DYNAMIC SECOND LANGUAGE DEVELOPMENT:
THE INTERACTION OF COMPLEXITY, ACCURACY, AND FLUENCY
IN A NATURALISTIC LEARNING CONTEXT

A Dissertation
Submitted to
The Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY IN EDUCATION

By
Elizabeth A. Hepford
May 2017

Examining Committee Members:
Dr. Aneta Pavlenko, Dissertation Advisor, Center for Multilingualism, University of Oslo
Dr. Elvis Wagner, Dissertation Advisory Committee Chair,
Department of Teaching and Learning, Temple University
Dr. Paul Toth, Dissertation Committee Member, Department of Spanish and Portuguese,
Temple University
Dr. Diane Larsen-Freeman, Dissertation Committee Member, Professor Emerita,
University of Michigan
Dr. Barbara A. Wasik, Dissertation Examining Committee Chair and External Reader,
Department of Psychological Studies in Education, Temple University
ABSTRACT

The purpose of this study was to examine the second language development of a native-speaker of Spanish learning English over a period of 15 months. More specifically, I explored the interaction of complexity (advanced forms of grammar and vocabulary), accuracy (grammatical and semantic), and fluency, commonly referred to as the CAF constructs. While findings in CAF literature tend to focus on one construct using experimental or cross-sectional studies (Bulté & Housen, 2012; Kormos & Dénes, 2004; Vyatkina, 2012), this case study investigated non-linear and interconnected CAF development, periods of fluctuation, and the effects of motivational factors on 14 variables. In order to explore the data as a system developing simultaneously, Complex Dynamic Systems Theory (CDST) (Larsen-Freeman, 1997; 2006) was applied as the theoretical framework. Through CDST’s theoretical lens and the tools developed for it (Verspoor, de Bot & Lowie, 2011), I found that knowledge variables (lexical diversity, accuracy, and elaboration) maintained consistent correlations, whereas their relationship with fluency variables (speed, repairs, and pauses) changed based on the cognitive strain the learner was experiencing at the time. I also found that the learner shifted his focus between the knowledge variables and that the complexity and accuracy variable on which he chose to focus appeared to be affected by changing motivational factors.
ACKNOWLEDGEMENTS

Much like a complex dynamic system, this dissertation is the result of many interacting agents. Without each of their contributions, my path would be considerably different and less successful in a thousand ways.

First and foremost, I would like to thank my adviser, Dr. Aneta Pavlenko, who patiently guided me as I explored research topics, learned the complexities of academic writing, and gradually transformed into a scholar. I am indebted to her for offering her well-timed advice while allowing me to follow my own path. She will always be a role model of dignity, grace, and determination and her wise words will guide me as a scholar and a friend throughout my career and life.

I would also like to thank the members of my committee for their time, thoughtful feedback, and encouragement. I am extremely grateful to Dr. Elvis Wagner for filling in as my administrative chair, pushing me to critically think about my methodology, and taking the time to discuss the more complex variables I chose. I’m thankful to Dr. Paul Toth for increasing my grammar analysis abilities through his classes and pointing me to the exact book I needed for the analyses conducted as part of this study. Finally, I am grateful to Dr. Diane Larsen-Freeman for agreeing to be on my committee, offering advice and feedback based on her vast knowledge of second language acquisition, and for introducing me and the rest of the linguistics world to Complex Dynamic Systems Theory. All of their contributions were invaluable to the completion of this project.
A great deal of my support system and assistance in gathering data came from my colleagues at Temple University. For helping me gather the data, I would like to thank Kimberly Mitchell, Nina Campellone, and Felicia Potter. In addition, I’d like to thank my second raters and translator for contributing their expertise: Anastasia Sorokina, Kimberly Mitchell, Yohana Gil Berrio, and Will Bevins. I will also always be grateful to my fellow doctoral students Sara Kangas, Anastasia Sorokina, Brooke Hoffman, and Lorraine Sova for years of support, proofreading, and friendship.

I am continually grateful to my mother, Marie Hepford, for forming me into who I am and being my cheerleader through the process. Without her example of determination and strength, I never would have arrived where I am. I will always be in her debt for the countless hours of proofreading this dissertation. I am also thankful for my Philadelphia family, Dolly Ketterer and Florence Fee, who kept me housed in desperate times, well-fed, and blessed with their friendship.

Next is the star of the show, “Juan”, who will remain anonymous. Thanks to his time, effort, and willingness to let me analyze his language and life for 15 months, I not only found a dissertation, but a friend.

Last, but definitely not least, my backbone, lifeline, and husband, Jim Poe. Without his support, patience, and endless proofreading, I never could have finished this endeavor. I am grateful for all the sacrifices he has made supporting me as my partner and my voice of reason along the bumpy road of our well-traveled life.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2 SECOND LANGUAGE DEVELOPMENT IN A NATURALIST CONTEXT LITERATURE REVIEW</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Complexity, Accuracy and Fluency (CAF)</td>
<td>4</td>
</tr>
<tr>
<td>2.1.1 Complexity</td>
<td>6</td>
</tr>
<tr>
<td>2.1.2 Accuracy</td>
<td>15</td>
</tr>
<tr>
<td>2.1.3 Fluency</td>
<td>21</td>
</tr>
<tr>
<td>2.2 Language Processing Models and Attentional Resources</td>
<td>27</td>
</tr>
<tr>
<td>2.2.1 Task-Based Learning</td>
<td>34</td>
</tr>
<tr>
<td>2.3 Naturalistic L2 Learning Context</td>
<td>39</td>
</tr>
<tr>
<td>2.4 Language Learning Motivation</td>
<td>45</td>
</tr>
<tr>
<td>2.5 Longitudinal Case Studies of Natural L2 Learning</td>
<td>50</td>
</tr>
<tr>
<td>2.6 Summary of the Literature Review</td>
<td>56</td>
</tr>
<tr>
<td>CHAPTER 3 THEORETICAL FRAMEWORK</td>
<td>57</td>
</tr>
<tr>
<td>3.1 Complex Dynamics System Theory CDST</td>
<td>58</td>
</tr>
<tr>
<td>3.2 CDST Methods and Studies</td>
<td>66</td>
</tr>
</tbody>
</table>
3.3 Gaps in the Current Research ................................................................. 72
3.4 Research Questions .................................................................................. 75

CHAPTER 4 METHODOLOGY ........................................................................... 77

4.1 Participants ............................................................................................... 77
  4.1.1 Juan ........................................................................................................ 77
  4.1.2 Comparison Group .............................................................................. 78

4.2 Research Design ....................................................................................... 79
  4.2.1 Data Elicitation ................................................................................... 79
  4.2.2 Data Elicitation Stimuli ....................................................................... 81

4.3 Data Coding and Calculations ................................................................. 86
  4.3.1 Sample Preparation .......................................................................... 86
  4.3.2 CAF Variable Coding ......................................................................... 90

4.4 Data Analysis ............................................................................................ 105
  4.4.1 Second Language Development Analysis ....................................... 106
  4.4.2 Global Variable Interaction .............................................................. 109
  4.4.3 Environmental Factors .................................................................... 113
  4.4.4 Task Effects Analysis ....................................................................... 115

CHAPTER 5 RESULTS ...................................................................................... 117

5.1 Research Question 1 Development of CAF ........................................... 117
  5.1.1 Emerging Changes Analysis .............................................................. 119
  5.1.2 Fluctuation Analysis .......................................................................... 129
5.1.3 Error Analysis ................................................................. 133

5.2 Research Question 2: Global Variable Interaction .......................... 138
  5.2.1 Comparisons of Variables with Periods of Fluctuation ............... 139
  5.2.2 Correlations Between Complexity and Accuracy Variables ......... 142
  5.2.3 Comparisons Between Proceduralization and Knowledge Variables ............................................................................. 149

5.3 Research Question 3: Environmental Factors ................................ 159

5.4 Research Question 4: Task Effects ................................................ 168

CHAPTER 6 DISCUSSION AND CONCLUSION ......................................... 173

6.1 Second Language Development in a Naturalistic Environment .......... 173
  6.1.1 Research Question 1: Development of CAF .............................. 173
  6.1.2 Research Question 2: CAF Global Variable Interaction ............ 177
  6.1.3 Research Question 3: Environmental Factors ......................... 179
  6.1.4 Research Question 4: Task Differences .................................. 184

6.2 Theoretical and Methodological Implications .................................. 186
  6.2.1 CDST Tools for Analysis ...................................................... 186
  6.2.2 Comparison Group ................................................................ 190

6.3 CAF Methodological Considerations ............................................ 192
  6.3.1 Complexity ........................................................................... 192
  6.3.2 Accuracy .............................................................................. 195
  6.3.3 Fluency ................................................................................ 196
6.4 Limitations of the Study........................................................................................................... 198

6.5 Conclusions and Future Directions for Research................................................................. 200

REFERENCES CITED ...................................................................................................................... 205

APPENDICES

A. MATERIALS FROM PICTURE DESCRIPTION SETS 1-3 .................................................. 218

B. CHAT COES FROM MACWHINNEY (2000)............................................................................ 221

C. RESULTS FROM ERROR ANALYSIS .................................................................................... 222

D. MOTIVATIONAL FACTORS TIMELINE.................................................................................. 225

E. SAMPLE TRANSCRIPTS FROM ANALYSIS ......................................................................... 229

F. EXCERPTS FROM INTERVIEWS DURING SESSIONS ..................................................... 238

G. PERMISSION TO REPRINT LEVELT’S 1999 MODEL OF SPEAKING ...................... 246
LIST OF TABLES

1. Summary of the Trade-off Hypothesis ................................................................. 34
2. Task list ................................................................................................................. 85
3. Analytical variables ............................................................................................... 90
4. Frequency of correlated data pairs and Spearman’s Correlations ...................... 150
5. Colloquial speech in Juan’s transcripts .............................................................. 160
6. Significant task differences in Juan’s data ............................................................ 168
7. Task differences between comparison group members ......................................... 169
8. Mean and Standard Deviation for Juan’s tasks ..................................................... 170
9. Mean and standard deviation for comparison group’s tasks .............................. 171
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Levelt’s 1999 Model of Speaking</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>Example of a TTR curve for a single sentence</td>
<td>92</td>
</tr>
<tr>
<td>3</td>
<td>Representation of the original TTR Curve Chart produced by vocd</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>Lexical diversity development 0 – 1 values</td>
<td>120</td>
</tr>
<tr>
<td>5</td>
<td>Subordination 0-1 development 0 – 1 values</td>
<td>120</td>
</tr>
<tr>
<td>6</td>
<td>Silent pauses development 0 – 1 values</td>
<td>121</td>
</tr>
<tr>
<td>7</td>
<td>Filled pauses development 0 – 1 values</td>
<td>122</td>
</tr>
<tr>
<td>8</td>
<td>Total pauses development 0 – 1 values</td>
<td>122</td>
</tr>
<tr>
<td>9</td>
<td>Second per silent pauses 0 – 1 values</td>
<td>123</td>
</tr>
<tr>
<td>10</td>
<td>False starts development 0 – 1 values</td>
<td>123</td>
</tr>
<tr>
<td>11</td>
<td>Speech rate development 0 – 1 values</td>
<td>124</td>
</tr>
<tr>
<td>12</td>
<td>Redundancies development 0 – 1 values</td>
<td>124</td>
</tr>
<tr>
<td>13</td>
<td>Total repair disfluencies development 0 – 1 values</td>
<td>125</td>
</tr>
<tr>
<td>14</td>
<td>Errors per AS-unit development 0 – 1 values</td>
<td>127</td>
</tr>
<tr>
<td>15</td>
<td>Elaboration development 0 – 1 values</td>
<td>127</td>
</tr>
<tr>
<td>16</td>
<td>Reformulations development 0 – 1 values</td>
<td>128</td>
</tr>
<tr>
<td>17</td>
<td>Articulation rate development 0 – 1 values</td>
<td>128</td>
</tr>
</tbody>
</table>
18. Min-max graph for lexical diversity 0 – 1 values ................................................................. 131
19. Min-max graph for filled pauses 0 – 1 values ........................................................................ 131
20. Min-max graph for errors per AS-unit 0 – 1 values ................................................................. 132
21. Min-max graph for subordination 0 – 1 values ...................................................................... 133
22. Past tense usage and errors .................................................................................................... 135
23. Present progressive usage and errors ...................................................................................... 136
24. Min-max graph for lexical diversity z-scores .......................................................................... 140
25. Min-max graph for elaboration z-scores .................................................................................. 141
26. Min-max graphs for errors per AS-unit z-scores ..................................................................... 142
27. Moving correlation between lexical diversity and elaboration .............................................. 144
28. Lexical diversity and elaboration trajectories .......................................................................... 144
29. Moving correlation between elaboration and accuracy .......................................................... 147
30. Elaboration and errors per AS-unit trajectories ....................................................................... 147
31. Moving correlation between lexical diversity and accuracy .................................................. 149
32. Moving correlations between accuracy and speech rate ....................................................... 151
33. Moving correlation between elaboration and total pauses ..................................................... 152
34. Moving correlation between accuracy and total pauses ......................................................... 154
35. Moving correlation between accuracy and total repairs ....................................................... 154
36. Moving correlation between elaboration and speech rate.............................155
37. Moving correlations between elaboration and total repairs..........................155
38. Moving correlation between lexical diversity and total pauses......................156
39. Moving correlation between lexical diversity and total repairs......................156
40. Moving correlation between lexical diversity and speech rate.......................1577
CHAPTER 1

INTRODUCTION

After four years of living in a Mexican community, I was comfortable, happy, and an honorary member of a large close-knit family. This is why I was so surprised when one of my friends asked me why I had not learned Spanish yet, even though we were having the conversation in Spanish. As Schmidt (1983) found with “Wes”, my grammar had not progressed beyond basic communicative functions despite my interaction with Spanish speakers and acculturation into the Mexican society. In my case, a busy English as a Foreign Language teaching schedule did not allow time for taking explicit Spanish instruction courses and most of my friends and acquaintances had stopped correcting me. This friend had not seen me in a few months and was reminded of how poor my Spanish grammar was when he heard me speak again. Even though I could converse on most topics, by his standards, I had not “learned Spanish”.

