 | 1 | 1 | 2 | 1 | 2 | | 3 | | 13 |
| B5 | | | | | 3 | 1 | | 1 | | | 1 | 6 |
| B6 | 5 | | | | 2 | 1 | | | | 1 | 1 | 10 |
| B7 | | | | | 2 | | | | | | | 2 |
| B8 | 1 | | | | 4 | 1 | | 3 | | | | 9 |
| B9 | | | 2 | | 3 | 2 | 1 | 2 | | 2 | | 12 |
| B10 | 5 | | | | 1 | | | | | | | 6 |
| B11 | | | | | | | | | | | | 0 |
| | | | | | | | | | | | | Total: 103 |

An analysis of errors between participants revealed that overall, Belinda had a higher frequency of errors in comparison to Michelle (Figure 1). This was the case for all but one of the eleven time periods examined in this dataset. The one exception, Measure 7, revealed that Michelle committed a higher number of errors than Belinda.

FIGURE 1
Error rates over time



Both participants displayed dynamic patterns of error rates, which tended to peak during the initial exchanges (M2-4) as well as the final exchanges (M8-10). Both participants displayed a drop in the number of errors they committed during M7. When assessed using a within participants approach, these interlocutors differed in the kinds of errors they committed. In totality, Michelle committed 61 errors. The majority of Michelle's errors were related to punctuation (51) and run on sentences (10). Belinda displayed a total number of 103 errors, with the majority of those related to punctuation (18), vocabulary choice (19), plurality (10), and tense (10). When compared, Belinda's errors were more diverse and occupied a broader range of error category types than Michelle.

Propositional Complexity

Overall, Belinda showed a higher frequency of major idea units while Michelle displayed a higher frequency for minor idea units (Figure 2). Both participants displayed similar peaks of major idea units around Time Measures (M) 1, 5-6 and 9-10. Michelle maintained a relatively consistent degree of total major idea units usage in her emails, with a spike in the number of major idea units supplied between M6-7. The pattern exemplifying the total number of major idea units for Belinda was much more erratic, with the highest number of major idea units produced in the beginning, two major declines at M4 and M8, and two major spikes in the total number of major idea units encoded between M5-6 and M9-10. Overall, both participants demonstrated higher frequencies for expressing minor ideas than major ideas (Figure 3).

FIGURE 2
Frequency of major idea units over time

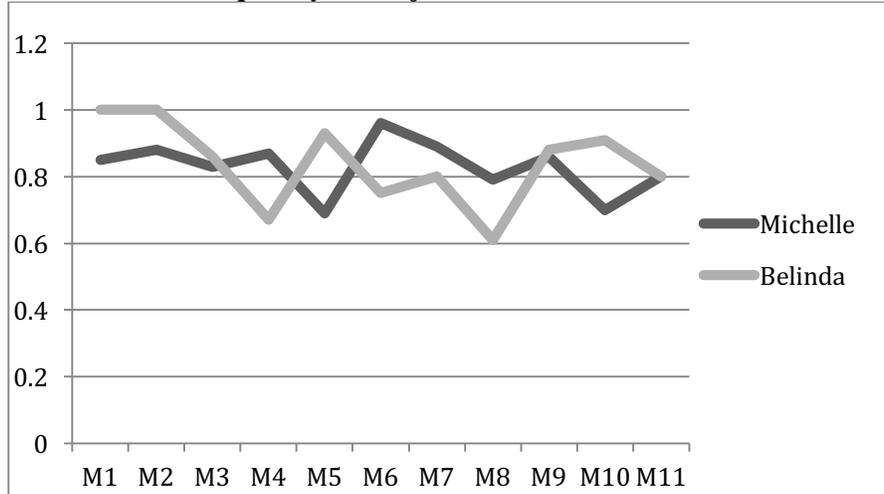
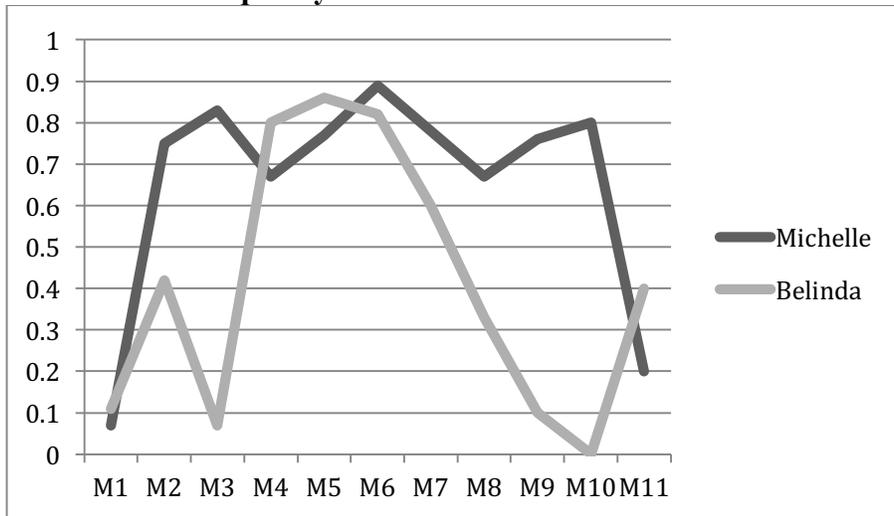