Like my friend, most people assume that being immersed in a second language (L2) will automatically result in acquisition of the L2. However, second language development (SLD) research shows that it is not that simple. Naturalistic language development may be affected by the necessity of speaking the language, the learner’s interest (or lack thereof) in integrating into the culture (Schumann, 1978; Shapira, 1978), their attitudes toward the language being acquired and those who speak it (Kinginger, 2008; Masgoret & Gardner, 2003), exposure to native-speakers (Elis, 2009; Regan, 2004), the learners’ motivation level (Dörnyei, 2005; Gardner, 1968) and opportunities for positive interactions with target language speakers (Norton, 2000; Piller & Takahashi, 2006).
Despite decades of research dedicated to L2 immersion, many questions about SLD through naturalistic immersion remain unanswered: Is it possible for an unexceptional, uninstructed, and immersed learner to develop grammatical accuracy, fluent speech, and a sufficient vocabulary? Do accuracy, fluency, and vocabulary develop simultaneously or separately? What factors affect the development of those variables?

To answer these questions, one recent trend is to explore the interaction between complexity, accuracy, and fluency (CAF) (Larsen-Freeman, 2006; Spoelman & Verspoor, 2010; Tavakoli & Foster, 2011). Complexity refers here to learners’ ability to incorporate a variety of grammatical structures, diverse vocabulary, and rich descriptive language (adjectives, adverbs, relative clauses, etc.); accuracy refers to the ability to produce grammatically and semantically correct sentences; and fluency refers to learners’ ability to produce uninterrupted fluid speech at an acceptable pace.

Another recent theoretical development in the study of how learners use a language is the adaption of Complex Dynamic System Theory (CDST) from the natural sciences. In the field of Second Language Acquisition (SLA), CDST works under the premise that language development is non-linear, self-organizing, and affected by multiple interconnected elements (Larsen-Freeman & Cameron, 2008). CDST provides a framework to explore developmental trajectories, interaction between variables, and the effect of the social environment and cognitive systems on L2 development. To explore this development, many CDST researchers have incorporated the CAF measures as dependent variables and called for longitudinal studies that provide detailed descriptions
of learners’ development, close examinations of the CAF variables, and the factors affecting them (Housen & Kuiken, 2009; Larsen-Freeman, 2006; Skehan, 2014; Verspoor, de Bot, & Lowie, 2011).

Following this call, this longitudinal case study of a native speaker of Spanish (Juan) learning English over 15 months addresses three underdeveloped areas of research: longitudinal CAF development, interaction between the CAF variables, and the effects of environmental factors on learners’ development and on interactions between the variables. When Juan arrived in the US, he was at a low-intermediate level of English. Study abroad research indicates that at this level, learners have enough L2 knowledge to benefit from the proceduralization effects of naturalistic learnings (Towell, Hawkins, & Bazergui, 1996; Regan, 2004; Kinginger, 2008). In other words, practice and exposure would lead to reinforcing his knowledge and increasing his fluency. In addition, due to the nature of individual case studies and frequent data collection, Juan and I discussed his struggles, achievements, life events, and attitudes related to his immersion experience. By placing motivational factors on a timeline, such as upcoming exam dates and struggles to integrate into his environment, it was possible to compare his CAF scores to those motivational factors. Thus, the results analyzed Juan’s development as a complex dynamic system with interconnected variables and as such contribute to the current literature on CDST and SLD.
In order to provide a comprehensive review of the elements affecting Juan’s complex language development, this literature review covers five areas. It begins with the focus of the study, the CAF variables. That will be followed by an overview of the factors that appeared to affect the participant’s CAF variables including cognitive strain during language processing, naturalistic learning environments, and motivation to learn an L2. Lastly, the benefits of longitudinal research will be discussed. While the variables are normally discussed within the methodology section, the scope and controversy related to the CAF variables necessitates a thorough review of the current literature. The literature review is followed by an explanation of the theoretical framework, Complex Dynamic Systems Theory (CDST), which was used to analyze the elements as a system. The CDST section will also contain a review of CDST studies on language development, which aid in explaining the theory and the tools utilized in CDST studies.

2.1 Complexity, Accuracy and Fluency (CAF)

The measures of complexity, accuracy, and fluency (CAF) emerged from researchers’ interest in developing a balanced evaluation of learner language development. While standardized ESL examinations such as the TOEFL and IELTS tend to separate proficiencies by the tasks of reading, writing, listening, and speaking, the
CAF constructs focus on the knowledge and skills necessary to complete those tasks successfully. Though CAF first became an area of interest in task-based learning research (Skehan, 1996), over time the measures have evolved into their own research areas in the study of L2 development (Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009). Since the CAF variables are primarily a measure, rather than a theoretical framework, their usage has rapidly expanded across different areas of research. Consequently, an effective measure in one research context may be ineffective in another. For example, increasing subordinate clauses has been shown to be a more effective developmental measure in oral production than written production in intermediate to advance learners. Moreover, new purposes for and variables within the CAF constructs have sparked debates among scholars regarding their validity and ability to adequately represent the construct (Housen & Kuiken, 2009; Lambert & Kormos, 2014; Pallotti, 2009). Housen and Kuiken (2009) point out that “critical surveys of the available tools and metrics for gauging CAF have revealed various problems, both in terms of the analytic challenges which they present and in terms of their reliability, validity, and sensitivity” (p. 464). Some of the current controversies, however, may stem from the diversity of modalities, purposes of the research, and frameworks in which they have been incorporated. Within this section, I address some of these controversies, define the constructs and the variables within them, and discuss the latest developments related to them. Even though the constructs will be discussed individually, many of the studies analyzed more than one construct. Thus, the
discussion will focus on the construct being reviewed and may additionally include the comparison between constructs.

2.1.1 Complexity

*Complexity* refers to “the extent to which learners produce elaborated language” and “the learner’s preparedness to use a wide range of different structures” (Ellis & Barkhuizen, 2005, p. 139). CAF scholars commonly differentiate between three types of complexity. The first, *structural or syntactic complexity*, includes variety in types of clauses, descriptive modifiers, and verb forms. The second, *lexical complexity*, refers to diversity and sophistication related to vocabulary. The last, *cognitive complexity*, refers to characteristics inherent to the task that may make it difficult for a speaker to perform the task. This section will cover the first two types of complexity, whereas cognitive complexity will be discussed in the section 2.2 on language processing.

2.1.1.1. Syntactic complexity. *Syntactic complexity* refers to “the range of syntactic forms that surface in language production and the degree of sophistication of such forms” (Ortega, 2003, p. 493). In theory, as L2 learners progress, they will incorporate more complex structures in their speech. Wolfe-Quintero, Inagaki, and Kim (1998) argue that written English complexity develops in stages. First, fragments become clauses. Then, learners add coordinated clauses, followed by subordinated clauses, and then reduced subordinate clauses. In the last stage, learners begin using nominalization. Ortega (2012)
suggests that the metrics of syntactic complexity show “a strong reliable relationship with proficiency” which is a “useful shortcut to gauge global proficiency” (p. 129). Although research into variables that reliably measure complexity continues (Bulté, Housen, Pierrard, & Van Daele, 2008; Vyatkina, 2012), most recent research includes two measures of syntactic complexity to measure development: clauses per unit of analysis (subordination) and words per unit of analysis (elaboration). This section will discuss the findings and controversies related to each of them.

Subordination is defined as the number of dependent clauses per measurement unit (Ellis & Barkhuizen, 2005). In Wolfe-Quintero et al.’s (1998) overview of CAF measures, they found that clauses per t-unit (independent clauses plus subordinate clauses associated with it) increased linearly related to learners’ proficiency levels regardless of task, target language, significance, or how proficiency was defined. Unlike subordination, elaboration tends to be measured holistically, in the form of total number of words per unit rather than a specific type of clause. While elaboration includes subordination, it also measures other types of complexity such as prepositional phrases, adjectives, adverbs, and non-finite verb phrases (Norris & Ortega, 2009). As an example of the need for this measure, the utterance “the shaggy dog joyously escaped from the confining fence” contains the same number of clauses per unit as the utterance “the dog ran away”. The first utterance is clearly more detailed, complex, and developmentally advanced even though the subordination variable would not calculate them differently.
Although SLD research has shown correlations between proficiency and both elaboration and subordination, continued research is needed to define and standardize the variables, analyze differences in written and oral complexity, and document the development of complexity. A comparison between Vyatkina (2012) and Bulté and Housen (2014) exemplifies the issues with differences between writing levels. Vyatkina (2012) found significant increased subordination (subordinate clauses per 100 words) among a class of beginning L1 English L2 German university students’ essays over four semesters (the exact number of students varied slightly between semesters). However, elaboration was less clear. While she found significant increased sentence length over time, words per finite verb phrases did not have significant increases. In contrast, Bulté and Housen (2014) found significant increases in elaboration (words per sentence and words per t-unit) and did not find increases in subordination (sub-clauses per clause), in essays produced at the beginning and end of a four-month study. However, their study included 45 intermediate to advanced ESL learners’ essays as opposed to Vyatkina’s (2012) beginning students. Since both studies found increases in the majority of their other complexity measures, such as mean length of sentence and measures related to coordinating conjunctions and clause length, their participants appeared to increase in overall complexity. Bulté and Housen (2014) suggest that advanced learners use less subordination as they learn to write in a formal academic style.

Ortega (2012) highlights style differences related to synoptic style (highly formal written text) versus dynamic style (oral, informal, and nontechnical). In her evaluation of
the current complexity measures, she posits that synoptic style, which is often used in academic writing, utilizes more nominalization (changing a verb to a noun) and noun phrases than subordination. Thus, in advanced learners’ essays, subordination may decrease, which limits the usefulness of subordination as a measure. Biber, Gray and Poonpon (2011) explored the differences between written and oral subordination by comparing 429 academic research articles and 723 conversation samples from L1 English speakers. They found that conversation contained significantly more subordinate clauses describing verbs (that, WH, ing, and to clauses), while academic writing contained more nominal clauses (that, of, and to clauses) and adjective clauses (that, ing, and to clauses). They concluded that “the kinds of complexity common in academic writing are fundamentally different from the kinds of complexity common in conversation” (Biber et al., 2011, p. 29).

Chan, Verspoor, and Vahtrick (2015) investigated this issue further by comparing two 15-year-old L1 Chinese L2 English identical twins learning English for eight months. The researchers coded 100 written texts and 100 oral texts for elaboration (mean length of t-unit) and subordination (dependent clauses per t-unit). They found that subordination and elaboration were significantly greater in the oral data during the first part of the study. After the 32nd sample, subordination and elaboration were nearly equal between written and oral samples in one twin. In the other twin, the written scores were close, but not equal to the oral sample. However, the learners were beginners and may have never reached a synoptic level of writing.
Although additional research to clarify the differences between oral and written complexity is needed, scholars agree that subordination and elaboration are measuring different types of complexity. As Lambert and Kormos (2014) point out, only measuring subordination “potentially obscures developmental processes” (p. 608) in that subordination does not account for a learner’s ability to use descriptive speech, such as adjectives, adverbs and prepositional phrases. As a last point, Pallotti (2009) reminds researchers that complexity also includes the grammar forms chosen by the learner. When describing an event, learners have multiple alternatives to express themselves. Their point may be made communicatively clear with the simple past, but more eloquently expressed with the past perfect. Since the past perfect form is acquired later, it is developmentally more complex. Elaboration may be able to capture this difference in English in the form of additional function words (e.g. have been + verb, will be + verb). However, in languages that indicate aspect as part of verb morphology, additional measures need to be developed. As a more thorough alternative, micro-analyses such as analyzing specific grammar forms may prove to be more effective.