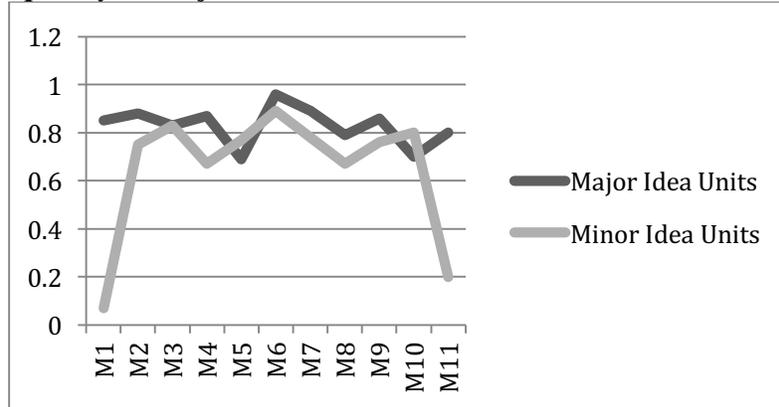
FIGURE 3
Frequency of minor idea units over time



For minor idea units, both participants experienced a peak in the number of minor idea units expressed from M4-M7 as well as low overall frequencies in the beginning of their email exchange.

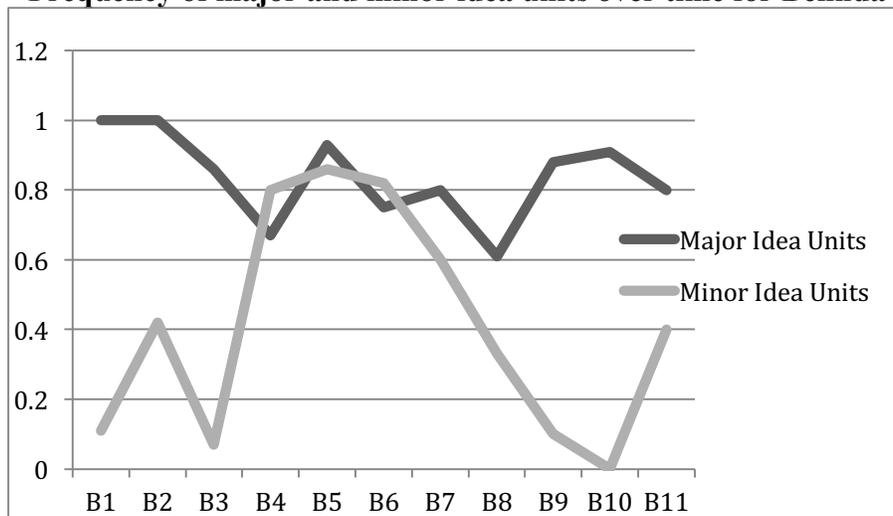
The participants experienced a tremendous amount of differences in the patterns of frequency when analyzed according to minor idea units. Michelle maintained a overall high rate of the minor idea units during the entire course of the project, only displaying a lower frequency of minor idea units during the beginning and end of her email correspondence (Figure 4).