To address the various forms of syntactic complexity, this study utilizes two measures: subordination due to its correlation with proficiency gains in previous studies; and elaboration to capture increasing descriptive language. In addition, a micro-grammar analysis that includes increasingly complex verb forms is included as part of the accuracy analysis.
2.1.1.2. Lexical Diversity. Though not new to SLA, lexical diversity is a relatively recent addition to the measures in the complexity construct. As the name implies, lexical diversity is a measure of vocabulary richness and refers to the number of unique words used by a speaker (Ellis & Barkhuizen, 2005; Malvern & Richards, 2002). A great deal of research in the last decade has been dedicated to finding the best measure of lexical diversity.

At present, there are mainly three types of lexical diversity measures: a type token ratio (TTR), a corrected type token ratio (CTTR), and TTR curve-fitting procedure. TTR, the most basic measure, divides the number of unique words (type) by the number of words in the sample (token). However, studies have shown that TTR scores reduce as the sample length increases since longer texts are more likely to repeat function words and words related to the topic. Consequently, texts of different lengths cannot be validly compared with TTRs (Jarvis, 2002; McKee, Malvern, & Richards, 2000; Malvern & Richards, 2002). Several algebraic transformations to correct for text length (CTTRs) have been attempted with varying levels of success (Bulté & Housen, 2014; Bulté et al., 2008).

A more recent and reliable measure is “D”, a TTR curve-fitting procedure (Jarvis, 2002; McKee et al., 2000). The principle of a curve-fitting procedure is based on a TTR curve, which is formed from a text’s TTR plotted on a graph at one word (type/token = 1) and at each point after that through the entire text. The scores decrease in a curved line depending on the number of repeated words. Thus, if a person never repeated a word, the
score would be one resulting in a straight line TTR curve. If they repeated the same word over and over, the score would drop quickly and stabilize at slightly above zero. $D$ is based on the difference between a text’s TTR and the TTR of random samples of text taken from throughout the transcript. Of the curve-fitting procedures, the measure $D$ appears to best resolve the problem of differing text length and distinguishes lexical diversity differences between L2 proficiency levels (McKee et al., 2000). The calculation of $D$ will be explained in greater detail in the methodology section.

To validate $D$, McKee et al. (2000) took 3 samples from 38 L1 transcripts of children between 27 – 33 months old. The first sample was taken from all the words, the second on the even-numbered words, and the last on the odd-numbered words. They calculated a series of two-tailed paired t-tests that confirmed that the three samples were not significantly different. Thus, even though the texts contained a different set of words from the same speaker, the lexical diversity scores were similar enough to demonstrate the same level of lexical diversity, which should be expected from the same writer.

In order to compare the measures of lexical diversity, Bulté et al. (2008) tested 22 different types of lexical diversity measures for improvement among 19 L1 Dutch L2 French students between 12 -14 years old living in Brussels. They analyzed unplanned oral narration to a wordless picture story, *Frog Where Are You* (Mayer, 1969), once a year for three years. The 22 measures were divided into three types. The first were standard measures of lexical diversity including the Uber Index, Guiraud Index, (both CTTRs) and $D$. The next type, class measures, included productivity measures (number
of nouns, verbs, or adjectives produced), and lexical diversity measures, such as TTRs within specific parts of speech (e.g. number of unique verbs/total verbs) and TTRs related to frequency of a part of speech (e.g. number of verbs/total words). The last type, frequency, measures lexical sophistication by distinguishing between basic and advanced words. They found significant differences between the pre-test and the post-test on 20 measures (two frequency measures were not statistically significant). However, the L2 learners did not reach the benchmark set by an L1 French group of 19 students also residing in Brussels. Furthermore, they also found that while significant differences existed between the first and third year, only 12 measures improved between the first and second year, and only two measures improved significantly between the second and third year. They tentatively proposed that task repetition may have decreased the participants’ motivation during the last year and suggested that future studies include motivational and attitudinal information.

Bulté et al. (2008) also found significant positive correlations between all the measures. Particularly strong correlations were found between the Uber Index, Guiraud Index and $D$ ($r = .95 - .99$). Since the measures should have distinguished between lexical diversity, lexical sophistication, and lexical productivity, they concluded that all of the measures “tap into similar aspects or dimensions of lexical proficiency, which poses problems in terms of construct validity” (Bulté et al, 2008, p. 292). They suggested that further research is necessary to increase conceptual clarity and develop measures that can distinguish between the types of lexical measures.
Jarvis (2002) compared the reliability and validity of several measures of lexical diversity that could be adapted to the curve-fitting approach including: $D$, Herdan’s C, Guiraud’s R, the Uber Index, and Zipf’s Z. These measures were applied to a corpus of L2 English narratives, elicited by a silent film from 140 L1 Finnish students, 70 L1 Swedish students, and a control group of 66 L1 English students. All were between the ages of 10 and 15 years. He found that the Uber Index and $D$ had between 97% – 98% of good fits, which were calculated by a formula comparing the original TTR curves to the curves produced by the measures. Additionally, when only the content words were included, the Uber Index scores were 98.15%, while the $D$ scores were only 90.77%. Jarvis (2002) concluded that “further work is certainly needed to clarify when a close-fitting curve is close enough” (p. 81). In terms of development, both $D$ and the Uber Index had significant correlations with L2 level, years of instruction, and vocabulary scores. However, no significant results were found related to holistic quality ratings of the L1 comparison group.

In contrast, Yu (2009) examined correlations between $D$ and the holistic scores given to students taking the Michigan English Assessment Battery Exam. The analysis of 201 compositions and 25 interviews concluded that $D$ was significantly and positively correlated with both proficiency and overall quality. $D$ was also a significant predictor of the overall quality rating of compositions. Additionally, significant differences were found between the topics. For example, $D$ was greater in impersonal topics and in topics
that were familiar to test takers. In contrast, the modality (written or oral) or L1 (Filipino or Chinese) did not produce significant differences in D.

While all the complexity measures in the literature review have shown connections to L2 development, some controversies still exist related to the validity of complexity measures themselves. This study contributes to the literature by providing longitudinal data for elaboration, subordination, and $D$.

2.1.2 Accuracy

Pallotti (2009) describes accuracy as the “the degree of conformity to certain norms” (p. 592). The notion of accuracy has a long history in the study of language including defining prescriptively correct standards for first language in the 18th century (Richards & Rogers, 2001). As second language learning became more relevant, methods for predicting errors a student might make developed. By 1960, research and pedagogy adopted contrastive analysis, which compares one language to another in hopes of predicting errors (Ellis & Barkhuizen, 2005; Richards & Rogers, 2001). This method is still used in modern classrooms as exemplified by Swan and Smith’s (2001) Learner English: A teacher’s guide to interference and other problems. However, the issue of prescriptive versus descriptive language has led to controversies in defining accuracy. For example, when analyzing the output of L2 English learners, which version of English is considered the “norm” – e.g., British, American, Indian, or Singaporean? Furthermore, which dialect within each of those versions is the “norm”? Researchers have tried to
resolve the issue by opting for acceptability over grammaticality. Ellis and Barkhuizen (2005) explain that current researchers base their error analyses on subjective evaluation that “often involves making stylistic rather than grammatical judgments” (p. 56). However, this method is not without flaws in that native-speakers, including researchers, often disagree on what is acceptable (Davies, 2003; 2011). Though most researchers acknowledge the flaws, CDST, task-based learning, and SLD studies including grammar analyses tend to define grammatical norms by using the researcher’s native-speaker intuition and a few additional raters (e.g., Ellis & Yuan, 2004; Polat & Kim, 2014; Tavakoli & Foster, 2011).

To define errors, Ellis and Barkhuizen (2005) suggest reconstructing erroneous sentences as closely as possible to the original sentences, then identifying where changes are necessary to correct the sentence. Each change counts as an error. This can be problematic in that the speakers’ intentions are not always clear. For example, in the case of English verb tense errors, the difference between an omitted “ed” indicating past tense and an omitted “s” on the third person singular depends on the context. In these cases, Ellis and Barkhuizen (2005) recommend asking the learner about their intention for an “authoritative reconstruction” (p. 59). However, since that is not always possible, researchers may have to rely on their own perception of the speaker’s intentions based on the situational context. As another option, a second native-speaking rater can contribute to identifying errors and offering corrections based on context as was done by Larsen-Freeman (2006), Mulder and Hulstijn (2011), and Polat and Kim (2014).
In contrast to traditional SLA research, which tends to focus on the development of specific grammatical forms (prepositions, past tense, negation, etc.), CDST and task-based learning researchers focus on global variables (frequency of all errors), the grammatical development of more than one construct (accuracy and complexity), or comparing the development of global and specific variables.

Task-based research analyzes the effects of tasks differences on the CAF variables. As part of that research, Skehan and Foster (1999) compared accuracy to complexity and fluency to show that task differences can cause cognitive strain which affects the constructs differently. In their study, 47 intermediate ESL students with various L1s added narration to videos in four task conditions. The conditions included: a) watch and tell, b) read a summary, then watch and tell, c) preview the video, then watch and tell, and d) watch then tell. They found that students were most accurate in conditions b and d, which allowed for pre-planning. In comparison with other constructs, condition d produced the least fluent and most complex narratives. They argued that global measures of accuracy are “more sensitive to detecting differences between experimental conditions” (p. 107), and therefore are more appropriate for task-based analyses.

Spoelman and Verspoor (2010) combined both specific error analysis and construct comparison to more closely analyze the relationship between accuracy and complexity as they developed over three years. They coded for errors related to the Finnish grammatical case system in an L1 Dutch L2 Finnish learner’s essays. They found that as the learner studied, variation in the amount of errors between essays decreased
indicating stabilization, but not necessarily improvement. Next, they reanalyzed the data for 11 specific cases and found that three cases (accusative singular, partitive plural, partitive singular) had considerably more errors than the other eight. They also analyzed sentence and word complexity finding variation over time. However, “no meaningful relationship” (p. 533) was found between accuracy and complexity in that the correlation between them was low. They concluded both variables are characterized by peaks and regressions in their development, but that the development of accuracy and complexity were not necessarily related.

Polat & Kim’s (2014) longitudinal case study also coded for two global variables (errors per 100 words and elaboration) and two specific variables (present simple errors and subordination). Present simple errors were chosen as the specific variable because it was “appropriate to the interview setting and to Alex’s [the participant] linguistic proficiency” (p. 193). The number of errors per 100 words was chosen to analyze general change over time. In contrast to other CDST studies which code for error-free units, they chose errors per 100 words since their participant was regularly making more than one error per unit. The results showed high variability and no clear patterns of development in either accuracy measure. Based on Larsen-Freeman and Cameron (2008), Polat and Kim (2014) conclude that variability in accuracy may be his current language state. While complexity showed improvement in lexical diversity and elaboration, subordination remained unchanged. They theorized that their participant improved in the areas necessary to communicate (lexical diversity) rather than grammatical correctness because
he was a naturalistic learner who was not receiving any type of explicit grammar instruction.

As illustrated in this review, current CDST and task-based learning research tends to include accuracy as a global variable to be compared against other variables. Specific variables may be included and differing units of measurement may vary based on the level of the participant and purpose of the investigation. Despite efforts to create variables that are valid and reliable, a few flaws merit discussion starting with reliance on native-speaker intuition and the use of global variables as a measure of development.

A problem with coding errors by native-speaker intuition is the fact that it may not reflect the speech to which the learner is exposed. Though the use of additional raters reduces this risk, the raters are likely to be as highly educated as the researcher. Consequently, their L1 intuition may result in unrealistically high standards for language norms. As evidenced by Mulder and Hulstijn’s (2011) study of 98 L1 speakers of Dutch, significant differences between L1 speakers with different socio-economic backgrounds, education, and peer groups can affect their CAF scores. While Mulder and Hulstijn (2011) coded with native-speaker intuition, they only included “obvious violations of grammar, ungrammatical even in spoken language” (p. 483). Specifically, they selected 11 violations of Dutch grammar to code including incorrect verb, word order, noun or adjective inflection, use of determiners or conjunctions, incorrect comparative constructions, omission of a word or constituent, and superfluous use of a word or constituents. However, they allowed for omissions of the subject and finite verb which is
common in spoken Dutch. They found that participants with higher education and/or professional status produced significantly fewer grammatical errors. Moreover, they found that “a large majority of the participants made serious grammatical errors, such as violations of subject-verb agreement in spontaneous speech” (p. 486). In addition to accuracy, elaboration, and lexical diversity were also affected by socio-economic backgrounds, education, and peer groups. They concluded that researchers studying the CAF variables should take native-speaker variability into consideration.