FIGURE 4
Frequency of major and minor idea units over time for Michelle



Conversely, Belinda lagged behind in the overall frequency with which she used minor idea units. The exception to this trend was towards the middle of the email correspondences (M4-M7) and at the very end of the exchange (M11) (Figure 5).

FIGURE 5
Frequency of major and minor idea units over time for Belinda



Qualitative Analysis

In addition to error analysis and an analysis of propositional complexity, a third type of analysis was used in order to gain a greater understanding of the patterns of language use and development in the dataset. The qualitative analysis set out to answer two questions, “what is the relationship between major and minor idea units and error rate?” and “are there any patterns in the kinds of errors that arise in major and minor ideas?”. In both research questions, “relationship” is thought of as any consistent, emerging pattern that is observed in the data. In answering the first question, the researcher found that there was no consistent pattern in the relationship between major idea units and error rate for both participants. For minor idea units, however, Belinda experienced a drop in the total frequency of errors when she expressed more

minor idea units. For Michelle, no a consistent pattern emerged in the relationship between the number of minor idea units and error rate. However as was mentioned in the previous analysis section on prepositional complexity, both participants experienced a reduction in error rate and an increase in minor idea units from M4-M7.

In order to explore the nature of these patterns, a qualitative analysis was conducted to offer some context and explanations for the data. When examined more closely, the context of the data reveals that in these emails, the participants frequently shifted topics as they provided answers to a number of questions mentioned in previous emails. The nature of the topics varied greatly between participants and were connected to the participants' unique histories and personal experiences. Overall, the topics and questions related to a discussion of movies, music, future goals, and descriptions of events in their hometown.

In answering the second research question, which sought to determine whether any patterns exist in the kinds of the errors that occur within major and minor idea units, it was found that Michelle, who exhibited the highest error rates during M3 and M8-9, committed a higher number of punctuation and run on sentences errors during this exchange. When compared to her patterns in expressing major and minor idea units, Michelle exhibited a fewer number of major idea units; however, a larger number of minor idea units were expressed during these time periods.

Belinda, on the other hand, exhibited a rich variety of errors, ranging from punctuation and run on sentences to errors related to plurality, tense, and possession. When compared to the number of major and minor idea units expressed, Belinda expressed fewer major idea units during the times when she committed the most errors (M4, M8-9). When examining the total number of minor ideas, this issue is more complicated, because during M4, Belinda expressed a high amount of minor ideas but later from M8-9, she expressed a lower amount of idea units. Based on this observation, a second look at the context was needed in order to offer an explanation. M4 is an email exchange that is characterized by a frequent shifting between topics. In the context of one email, Belinda shifts from commenting on Michelle's news that she was sick ("I am sorry to hear that you got sick last week. Are you feel all right now?"), to the Japan disaster ("As to the hot topic of Japan disaster, according to official announcement, we are not affected by Japan's nuclear leakage by now"), to movies ("And movies.....", "I love Romance"), and finally to recommending movies to Michelle ("Ah, yes, I recommend you two movies I saw yesterday"). In contrast to the M4, M8-9 are characterized by a relatively homogenous discussion of topics. In these exchanges, Belinda was not shifting topics as frequently as can be seen in M4, instead she remained committed to discussing a central theme (for example, M8: holidays in China, M9: Mount Lao adventure).