In order to make that consideration, researchers throughout SLD often incorporate comparison groups. By using a comparison group, researchers may be able to classify some errors as acceptable due to the L1 speaker’s usage and eliminate them from the learners’ analysis. Moreover, comparison groups from the learners’ speech community may also be useful to examine non-standard grammar features present in the speech community. Once identified, the learner’s development of those features can be analyzed as a form of development (Howard, 2005; Kinginger, 2008; Regan, 2004).

In addition, task-based research has also shown that both native and non-native speakers are affected by cognitive processing speed and strain. Therefore, the context of the error may need to be considered. Rather than labeling specific errors as acceptable, it may make more sense to compare the average number of errors a native speaker is making to the average number of errors a non-native speaker in a task where cognitive strain may be particularly high. Such analyses may also add insight to which errors are truly mistakes unrelated to the speakers’ knowledge of the language.
Lastly, errors in more complex forms often indicate development, which is difficult to see with a global measure. As Skehan and Foster (1999) explained, decreased accuracy may indicate that the learner is taking risks and experimenting with new forms. Additionally, Pallotti (2009) argues that the ability to communicate should be considered as well. He explains that “one can have perfectly accurate but communicatively inadequate messages (colorless green ideas ...) or perfectly intelligible messages violating various L2 norms (me no likes go dance)” (p. 590). Some of these issues are addressed with specific accuracy measures and complexity measures (elaboration and lexical diversity), as well as micro-analyses and qualitative analyses, such as the ones done by Larsen-Freeman (2006) and Polat and Kim (2014).

The present study aims to contribute to the research on accuracy by comparing a global variable (errors per unit) and two specific variables (errors related to the past tense and omission of obligatory auxiliary verbs) and by comparing the main participant to a group of native-speakers.

2.1.3 Fluency

*Fluency* refers to the ease or difficulty a speaker has producing language (Ellis & Barkhuizen, 2005; Skehan, 2003). Unlike complexity and accuracy, fluency reflects how quickly a speaker can access language knowledge rather than just possessing knowledge of the language. As Housen and Kuiken (2009) explain, “fluency is primarily related to learners’ control over their linguistic L2 knowledge” whereas accuracy and complexity
“are primarily linked to the current state of the learner’s (partly declarative, explicit and partly procedural, implicit) interlanguage knowledge” (p. 462). Thus, slower and more disfluent speech may occur when L1 or L2 speakers are struggling to plan a phrase, retrieve a word, or produce the correct morphology, syntax, and/or pronunciation. Accordingly, studies have found that memorized exemplars (lexical chunks) and automatized restructured morphology and syntax result in increased fluency (Skehan, 1996; Towell et al, 1996, Wolfe-Quintero et al., 1998).

During the 1990s and early 2000s, researchers developed three measurement constructs for fluency variables: speed, breakdown, and repair (Kormos & Dénes, 2004; Skehan, 2003; Skehan & Foster, 1999) Speed fluency refers to how quickly the speaker produces speech. Speed measures generally include words (or syllables) per second (or minute) and length of silent pauses. The second construct, breakdown fluency refers to number of filled pauses (uh, um, err, etc.) or silent pauses. Lastly, repair fluency is measured by the frequency of disfluencies including false starts, reformulations (self-corrections), and repetitions (repeating sounds, words, or phrases). Skehan (2003) argues that the separate constructs are necessary to give a comprehensive picture of fluency performance. Literature prior to 1999 and some current literature combine repair and breakdown fluencies into the category of hesitations. Despite progress in research, fluency remains a controversial area due to the lack of consistent definitions, consistent standards, and mixed results in the research.
Although L1 speakers are considered “fluent by default” (Bosker, Pinget, Quené, Sanders & de Jong, 2014b, p. 580), the amount of variation among L1 speakers causes difficulties determining what speed is acceptable, how often a speaker should pause, and what acceptable amounts of repairs are. Consequently, traditional research techniques, including L1 benchmarks, are problematic for fluency research. For example, Mulder and Hulstijn (2011) found that native-speakers’ hesitations (disfluencies) varied from 0 – 2.15 hesitations per t-unit and were not correlated with education, performance, or communicative adequacy. Thus, even within a carefully selected comparison group, fluency benchmarks are difficult to obtain. Moreover, Bosker et al., (2014b) found that even after controlling for length of pauses and rate of speed, non-native speakers were still rated less fluent than native-speakers; and Towell et al. (1996) found a correlation between L1 and L2 speech rates, which indicates that learners will transfer their L1 speech rates into their L2. As such, it appears that there are elements of fluency that are difficult to explain with quantitative comparisons between native and non-native speakers. Towell et al. (1996) summarize some of the possible factors that could affect both native and non-native speakers in reference to the temporal variables in their study:

Change in such measurements may be caused by the demands of a particular task, it may be a characteristic of an individual, it may be a sign of individuals knowing how to verbalize what is already in the mind, it may be to do with establishing the correct ‘balance’ between length of utterance and linguistic structure. (p. 93)

As an alternative to native-speaker comparisons, many fluency studies rely on native-speakers as judges of fluency rather than exemplars of it. Derwing, Munro, and
Thomson (2007) had 33 L1 English speakers participate as fluency raters in a study of comprehensibility and general fluency. They listened to samples of 16 Slavic learners (L1 Russian or Ukrainian) and 16 L1 Chinese learners of L2 English from three data collection points over the period of two years. The post-test results showed that the Slavic participants, who were more integrated into the community, became more understandable and generally more fluent by the native-speaker judges than the Chinese participants. Since self-reported native-speaker interaction and responses during interviews indicated higher levels of native-speaker interaction among the Slavic speakers, Derwing et al. (2007) theorized that increased practice with native-speakers among Slavic speakers most likely led to their improvement. Other variables such as native-speaker bias against Chinese accents or cross-linguistic differences affecting learning were given less weight. Derwing et al. (2007) explained “While some previous studies have revealed biases against second language users’ speech … bias cannot explain the differential effects seen here, since the groups were rated the same at T2 on both measures” (p. 373). However, cross-linguistics and cross-cultural differences may have affected the learners’ ability to advance. For example, the Slavic speakers did not need to overcome the l/r pronunciation difference which creates difficulty for Chinese speakers (Swan & Smith, 2001).

While developing quantifiable fluency variables, scholars have also relied on the holistic judgments of expert judges, such as instructors, to add validity to their quantitative results. The results have been mixed, indicating that some variables are more noticeable to judges than others. For example, Kormos and Dénes (2004) had three L1
English and three L2 English speaking instructors’ rate 16 L1 Hungarian L2 English learners orally adding narration to a cartoon strip. The judges’ holistic measures (flow of speech, lack of pauses, self-corrections, speed of delivery, and efficiency in handling breakdowns) were compared to the quantifiable measures of speed (speech rate, articulation rate, phonation time ratio, mean length of run, and length of pauses), breakdowns (silent and filled pauses per minute), and repairs (number of disfluencies per minute). They found that the speed variables correlated with the judges’ scores, whereas breakdowns and repairs did not. Additionally, they found correlations between the judges’ holistic scores with accuracy (error-free clauses per total clauses) and with lexical diversity ($D$). They concluded that “fluent performance entails the application of efficient and accurate processing mechanisms” (p. 161).

In a similar study, Bosker, Pinget, Quené, Sanders, and de Jong (2012) used multiple linear regressions to find the quantitative fluency variables most likely to predict the holistic fluency scores of 80 L1 Dutch judges. The participants, 15 L1 English and 15 L1 Turkish learners of Dutch, completed eight speaking tasks. The transcripts were analyzed quantitatively by the researchers for speed, breakdown, and repair disfluencies. Then, the L1 Dutch judges listened to the samples and rated them qualitatively on overall fluency, which was defined by use of pauses, speed of delivery, hesitations, and corrections. Analyses showed that breakdown and speed disfluency were statistical predictors, whereas repair disfluencies were not. However, in additional analysis where judges were asked to direct their attention to a specific fluency construct, all three
constructs showed some predictive power in order as follows: pauses, speed, and repair. Bosker et al. (2012) suggest the repair disfluency may need to be further distinguished to produce more predictive power. For instance, repetitions may produce more predictive power than self-corrections. They concluded that all three sub-dimensions (including repair) “play a role in fluency perceptions and none should be disregarded” (p. 171).

Despite issues with the variables, SLD studies have shown improvement in fluency over time (e.g., Ferrari, 2012; Kinginger, 2008; Segalowitz & Freed, 2004). For example, Ferrari’s (2012) study of six L2 Italian learners with various L1s studying abroad found linear decreases in length of pauses (speed fluency) and hesitations (number of repairs and pauses combined) over four data collection points during three years.

Towell et al. (1996) looked at 12 intermediate L1 English L2 French students before and after a year of study abroad. They hypothesized that the average length of pauses between utterances would increase allowing planning time for the conceptualization of ideas. In contrast, the phonation time rate (time speaking/total time) would decrease indicating less frequent pauses to perform lexical and syntactic searches during speech performance. Their participants retold a film in French during their 2nd year of study at home and upon return from six-months study abroad during their third year of study. They found significant increases in speaking rate (syllables per minute), articulation rate (syllables/minute minus pauses), and mean length of run (syllables between pauses) between the first and second narrations. However, there were no significant differences between phonation time rate or average length of pause as
expected. A qualitative analysis showed that rather than increasing pauses between utterances to think, learners began using fillers, modifiers, collocations, and lexical and syntactic phrases which allowed them time to think and improved their speaking rate, yet did not significantly affect the length or frequency of their pauses. They concluded that these proceduralized phrases, rather than changes in pauses, increased their fluency scores.

If fluency is related to automaticity of language production as Towell et al. (1996) suggests, one might expect a cyclical pattern related to learners gaining and automatizing explicit knowledge as Verspoor, Lowie & van Dijk (2008) theorized with accuracy. Task-based research, which compares fluency, accuracy, and complexity, offers some insights into how cognitive strain, which may be related to automaticity, affects all three variables. The next section discusses such research.

2.2 Language Processing Models and Attentional Resources

During the 1990s, the interest in communicative tasks for the L2 classroom increased, and researchers began to explore the effects of different tasks on CAF variables (Skehan, 1996; Skehan & Foster, 1999). Skehan (1996) suggested that different tasks may ease or increase strain on attentional constraints related to processing. His theories are largely based on Levelt’s (1989) model of speaking. This section discusses Levelt’s (1989) original monolingual model of speaking, suggestions by de Bot (1992) to adapt the model for bilingual processing, and Levelt’s (1999) revised model.
Additionally, I will briefly review research in task-based learning as it relates to the CAF variables competing for attentional resources.

Levelt’s (1989) model of speaking describes the process by which speech moves from conceptualization to production and predicts points in the process that are more susceptible to strains on speakers’ cognitive resources. Levelt (1989) theorized that speech begins in the conceptualizer where preverbal ideas are gathered and context is considered. The formulator then retrieves lemmas from the speakers’ long-term memory. Next, the necessary morphology and syntax are added to the lemmas in order to relay the conceptual message. That message is transformed into a phonetic plan, which is passed to the articulator. The articulator executes the instructions, resulting in oral production. Lastly, the speech comprehension system allows the speakers to self-monitor their inner and overt speech which may result in corrections and disfluencies. The processes, in Levelt’s view, run parallel to each other and rely on working memory to allow buffering between them when needed.

Levelt (1989) also compares executive or central control to automatized speech. Processes that are under central control are automatized and do not require conscious control. For example, syntax, morphology, or pronunciation in speakers’ L1 are generally automatized. Processes under executive control are intentional and demand the speakers’ attention. Executively controlled processes, such as such as meaning construction and accessing newly learned vocabulary, generally occur in the conceptualizer. Executively controlled processes draw on speakers’ limited cognitive capacities and may result in
cognitive strain. Thus, strain on the conceptualizer is more likely to produce disfluencies than the other areas. Additionally, disfluencies may occur when speakers are self-monitoring their speech, which can occur at any point in the process.

Levelt’s (1989) model advances our understanding of L1 processing but did little for understanding the simultaneous processing of two or more languages. De Bot (1992) suggested two bilingual adaptations to Levelt’s (1989) model. His first modification suggests separating the conceptualizer into a non-verbal macroplanning level and a language specific microplanning level. He theorized that the first part of the preverbal message (macro-planning) is not language-specific in that it is gathering knowledge (concepts, context, deixis, etc.) rather than linguistic data. He then proposed that bilinguals evaluate language choice in the same manner as they consider register and context. After this decision has been made, they begin the microplanning stage where the appropriate lexicon is selected. He also proposed that for bilinguals, the mental lexicon contains vocabulary from all the languages available to the speaker. Depending on the level of activation or dormancy of the language, lexical retrieval may be more difficult. If the appropriate word to express the concept is dormant, unknown, or nonexistent in the L2 (but accessible in the L1), speakers may experience problems moving to the formulator stage where the sentences are formulated.

Levelt (1999) made adjustments to the model based on de Bot’s (1992) suggestions by adding a surface structure level between the conceptual area and the formulator/articulator area (see Figure 1). The *surface structure* gathers the lemmas and
places them in the proper order. Once the lemmas have been placed in the surface structure, the morphosyntactic coding is activated. The morphosyntactic and phonological information is encoded onto the surface structure in the *phonological/phonetic system* (see Figure 1). In Levelt’s (1999) model the formulator and articulator are combined.

Some recent studies challenge the idea of a language-neutral conceptualizer and suggest that crosslinguistic differences, particularly in the area of motion verbs, may affect the details that speakers include in their speech (Bylund, Abrahamsson, & Hyltenstam, 2009; Pavlenko, 2014; Slobin, 1996; 2006). Therefore, the pre-verbal concepts would be affected as well as the lexical choices. In light of this research, Levelt’s (1999) model may need another update.

De Bot’s (1992) second suggestion posits that morphosyntactic coding is language specific. Once the word (and therefore the language) has been selected, a set of language specific morphological information is selected and available to the speaker. However, since second language learners’ grammar may be incomplete, they may not be able to encode the sentence correctly. Additionally, de Bot (1992) suggests that declarative knowledge (explicit) rather than proceduralized knowledge (implicit) may slow the process or create disfluencies.
Figure 1. Levelt’s 1999 Model of Speaking

Reproduced with the author’s permission (see Appendix G)
Lastly, the phonetic information is encoded. Levelt (1989) theorized that phonemes are coded into syllables which are stored in speakers’ memory. In this case, de Bot (1992) argued that storing two sets of syllable inventory would be redundant since syllable combination in most languages overlap at some level. Therefore, he proposed that this information is stored in one unit, which Levelt (1999) referred to as the syllabary (see Figure 1). Once the syllables are coded, the articulatory information is activated to produce the language (for a more detailed explanation of how the phonemes are encoded into the syllabary refer to Levelt (1999)). All three models stress that while these processes usually occur chronologically, they are also happening simultaneously. For example, while the speaker may be adding morphological endings to a verb, they may also be macroplanning for their next utterance. Based on the model in Figure 1, Segalowitz (2010) suggested six critical points where difficulties in underlying processing could be associated with L2 speech dysfluencies. These include: a) microplanning, b) grammatical encoding, c) selecting lemmas, d) morpho-phonological encoding, e) phonetic encoding, and f) articulation.

Skehan (1996) applied Levelt’s (1989) model to the CAF variables. He suggested that complexity, accuracy, and fluency compete for the learners’ limited attentional resources while moving through the speech production process. In later work, Skehan (2009) and others refer to this phenomenon as the Trade-off Hypothesis (e.g. Ellis, 2009; Tavakoli & Foster, 2011). This hypothesis predicts that when “committing attention to one area, other things being equal, might cause lower performance in others” (Skehan,
Thus, when concentrating on creating correct grammar forms (accuracy), a learner may pause more in order to think (fluency). VanPatten (1990) arrived at similar conclusions regarding competition between conscious attention to form and conscious attention to meaning. He found that L1 English L2 Spanish learners’ recall of ideas reduced significantly when asked to focus on bound morphological forms (verb suffixes) over unbound lexical forms that contributed to meaning. These findings suggest that meaning and form were competing for attention.

Skehan (1996) also noted a difference in learners’ choices of foci that may result in tension between complexity and accuracy. While a conservative learner may choose to focus on grammatical structure rather than complexity, a risk-taking student may try to incorporate new forms and thus focus on complexity rather than accuracy. Conversely, Robinson (2011) has argued that the tension between complexity and accuracy do not necessarily result in competition. He theorized that carefully constructed tasks should result in positively correlated increases in complexity and accuracy.

2.2.1 Task-Based Learning

Task-based learning research examines attentional resources by manipulating tasks to create strain during different phases in Levelt’s (1989) speech model. Skehan (1996) proposed that CAF outcomes would be affected by task differences related to pre-planning, the cognitive complexity of the task, and time allowed to complete the task.
Table 1 (below) briefly summarizes the theorized effects of focusing on one area over the others.

### Table 1
**Summary of the Trade-off Hypothesis**

<table>
<thead>
<tr>
<th>Prioritization</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Accuracy may be increased by familiar tasks and tasks with concrete structures in L2 production. Focus on accuracy may encourage avoidance of new structures and reduce fluency.</td>
</tr>
<tr>
<td>Complexity</td>
<td>Complexity may be increased by tasks that involve interaction, pre-planning, or simultaneous events. Focus on complexity may increase the range of structures but compromise accuracy and fluency.</td>
</tr>
<tr>
<td>Fluency</td>
<td>Fluency may be increased by communicative tasks which encourage the learner to focus on meaning. Focus on fluency may decrease development of new structures and result in over-lexicalized performance.</td>
</tr>
</tbody>
</table>

*Note: Adapted from Skehan (2014) and Skehan and Foster (1999)*

Task-based learning researchers manipulate the task to produce different amounts of cognitive strain. They do so by allowing planning time (Ellis & Yuan, 2004; Skehan & Foster, 1999; Skehan & Shum, 2014), or changing the complexity or structure of the task itself (Foster & Tavakoli, 2009; Skehan & Foster, 1999; Tavakoli & Foster, 2011).

Researchers focusing on planning, experiment with two types: pre-planning and *on-line planning*, which refers to planning while performing the task (Ellis & Yuan, 2004; Wang, 2014). Pre-planning activities may teach skills necessary to complete a task, activate schemas, or activate salient vocabulary and structures that will be incorporated
throughout the task (Skehan, 1996). Depending on the task, pre-planning may ease the strain on *microplanning* (conceptualizing the ideas), or *macroplanning* (adding the lexicon). For example, cognitively complex tasks, such as describing multiple events or predicting outcomes, can be improved by clear instructions on what will be described. However, pre-planning has not been found to improve accuracy. Yuan & Ellis (2003) found that in 42 L1 Chinese L2 English learners, accuracy did not significantly improve with pre-planning, but did improve with additional on-line planning. They theorized that pre-task planning promoted formulations, aiding with vocabulary and ordering the events, but unpressured on-line planning provided additional and improved monitoring, leading to improved grammar.

Common pre-planning variables include allowing planning time before writing or speaking (Ellis & Yuan, 2004), allowing participants to preview the task before completing it, or giving participants oral plot summaries before asking them to add narration to it (Skehan & Foster, 1999; Skehan & Shum, 2014). All of these studies found that pre-planning generally increased CAF scores. The most significant results were found between learners allowed to watch a video then summarize the story compared to learners with no pre-planning describing the video while watching it (Wang, 2014).

Another focus of recent research is on-line planning versus pre-planning. On-line planning is generally tested by allowing the learners additional time to complete the task. Wang (2014) compared participants adding narration to a video at full speed versus allowing the learner to pause the video. Yuan and Ellis (2003) compared oral narrations
between participants with a time-limit versus participants with unlimited time to complete the task. Both Yuan and Ellis (2003) and Wang (2014) found significant improvements in accuracy and complexity in conditions that allowed online planning rather than pre-planning or no planning.

Other researchers have experimented with the task itself (Skehan & Foster, 1999; Skehan & Shum, 2014; Tavakoli & Foster, 2011). Skehan (1996) theorized that if the task is too difficult, an excessive amount of resources may be devoted to communicating meaning. This difficulty, in turn, may result in increased fluency with decreased accuracy. If the task is too easy, the learner may not be challenged in terms of restructuring or developing automaticity since the resources necessary to complete the task already exist in learners’ long term memory. The structure or characteristics of the task have been researched in two areas, storyline complexity and inherent structure. Both have been found to affect the areas of complexity and accuracy (Foster & Tavakoli, 2009; Tavakoli & Foster, 2011). Storyline complexity refers to the inclusion of background events, which theoretically increases the use of subordination (complexity). Inherent structure refers to the ease of connecting one event to the next. For example, if a story has an obvious ending, it is easier for the learner to add narration to it since they are not straining the conceptualizer to create connections between events. Theoretically, increased inherent structure leads to higher accuracy scores due to the decreased cognitive load on the conceptualizer.
Foster and Tavakoli (2009) and Tavakoli and Foster (2011) found significant differences between tasks with differing storyline complexity and inherent structure. Tavakoli and Foster (2011) elicited narratives from four cartoons, each containing a different combination of inherent structure and background events (2 x 2 factorial design). They collected narratives from 60 L2 English learners living in Tehran and 40 L2 learners living in London. Complexity was measured with mean length of utterances (words between pauses) and clauses per AS-unit. An AS-unit, or analysis of speech unit, is comprised of an independent clause or subclausal unit plus all subordinate clauses associated with it. Accuracy was measured by the percentage of error-free clauses. The narratives with background events showed significantly increased complexity, regardless of their inherent structure. While London learners’ narratives were more complex than the Tehran group, the differences were not significant. In terms of accuracy, narratives with inherent structure had significantly more accurate results regardless of storyline complexity or learning context. Thus, as expected, inherent structure increased accuracy, and storyline complexity increased complexity. In addition, Foster and Tavakoli (2009) gave the same tasks to 100 L1 English speakers and found that their complexity scores were significantly higher with the inclusion of background events as well. However, inherent structure did not have a significant effect on accuracy as it did with L2 learners.

This overview shows that task-based learning research still has areas in need of further research. To begin, narrative research focused on storyline complexity and inherent structure has explored complexity and accuracy, yet has not explored fluency.
Furthermore, most task-based learning research focuses on one task and slight variations of it. Most commonly, the researchers chose one of the following types of oral stimuli to gather CAF data: story books (Pang & Skehan, 2014), brief videos (Ferrari, 2012; Skehan & Foster, 1999), or cartoon strips (Ellis & Yuan, 2004; Foster & Tavakoli, 2009). Since many stimuli are used in research and pedagogy, the field would benefit from research that compares performance across different tasks. The present study contributes to filling this gap by comparing CAF scores from four different tasks (conversation, picture descriptions, narratives from picture books, and narratives from videos).

Since this study is longitudinal and investigates language learning as a complex system, quantitative proficiency measures may not show the complete picture. In the following two sections, two areas related to Juan’s environment that were shown to affect his learning trajectories are reviewed: naturalistic language learning and motivation to learn an L2.

2.3 Naturalistic L2 Learning Context

Even among native speakers, language use varies due to individual differences such as age, education, or dialect (Mulder & Hulstijn, 2011). Native speakers may also use speech as a marker of identity and adjust it in order to conform to the patterns of their peers, or those they want to be their peers (Lippi-Green, 2012). Therefore, it makes sense that L2 speakers would also adapt their speech to conform to the patterns of their L2 community if they want to integrate into that community. As such, Regan (2004) argued
that for learners wishing to integrate into the L2 community, knowledge of native-speaker variability, and including less prestigious forms of the language ought to be considered part of L2 speaker competence. However, classroom instruction is limited in this area. As Kinginger (2008) explains, “classroom language is by definition a representation of language use, where the forms and functions under scrutiny are always preselected according to rules of decorum and pedagogical considerations of relative usefulness” (p. 29). In contrast, immersed learners have the opportunity to adapt to community norms. Much of the recent research in this area is related to study abroad programs which are growing rapidly. This section discusses study abroad research in relation to academic achievement, proceduralizing previously acquired explicit instruction, and length of stay.

While the study abroad context and immigrant immersion have a great deal in common, study abroad students have different motivations, an academic environment, and a pre-determined amount of time. Over the last few decades, participation in study abroad programs has increased worldwide (Collentine & Freed, 2004; DeKeyser, 2014; Kinginger, 2008). As such, interest in the topic has also increased. In 2014, 886,052 international students were attending colleges in the United States, an increase of 72% from the year 2000 (Institute of International Education, 2015). Additionally, Europe’s multilingual education policies have encouraged growth in the long-standing exchange programs, such as Erasmus, Socrates, Comenius, and Leonardo (Pérez-Vidal, 2011).
Due to pressure on study abroad programs to prove their effectiveness, early research was dedicated to providing evidence that study abroad programs were beneficial. Studies in the 1990s showed that when study abroad participants were compared to students receiving formal instruction in their home classrooms, the study abroad students showed significant advances in fluency, communicative strategies, and repertoires of style (DeKeyser, 2014; Kinginger, 2009; Towell et al., 1996). For example, Towell et al. (1996) found that L1 English L2 French learners in the study abroad context increased their knowledge of formulaic expressions, syntactic strings, organizers, and fillers. They theorized that enabling these structures allowed learners to increase their amount of speech and speed of production resulting in increased fluency scores.