DISCUSSION AND CONCLUSION

The results of this study revealed intriguing patterns in error rates and propositional complexity over time. Perhaps the most interesting finding from the results was that both participants experienced a drop in error rate and an increase in the production of minor idea units from M4-M7. When examined more closely, there was a negative correlation between error rate and frequency of minor idea units for Belinda during this time, which was unusual for her. Outside of the M4-M7 time period, Belinda did not display a high frequency of minor idea units, and overall had a higher percentage of error rates during the times before and after M4-M7.

From a CDST perspective, this could be interpreted as potential evidence of competition within the linguistic subsystems of the learner. A qualitative analysis was unable to reveal an explanation that could sufficiently account for a rise in minor idea units during this period; however, there are several possible explanations that could account for this.

First, a learner's knowledge about a particular topic may have allowed her to produce more varied and complex language, which might account for the rise in minor idea units. For example, Belinda's discussion of holidays in China in M8 may have included more minor idea units because she possessed more knowledge with which to comment and was able to add to the details related to major idea units proposed in this message. An additional explanation that could account for the complexity that was observed in Belinda's writing at this time is related to the idea of learner investment. An increase in Belinda's minor idea units might have been prompted by her motivation to share additional details with her conversation partner due to her investment in the topic. For example, Belinda's description of her Mount Lao adventure in M9 could have included more minor idea units because she felt strongly motivated to share and communicate her experience to her partner. This could have been attributed to an intrinsic reason on the part of the learner (i.e. desire to convey a meaningful experience) or due to properties of the conversation itself (i.e. the affective and social environment encouraged the participant to share and elaborate more frequently).

A third explanation concerns the influence of the interlocutor in the exchange. In contrast to Belinda, Michelle displayed relatively stable trends in the frequency with which she used minor idea units. Perhaps influenced by the nature of the interactional exchange, Belinda began to model the writing style of her interlocutor by employing the use of minor idea units in her writing. However, further interactional data is needed to test this claim.

Using three different methods of analysis-error analysis, a measure of propositional complexity, and a qualitative analysis- data from an asynchronous mediated exchange between a native and nonnative speaker of English revealed several interesting findings when analyzed from a CDST perspective. Consistent with previous research examining accuracy and complexity from a CDST perspective, language production exhibited dynamic and complex patterns of variation overtime, displaying fluctuating changes in accuracy and propositional complexity over time. These patterns of change were visible between individuals as well as within individuals in the study. Furthermore, this dataset revealed an interesting negative correlation between error rate and minor idea units, suggesting that complexity and accuracy measures could be connected growers in a complex system. Connected growers are a key concept in CDST and are defined as linguistic subsystems that develop symbiotically as they interact. Perhaps the fluctuations observed in error rate and minor idea units represent evidence of a system in transition; one that reflects how input from interaction can alter patterns of language use.

One of the most intriguing patterns of interaction in this dataset concerned the discovery of an overall decrease in error rate and an increase in minor idea units from M4-M7 for both participants, as well as the negative correlation between error rate and frequency of minor idea units for Belinda. One potential explanation for this finding lies in the nature of the correspondence between the participants. Perhaps after several turns of electronic correspondence, Belinda might have been linguistically primed to assume the style of her interlocutor, namely correspondence characterized by a high overall frequency of minor idea units. Influenced by linguistic priming, Belinda's style of writing may have resulted in a number of changes, including shifts in propositional complexity and accuracy. Claims related to the role of linguistic priming in Belinda's writing and questions regarding why only one participant, and

not both, were influenced by this effect would require a lengthier account of the email correspondence. In all, future research is needed to examine these claims and refine the application of CDST to learner data.

REFERENCES

- Corder, S. P. (1974). Error analysis. *The Edinburgh course in applied linguistics*, 3, 122-131.
- De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and cognition*, 10(01), 7-21.
- Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford University Press, USA.
- 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), 590-619.
- Levelt, W. J. H. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 585-604.
- Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied linguistics*, 35(2), 184-207.
- Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31(4), 532-553.
- 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.