In the area of accuracy, study abroad students do not necessarily show significant gains in syntax or morphology over students studying in their home country’s classroom (DeKeyser, 2014; Freed, 1995). However, certain linguistic forms appear to develop more in a study abroad context. Howard (2005) found that six L1 English L2 French students in a study abroad program in France produced more sophisticated verb tenses than six students studying at home at the same level and six more advanced students who had not studied abroad. In oral interviews, the learners who studied abroad produced the past tense, including the perfect and imperfect distinction, significantly more accurately and more frequently than those who did not study abroad.

The participants in Howard’s (2005) longitudinal study were advanced learners of L2 French who had studied the difference between past and past imperfect. Therefore,
they previously had explicit instruction on how to form these tenses, but they needed to
learn to use them in context and fluently. Through repeated practice and increased L2
exposure, they began to automatize their explicit L2 knowledge, leading to more implicit
knowledge, higher fluency, and contextual accuracy. Accordingly, most study abroad
scholars agree that the intermediate level is ideal for study abroad participation, in that
learners should have gained enough explicit language knowledge to be ready to
automatize it (Collentine & Freed, 2004; Pérez-Vidal, 2014; Towell et al., 1996). As
Perez-Vidal (2011) explains, study abroad “is beneficial if learners have reached an upper
intermediate functional level of language competence at which they no longer need to
learn procedures but rather make their language more automatic” (p. 13).

Regan (2004) longitudinally studied five advanced L1 English L2 French students
for three years, including a year during which they studied abroad in a Francophone
country. Regan (2004) focused on a mildly stigmatized variation from formal French, the
omission of the first negation particle *ne* when forming the negative in French. Thus, *je
ne* parle *pas* français would be *je parle pas* français. She found that the overall rate of *ne*
deletion increased dramatically and approached native-like norms in all but one student
after a year abroad. She also found that participants followed the same constraints on
deleting *ne* as native speakers such as the levels of formality, phonological constraints,
and the syntactical nature of the verb. Furthermore, in the speech samples taken a year
after the participants returned, the *ne* variation remained stable or decreased only slightly
even though they were no longer exposed to the L2 speech community. The students in Howard (2005) also maintained their native-like speech habits even after returning home.

In another study of L1 English L2 French students, Kinginger (2008) explored how 24 American learners of L2 French developed three aspects of sociolinguistic competence: acquisition, usage, and awareness of colloquial words and phrases; appropriate address forms; and the ability to perform speech acts for leave-taking situations. Students were formally tested for their proficiency at the beginning and at the end of their year abroad with a variety of tests, including a standardized grammar exam (Test de Français International), language assessment interviews, and tasks involving role play. Though development of the forms indicated individual differences, she found that on average learners improved in all of the sociolinguistic areas and that all of the students except one showed significant general proficiency gains. While the purpose of this study was to investigate L2 learning during study abroad, many of Kinginger’s (2008) conclusions related to improving the quality and frequency of learners’ interactions with native speakers. In the case of one unsuccessful learner, she suggested that factors limiting the frequency of exposure to and interaction with native-speakers may have affected her development. Due to the distance between her housing and the city, the learner had to leave the city early and could not interact with locals in the evenings. Additionally, she rarely spoke with her host family.

While these studies have shown that immersion has an effect on language acquisition, it is still difficult to narrow down the specific factors regarding the
interaction with native speakers, or even identify best methods for their study. Kinginger (2008) gained some of the previously mentioned insight about the learner’s housing through a diary kept by the participant. Other researchers have relied on surveys or interviews regarding L2 exposure. For example, Valls-Ferrer and Mora (2014) found two noteworthy correlations related to effects of L2 exposure: a) self-reported interaction and perceived improvement and b) watching L2 TV and fluency scores. However, since most participants interacted with native speakers and watched L2 TV, the results cannot be separated into two causative factors. Ranta and Meckelborg (2013) separated some factors using daily activity logs. In a longitudinal study of international teaching assistants, they found that 17 Chinese students appeared to have an inverse relationship between native-speaker interaction and other activities in English. Students that engaged in conversation and academic speaking, read and watched TV less and vice versa. However, this study did not compare interaction with proficiency still leaving the effect of the interaction unanswered.

Whether students benefit from explicit instruction while studying abroad remains a subject of debate. If students are studying academically, they are most likely receiving corrections to their writing and possibly their speech. Thus, the variables are difficult to separate. Collentine and Freed (2004) point out that study abroad learners reported consciously attempting to use their new explicit knowledge outside of the classroom. In contrast, the learners in Miller and Ginsberg (1995) did not value the classroom instruction and in some cases rejected it.
Though the studies mentioned in this section show considerable changes in learners’ L2 after immersion, all of them mention variation within the sample. Due to the autonomous nature of study abroad programs, students can choose whether or not to take advantage of the opportunities in their environment. As found with many SLD and motivation studies (Gardner, 1968; Schumann, 1978; Shoaib & Dörnyei, 2005), motivation, purposes for learning, outgoingness, and the learners’ willingness to interact with speakers of the target language affect their development (Howard, 2005; Kinginger, 2008; Segalowitz & Freed, 2004). Consequently, scholars are calling for further investigation into the effects of motivation on study abroad students. DeKeyser (2014) called for documentation of “how individual students are motivated for acquiring advanced language proficiency, how this motivation increases or decreases during their stay abroad, and how it is both a cause and consequence of their level of language proficiency” (p. 318). He additionally mentions the need for more studies utilizing the CAF variables combined with a qualitative analysis. Collentine & Freed (2004) concurred with his suggestion by calling for more studies including linguistic features and the effects of social conditions. This study contributes to research in this area by comparing the participant’s CAF scores to his self-reported motivation and the effects of L2 exposure and interaction. The next section covers motivational factors relevant to this study.
2.4 Language Learning Motivation

When studying a language, motivation can be a powerful influence on a learner. As Dörnyei (2005) explains, motivation “provides the primary impetus to initiate L2 learning and later the driving force to sustain the long and often tedious learning process” (p 65). However, researching motivation presents a variety of problems in that motivation is difficult to quantify and remains in a constant state of flux. Consequently, motivation research generally relies on self-reported measures, such as questionnaires and interviews, conducted in a retrospective fashion. In recent research, scholars studying motivation have begun to explore fluctuations in learners’ levels of motivation in comparison to their language development with a CDST framework (Dörnyei, 2009). The following section briefly summarizes the findings of motivation research that involves naturalistic contexts. I will not discuss research on motivation in the FL/L2 classrooms, as it is outside the scope of this study.

Integrating into an L2 community as a motivator first appeared as a research topic during the 1960s and 1970s. During that period, Gardner (1968) began his seminal work with Canadian children learning English and French and introduced the concept of integrative motivation. Gardner theorized that the desire to integrate into a community is a primary motivation for learning a language. Gardner and Smythe (1981) expanded the scope and specificity of those motivational factors which they developed into the Attitude Motivation Test Battery (AMTB), a quantitative standardized questionnaire still used in current research (Dörnyei, 2005; Masgoret & Gardner, 2003). The AMTB categories
include attitudes toward the learning situation, integrativeness, motivation, integrative orientation, and instrumental orientation. Masgoret and Gardner (2003) define *integrative orientation* as identification with the community and *instrumental orientation* as practical reasons for learning a language. They emphasize that a particular orientation does not necessarily imply motivation. One may have a reason to learn a language, but lack the motivation to act on that reason.

Masgoret and Gardner (2003) conducted a meta-analysis combining the results from 75 independent samples or 10,489 individuals to explore the AMTB factors effects on achievement. They found that achievement (grades, self-ratings, and objective measures) was positively correlated with reasons to identify with the community, an openness to integrate, and motivation (defined as goal directed behavior). Of those three, motivation was the most strongly correlated with achievement.

While integrative and instrumental motivation differentiate between the reasons for learning the L2, Dörnyei (2005, 2009) added the distinction between internal and external motivation, referred to as the L2 Motivational Self System. Within this framework, the *ideal L2 self* refers to the L2 attributes learners wish to possess in the future. The theory assumes that becoming an L2 speaker is positive, and thus characteristics of the ideal L2 self tend to be positive, such as successfully integrating into an L2 community’s culture, achieving individual proficiency goals, or doing well in an L2 class. The *ought-to self* refers to pressure from external sources to learn a language such as encouragement or discouragement by significant others, standards imposed by
peers in the L2 community, or academic standards (Dörnyei, 2005; 2009; Dörnyei & Chan, 2013). Csizér and Lukács (2010) found that the ideal L2 self and language learner attitudes had the highest mean values of the 17 motivational variables they tested in 623 L1 Hungarian L2 English learners (202 secondary students, 230 university students, and 64 adults). In contrast, factors relating to the ought-to self were inconclusive.

While categories of motivation are useful in investigating the characteristics of motivation, studies incorporating them tend to neglect the changing nature of motivation. As Dörnyei (2005) explains:

> when motivation is examined in its relationship to specific learner behavior and classroom processes, there is a need to adapt a process oriented approach paradigm that can account for the daily ups and downs of motivation to learn, that is the ongoing changes of motivation over time. (p. 83)

To account for the changing nature of motivation, Dörnyei (2005) introduced the Process Model of L2 Motivation. The model describes a cyclical pattern for motivation during the learning process involving three stages. The first is the Preactional Stage where learners set goals and create a plan of action. Learners’ motivations might include attitudes towards speakers of the L2, expectancy of success, instrumental or integrative goals, and support from their environment. The Actional Stage involves completing the necessary tasks to achieve their goal, self-regulation, and ongoing appraisal. In this stage, motivation may appear in the form of satisfaction with the learning environment, a sense of autonomy, positive or negative feedback from classroom teachers or their L2 community, and self-motivating strategies. Lastly, the Postactional Stage completes the
cycle. Learners evaluate progress, find causal attributions for successes or failures, and continued planning. During this stage, learners may be motivated by self-concepts and beliefs, positive or negative feedback, or attributional factors. Ideally, after evaluation, they would return to the first stage and set new goals. At this point, empirical evidence documenting the process is unavailable. Consequently, it remains a theoretical model applied to studies that investigate change retrospectively.

That is not to say that changing motivation has not been investigated. Shoaib & Dörnyei (2005) interviewed 25 L2 English learners between 18 and 34 that began learning English at a young age. From these semi-structured interviews, they collected data on changing motivational factors that influenced the subjects throughout their English language learning experience. They classified the data into seven motivational dimensions: Affective/Integrative, Instrumental, Self-Concept, Goal-Orientated, Educational-Context, Significant-Other, and Host-Environment. Learners varied in frequency of times they mentioned each of the dimensions, thus indicating individual variance. The most frequently mentioned positive and negative influences were related to the Self-Concept Dimension. Learners generally mentioned their attitudes regarding learning the language, the language itself, and its speakers. Since the interviews covered a life-time of language learning, they often mentioned life-changing events that affected their motivation, which Shoaib & Dörnyei (2005) referred to as motivational transformational episodes. They classified these episodes into six categories including maturing, stand-still periods, moving into a new life phase, internalizing external goals
and visions, relationships with significant others, and time spent in the host environment. While these types of events do not always affect learners’ language development, learners’ mentions of them in connection with language learning indicate their saliency to the learners’ individual motivation trajectories. Just as proficiency development is individual and non-linear, their findings show that learners also have individual non-linear motivational trajectories.

Recently, motivation research has devised methods to document fluctuations and non-linear trajectories in motivation. Researchers still incorporate self-reported information from surveys or interviews, but rather than relying entirely on learners’ retrospective evaluations of their motivation, researchers have begun collecting data two or three times during the learners’ immersion experience. For example, Campbell & Storch (2011) tracked the motivation of eight L1 English L2 Mandarin learners over a semester. Through semi-structured interviews conducted at weeks three, seven, and ten during the twelve-week semester, they identified eight categories of motivational factors: language learning history, language related enjoyment, personal satisfaction, external pressure, influences or incentives, positive feelings regarding China, personal goals, desired L2 competence, and identity. They found that students had different motivational factors and motivational trajectories. Three participants felt their motivation had weakened, while five felt it became stronger. In some cases, the same variable would lead to an opposite reaction in a different participant. As an illustration, the increasing difficulty of the course motivated some students while demotivating others.
By documenting the changing levels, motivation research can be incorporated into the study of a dynamic system. The present study has the unique advantage of comparing the participant’s CAF scores to the motivational events at the time both occurred. Since Juan and I discussed his struggles and triumphs throughout the study, these motivational events were documented in session transcripts. Rather than considering motivation as a separate variable to be correlated with achievement, this study considers motivation as part of a dynamic system as is called for by Dörnyei (2009).

2.5 Longitudinal Case Studies of Natural L2 Learning

The last element of Juan’s dynamic development in this study is time. Since second language development is an individual process that may take place over a lifetime, a great deal of knowledge can be gained from following an individual’s development over time. Longitudinal case studies provide information that is often unavailable in cross-sectional and experimental studies and allow researchers to follow the development of learners, instead of focusing on the end result. Duff (2014) defines a case study as “a phenomenon being examined closely within the context of the case-in-context and against the backdrop of existing theory and research” (p. 237). As such, case studies may add explanation to, show flaws in, or provide evidence for or against the current theories of second language development (SLD). In some cases, individual case studies influenced major shifts in SLD theory.
During the 1970s, SLD researchers generally limited the scope of their research to the classroom and disregarded language learning taking place in naturalistic context. The seminal study by Schumann (1978) followed “Alberto”, an L1 Spanish speaking immigrant from Costa Rica, during his 10-month immersion in English in the United States. Schumann (1976, 1978) documented Alberto’s inability to develop beyond pidginized (basic communicative) language forms in the areas of negation and wh-questions. He also showed little improvement in the areas of possessives, past tense, pluralization, and progressive aspect. Schumann (1976, 1978) theorized that Alberto’s lack of acculturation led to psychological distance between Alberto and the L2 community, which resulted in a lack of motivation to learn English. Alberto lacked opportunities for interaction with L1 speakers, lived in an enclosed community of mostly Spanish speakers, and was planning to return home. Schumann (1976, 1978) argued that Alberto had little instrumental or integrative motivation. By showing the importance of L2 interaction and integration, Schumann (1976, 1978) encouraged other researchers to investigate learners’ amount and quality of interaction rather than assuming all naturalistic learners were equally exposed to their L2.

In response to Schumann’s (1976, 1978) claims regarding psychological distance and lack of acculturation, Schmidt (1983) introduced his study of “Wes”, a learner who acculturated, was comfortable in his environment, and had communicative competences (e.g., politeness, memorized scripts, and persistence in communicating). Despite his acculturation, positive interactions with native speakers of English, and pragmatic
improvements, Wes did not develop basic grammatical forms. For example, he did not consistently or properly use the regular or irregular past tense, during the two-year study. Thus, Schmidt (1983) argued that learners need more than acculturation or motivation to acquire grammatical accuracy; they need attention to grammatical form.

As a follow-up, he put forth the Noticing Hypothesis (Schmidt, 1990), which suggests that close attention to language forms (i.e. noticing) are necessary for acquisition of grammatical forms by adult L2 learners. He explained that noticing is encouraged by explicit instruction or tasks that require focus on the language form to complete them. Schmidt’s (1990) Noticing Hypothesis contributed to current Focus on Form research such as Mackey (2006) and Spada & Lightbown (2008).

In both Schumann (1976, 1978) and Schmidt’s (1990) longitudinal case studies, researchers documented the learner’s syntax, vocabulary, and evidence of acculturation to provide examples in support of or against the central debates surrounding second language “input” at the time. During that period, Krashen (1987) was arguing that learners only required input slightly above their level (i+1) to acquire language implicitly. In contrast, Schumann (1976, 1978) showed that learners development requires motivation to acculturate, and Schmidt (1983) showed that learners needed to focus on form in addition to meaning.

Ortega and Iberri-Shea (2005) argued that present day SLD research also needs more longitudinal work to “contribute meaningful characterizations of the gradual process of attaining advanced second language and literacy competencies across various
contexts” (p. 28). They identified four types of longitudinal research that could make such contributions: L2 program outcomes, L2 instructional effectiveness, qualitative research, and descriptive quantitative results. The first analyzes large amounts of data over time to evaluate policies and practices in L2 curricula for the purposes of evaluation and change. The second, L2 instructional effectiveness, uses quasi-experimental techniques repeated over time to investigate the effectiveness of L2 instruction techniques. The third, qualitative research, relies on sociocultural approaches, ethnographic perspectives, and ethnographic techniques. Lastly, descriptive-quantitative studies rely on descriptive statistics to describe development in individuals or small groups.

Since the present study relies heavily on descriptive-quantitative techniques, I will discuss this approach in more detail. The purpose of descriptive-quantitative research is to document quantitative measures over time (e.g., proficiency scores, average number of pauses, frequency of errors). Descriptive statistics from those measures are developed into learner trajectory graphs and analyzed for trends. Researchers may also add a qualitative analysis to descriptive statistical studies including discussions regarding life events, motivating factors, explicit instruction, and interaction with target language speakers (e.g., Larsen-Freeman, 2006; Polat & Kim 2014; Vyatkina, 2012). Ortega and Iberri-Shea (2005) describe the main strength of descriptive-quantitative research as providing “evidence for time relationships among variables, thus helping to explain causes and effects by looking at antecedent-consequent relationships” (p. 30).
Vyatkina (2012) used the descriptive-quantitative method in her study that compared two L1 English L2 German learners with their classmates using 14 writing samples collected over 4 semesters. The purpose was to compare learner development in six areas of language complexity: mean sentence length, distinct finite verb units, words per finite verb, coordinating conjunctions, subordinating conjunctions, and vocabulary type/token ratios. For the quantitative data, she created three graphs for each measure, one for the class averages and one for each of the two target participants. She plotted the scores of the target participants creating trajectory graphs and adding trendlines. Then, she added the 95% confidence interval from the class averages. She found that in all but two variables, words per finite verb-unit and coordinating conjunctions, the trendlines moved in the same direction as the class. However, the target participants developed non-linearly – their graphic trajectories often included large fluctuations outside of the 95% confidence interval. This directed her to closely analyze the differences.

Vyatkina’s (2012) analysis of sentence length revealed developmental differences in her target learners. She found that the first target participant scored higher than the 95% confidence interval at the beginning, then dropped below the class during the middle of the study, and ended within the 95% confidence interval of the class. The second target participant remained close to the 95% confidence interval at the beginning of the study, then gradually fluctuated further and further above and below the 95% confidence interval. When the researcher looked at the how they were lengthening their sentences, she found that the first target participant used more finite verb phrases, while the second
included more clausal units per sentence. Since clausal units are usually found later in development, she argued that the second participant showed greater development even though her trendline fell below the class average. If she had chosen to only examine the whole group, she would have found that the class showed a linear trajectory of improvement. However, her results revealed that individuals develop differently and that a global measure such as sentence length may not reveal the whole picture. These findings illustrate one of the key advantages of descriptive quantitative method, the ability to investigate variation at an individual level.

Duff (2014) explains the benefits of case studies as the ability to understand “individuals’ experiences, issues, insights, developmental pathways, or performance within a particular linguistic, social, or educational context” (p. 233). She also mentions promising new possibilities for case studies in which linguistic measures combine with sociocultural aspects to provide a “more balanced description of the subjects’ oral and written linguistic performance and their personal perspectives on their learning or use of language” (p. 235). The present study answers the call for this type of research by comparing Juan’s CAF scores to his self-reported perspectives concerning his successes and failures, major life-events, native-speaker exposure, and motivation.

2.6 Summary of the Literature Review

The literature reviewed here reveals several gaps in research to date, starting with the scarcity of longitudinal studies of L2 development. While experimental and cross-
sectional studies have shown effectiveness in CAF measures, differences in task modalities, and improvement among language learners in the areas of CAF variables, little is known about individual development over time. In addition, even though the effects of motivation and native-speaker exposure have been shown to affect language development, little is known about how changes in those variables affect CAF development over time. Furthermore, the complex and dynamic nature of CAF, native-speaker exposure, and motivation remains unexplored. To examine these issues, researchers need a theoretical framework that treats language development as a system that changes over time. In the next chapter, I will introduce such a framework, Complex Dynamic Systems Theory, and explain how the tools developed for it are able to measure change and interaction over time.
CHAPTER 3
THEORETICAL FRAMEWORK

A pioneer of Complex Adaptive Systems, mathematician Holland (1998) explains dynamic systems by comparing it to a chess game. Before the game begins a certain set of options are available to the players. As soon as the first move is made, the set of available options changes. The second move results in a new set of options, and so on. Each move creates a new set of circumstances from which the next move is selected. Consequently, predicting how the game will progress is nearly impossible at the beginning of the game. Even though there are a finite number of possible endings (the king is checkmated, a player resigns, or the game ends in a draw), the path to arrive at the end may be very different each time. While traditional research focuses on the endpoints, researchers operating under a complex dynamic systems framework analyze the process of getting there.

The previous chapter shows that the process of SLD is affected by many factors, yet traditional research techniques limit the researcher’s ability to explore how those factors change and interact over time. For example, while we know that motivation influences a learner’s results, researchers rarely consider how changes in the learner’s motivation over time relate to changes in the learner’s proficiency over time. Rather, traditional SLD methodologies compare the initial results of a limited number of factors to the end results. If there is a statistical difference between them, then the factor is labeled as causative. While such research has advanced the understanding of SLD, the
results tend to be reductionist in that other factors which were not considered that may also be contributing to the results. As Larsen-Freeman (2013) explains, “Such approaches too readily dismissed variability as noise or measurement error or attributed it to ‘outliers’. They treated context as a backdrop, removed from the main action” (p. 369). CDST offers an alternative theoretical lens operating under principles that allow researchers to analyze how multiple variables affect each other and develop over time. Larsen-Freeman (1997) outlines those principles stating that language a) is dynamic, complex and nonlinear, b) is chaotic, unpredictable, and sensitive to initial conditions, c) is open, self-organizing, feedback sensitive, and adaptive, and d) has attractors and a fractal pattern. In what follows, I will further explain the basic theoretical principles behind the CDST, present tools unique to CDST, and review studies that utilize them.

3.1 Complex Dynamics System Theory (CDST)

Just as in the chess game described by Holland (1998), each change within a language system creates a new set of circumstances. In order to analyze changes in the system, CDST researchers establish an initial state which represents the learner’s current circumstances, including the learners’ previous language background, proficiency, and environmental circumstances at the beginning of the study. The learner’s initial state acts as a point for departure from which to begin their analysis. CDST defines two categories that may bring about changes in the learner’s initial state: the social ecosystem and the
cognitive ecosystem (de Bot, Lowie & Verspoor, 2007; Larsen-Freeman & Cameron, 2008). The social ecosystem refers to external elements which could include classroom instruction, language exposure, and L2 exposure. In contrast, the cognitive ecosystem refers to factors within the learner such as aptitude for language, motivation, the learners’ L1, and the learner’s proficiency level. The initial state is constantly being acted upon by many agents from the social ecosystem, which results in constant change in the cognitive ecosystem and unpredictable behavior throughout the entire system (de Bot et al., 2007; Larsen-Freeman & Cameron, 2008). In the same manner as the chess game, learners’ developmental paths vary greatly depending on changes within their social and cognitive ecosystem. Changes to any of these elements singularly, or due to interaction between them, may result in changes in the learner’s language development (de Bot et al., 2007).

As an example, an ESL student learns the word for “peach,” her favorite fruit, in class one day. She is excited to try it, and goes to the farmer’s market which operates in her L2 near her house the next morning. When she tries to ask for a peach, she gets confused and pronounces it /piʃ/. The native-speaking clerk corrects her, and she is embarrassed. When she returns to class that afternoon, she refuses to say the word in fear of embarrassing herself again. In this scenario, there are two social elements (the class and her interlocutor) and three cognitive (including emotive) elements (knowledge of new word, excitement, and embarrassment). These elements appear and interact in unpredictable ways. If any of these events had happened slightly differently, the results would have been quite different. CDST researchers refer to this concept as the butterfly
effect, based off the idea that a butterfly flapping its wings (a small element) could change the weather in a distant location (a large effect). The term was introduced in a Ray Bradbury story, *The Sound of Thunder* (Bradbury, 1952), in which a time traveler accidently crushes a prehistoric butterfly causing a series of events that change the future. His small error caused changes including people’s behavior, language, and even changed the outcome of a presidential election. It was adapted by Edward Lorenz, a meteorologist who found such an effect in the 1960s while making minute changes to atmospheric system models designed to predict the weather (Resler, 2016).

Although the interactions between variables and the results of those interactions are often unpredictable, systems self-organize and may move towards a new attractor state (Holland, 1998; 2006; Larsen-Freeman, 1997; Verspoor & van Dijk, 2011). *Attractor states* represent the options available to the learner at that moment and may be limited by cognitive factors such as attention, motivation and memory (de Bot, et al. 2007). As such, the new attractor state is not necessarily the best option, but one of the many options with which the system experiments. In our previous scenario, /piʃ/ represents an unsuccessful attractor state. Since our learner likes peaches, she is likely to try again until she achieves the correct pronunciation, which represents an alternative attractor state. In this case, the learner is motivated to move from one attractor state to another by her social ecosystem, which provided the correct form and the motivation to learn it.
Kauffman (1995), a biologist who also studies CDST, explains biological evolution based on attractor states. He describes organisms progressing towards new traits by making slight mutations. Those mutations may be random, but the mutations lead to additional mutations based on success in the environment, or a form of self-organization. Moreover, changes in the initial state of their ecosystem, may also lead to mutations. A small change may result in slight adjustments (mutations in a digestive enzyme for a new food source) or large changes over time (development of a specialized beak to dig a new form of insect out of tree bark). In the latter case, organisms will produce several strings of mutations to find a path to stability. In the former case, small mutations will make slight adjustments in order to maintain stability. Kauffman (1995) refers to the balance between stability and instability as “the edge of chaos” (p. 86). If a system is too stable (rigid), the introduction of a new element may send the organism into extinction as it will not be able to adjust. If the system has some instability, it is able to mutate and make the necessary changes to survive.

The evolution of a learner’s language development can be viewed in a similar manner. Learners’ initial states interact with changes in their cognitive ecosystem or their social ecosystem (or both at the same time) as they are moving towards an attractor state (new pronunciation, grammar forms, or vocabulary). A learner’s determination to achieve a goal (e.g., passing a class or communicating with others), may direct and organize the system (self-organization). As in evolutionary processes, learners use trial and error to move to different attractor states before finding one that suits their cognitive abilities and
communicative needs. For example, students learning the past tense may overgeneralize *(hitted)* until they learn the correct irregular verb form *(hit)*. In this case, both *hitted* and *hit* represented an attractor state in the evolutionary process. Even when learners have explicitly learned a new form, they may not use it correctly every time. They may make mistakes, try another form to express the same content, or forget it temporarily and begin using it again months later. Van Geert and van Dijk (2002) suggest yet another option – *bimodality*. In these cases, the learner may sometimes follow the “old rule” and other times follow the “new rule” (p. 365). The system may eventually stabilize or it may remain in a bimodal attractor state indefinitely. For example, the learner might use memorized lexical chunks to produce the third person singular accurately sometimes, while omitting the “s” during creative production. Consequently, the learning trajectory for this form is non-linear in that some speech samples may contain many errors, while others contain few or none.

Moreover, a small change could result in large changes to the entire system. For instance, if an L2 English student learns the word *care* as a verb, that student potentially also learns *care* as a noun and may quickly add *caring* (adjective) and *caringly* (adverb) depending on whether the learner is exposed to it and has a need for it. In addition, the student may add prefixes (uncaring) and affixes (careful) and learn phrases such as “handle with care” and grammar forms such as “care for” or “care about”. Thus, a small addition to the learner’s vocabulary has the potential to create changes to that learner’s lexical diversity and grammatical accuracy. Furthermore, fluency may improve in that the
student can quickly access the word rather than talking around it and utilize lexical chunks thereby increasing the automaticity of their production.

One element of language building on the progress of another element is often investigated in CDST. These relationships are referred to as connected growers (Verspoor & van Dijk, 2011). For example, children learning their L1 must reach a certain level of vocabulary before they begin to acquire syntax. As learners advance, they begin to convert nouns into other parts of speech. Hence, more complex syntax structures and morphology appear as more vocabulary is learned (Lowie, Caspi, van Geert, & Steenback, 2011). In this case vocabulary and syntax are conditional growers, meaning one must exist or reach a certain level of proficiency before the other occurs. After learners have sufficient vocabulary and grammar, they may focus on one or the other while producing language. Therefore, while concentrating on accurate grammar production they may use simple vocabulary and while they are concentrating on vocabulary they may use simple grammar. In this case, vocabulary and grammar are competitive growers, meaning that while one increases in usage the other decreases. Verspoor et al. (2008) would argue that vocabulary and grammar are competing for attentional resources in the cognitive ecosystem. While one aspect is receiving those attentional resources, the other is neglected, resulting in a negative correlation. The last type of connected growers, cooperative growers, improve simultaneously. For example, in a case study of an L1 Dutch L2 Finnish learner, Spoelman and Verspoor’s (2010) found a positive correlation between word complexity and sentence complexity. For this
learner, when words became longer, sentence complexity also increased resulting in a positive correlation.

The last two theoretical concepts I will cover relate to how and when changes in the learner’s system occur, an under-developed area in traditional SLD research. CDST research supports two options: slowly emerging changes or periods of extreme fluctuations followed by stabilization. Emerging changes may appear due to iteration, or exposure to repeated patterns (de Bot et al., 2007; Larsen-Freeman & Cameron, 2008). The new form may be learned by repetition in explicit instruction or by noticing and imitating phrases in the social ecosystem (Ruhland & van Geert, 1998). To illustrate this point, Ellis and Larsen-Freeman (2009) used computer models to show how common prototypical words in the English language appear as exemplars in second language learners’ speech due to exposure. These common exemplars (go, put, get, and give) tend to be overused until learners develop enough vocabulary to add more specific, and consequently less common, words (walk, run, drop, place, pass, etc.). Thus, repeated exposure in the environment leads to the gradual emergence of the most common forms followed by less common forms. The same concepts apply to grammar as Ellis and Ferreira-Junior (2009) show in their study of English verb-argument constructions.

While some forms appear gradually, other forms may appear suddenly and with great frequency. As Ruhland and van Geert (1998) explain, “language production is the expression of underlying syntactic rules that are … dependent on performance limitations that must be overcome somewhere along the developmental path” (p. 69). Whether those
performance limitations are based on shyness or cognitive understanding of the form, learners have been building to a point of production. Once the necessary change to produce the form occurs, learners overcome the obstacle, and suddenly begin producing the form. Larsen-Freeman (1997) refers to this phenomenon as the camelback effect, referring to the metaphor of the straw that broke the camel’s back. It implies that one additional element, or piece of straw, is enough to cause considerable changes in a learner’s language system. However, rather than producing the form correctly every time, the learner may fluctuate between the new form and the form they previously used to communicate the same idea.

The process of learning a new form, fluctuating between the old and new form, and finally settling into a restructured new form is referred to as a phase shift (Larsen-Freeman, 1997; Larsen-Freeman & Cameron, 2008). Unlike emerging forms, phase shifts involve a stage of fluctuation followed by stabilization, rather than steady growth. If learners acquire the new form, it becomes the stronger attractor state, the language stabilizes by restructuring, and the previous usage disappears (Polat & Kim, 2014). If learners do not restructure their language, the production will return to the previous attractor state after the period of fluctuation. This may happen for a variety of reasons, such as the explicit instruction ceasing too soon or decreased exposure to native speech. The result will be that the learner’s language stabilizes again but without restructuring. Therefore, the phase shift is incomplete.
In sum, CDST emphasizes the importance of considering the whole system, especially interaction between more than two variables. Unlike traditional research, CDST tends to longitudinally examine the development of multiple variables in individuals or small groups. From this information, researchers try to explain the development process and explore the relationship between variables. However, Hiver and Al-Hoorie (2016) point out the need for a “practical blueprint to ensure compatibility between its theoretical tenets and empirical SLD research designs” (p. 741). Without limiting the variables and parameters of the language system under investigation, CDST research would become impractical based on the sheer volume of data necessary to conduct the study. Thus, their article proposes a template they refer to as the Dynamic Ensemble. It suggests CDST scholars develop parameters for their research based on operational considerations (validity, agents, and time); contextual considerations (environment and interconnected systems); macro-system considerations (variability and changes in the system); and micro-structural considerations (components of the system and their interactions). The next section will further explain the tools developed by CDST researchers and provide examples of how researchers have focused their studies to investigate the data as a system rather than separate variables.

3.2 CDST Methods and Studies

Language development researchers using CDST collect longitudinal samples of data in order to examine emerging changes, phase shifts, and interactions between
variables as they develop. As such, CDST researchers prefer written or oral samples with enough text to analyze multiple variables and “where all aspects of the linguistic production process … are, as far as possible, fully under the control of the learner” (Schmid, Verspoor & MacWhinney, 2011, p. 39). Additionally, Larsen-Freeman (2006) states that to capture changes in the system, data collection should be “frequent enough to capture the relevant properties underlying the developmental process” (p. 595).

Depending on the variables being explored and the level of the student, data collection frequency and modality may vary considerably. Consequently, the studies reviewed below have differences in the length of the study (six months to three years), the sample modality (written essays or spontaneous speech) and the intervals in which the data were collected (two weeks to one year). Researchers using this theoretical framework must acknowledge that some variables are unknowable, unmeasurable, or beyond the scope of the study, and that due to unforeseeable changes within learners’ environment, predicting a learner’s path is highly unlikely (de Bot, 2008; Larsen-Freeman, 1997). Thus, most CDST studies look at data retrospectively rather than making predictions based on statistically significant results of large groups of learners. Accordingly, they have developed analytical tools capable of such analyses. This section explains those tools and introduces studies that have used them.

Polat and Kim (2014) exemplify phase shift analysis at a microgenetic level using tools developed by Verspoor et al. (2008) including moving averages and min-max graphs. They analyzed an L1 Turkish L2 English naturalistic learner’s development of
accuracy and complexity. They collected spontaneous conversation samples every two weeks for a year (24 samples) which they coded for four complexity measures ($D$, words per clause, mean length of clauses, and number of clauses per AS-unit) and two accuracy measures (percentage of error free clauses and errors with present simple accuracy). To analyze the data for development and phase shifts, they utilized trajectory graphs, polynomial trendlines, moving averages, and min-max graphs. *Trendlines* are mathematically calculated lines added to graphs to display a tendency in the data. Polat and Kim (2014) used trajectory graphs with trendlines to examine emerging development. In order to obtain a clearer picture, they also utilized moving averages. *Moving averages* create new data points by averaging the scores of a data point and two or three data points on either side of it. Moving averages are calculated on every data point resulting in a new trajectory that is smoothed, but avoids averaging all the data which would eliminate the ability to observe change. Moving averages are generally used when the pattern of the data is difficult to see.

Next, Polat and Kim (2014) added min-max lines to the moving average graphs, which create a min-max graph. *Min-max graphs* display the average of the minimum and maximum scores of the data point and two to five data points on either side in a similar manner to a moving average. The maximum line is plotted above the actual data points and the minimum below the actual data points on the same graph. Thus, the min-line and the max-line frame the actual data (for examples see Figures 18 – 21 in section 5.1.2). When the distance between the minimum and maximum lines widens suddenly and
maintains that distance for a substantial period, it indicates a period of fluctuation. A phase shift occurs if the fluctuation is followed by a smaller distance between the min-max lines that is higher or lower than the scores before the fluctuation period (Larsen-Freeman & Cameron, 2008; Larsen-Freeman, 2006). This pattern of stable scores moving to fluctuation then moving to a different level of stable scores indicates that the learner has shifted from one stable attractor state to another stable attractor state. In short, the language has been restructured by the learner.

By using min-max graphs, Polat and Kim (2014) found periods of fluctuation in clauses per AS-unit, words per clause, global accuracy, and present simple accuracy. The first three (clauses per AS-unit, words per clause, and global accuracy) did not result in restructuring but returned to their previous attractor states. Meanwhile, present simple errors appeared to remain in a period of fluctuation during the entire study. Polat and Kim (2014) also found emerging improvement from the trajectory graphs with trendlines in words per AS-unit, $D$ (lexical diversity), and words per clause.

Polat and Kim (2014) suggest that these outcomes are explained by the learner’s untutored naturalistic learning. Without the benefit of explicit instruction, the learner’s language only progressed for communicative functions. Therefore, while vocabulary and elaboration increased, accuracy did not. They also suggest that the phase shift related to present simple accuracy represented a bimodal attractor state, meaning that the learner was shifting between two attractor states consistently and long enough to be considered a stable bimodal attractor state. In terms of CDST, Polat and Kim (2014) showed that the
traditional trajectory analysis does not reveal the whole picture of language learning. While their learner steadily improved in some variables, others worsened or remained in periods of fluctuation. Their study indicates the need for more microgenetic exploration of language development including variation analysis.

Larsen-Freeman (2006) took a broader approach by comparing four essays about the same topic written over a period of six months by five L1 Chinese L2 English learners. Her purposes were considerably different than Polat and Kim (2014) in that she investigated the interaction between the CAF constructs and intra-individual differences between the five learners’ development. For the CAF constructs, she measured complexity (clauses/t-unit), accuracy (error free/total t-units), lexical diversity (corrected type/token ratio), and writing fluency (words/t-unit). When the learners’ scores were averaged together, all constructs showed gradual development. However, when she analyzed the scores separately, each learner showed unique developmental trajectories, some improving while others worsened in the same constructs.

To investigate intra-individual development further, Larsen-Freeman (2006) calculated the z-scores for each of the CAF variables, then plotted each learners’ z-scores onto one graph. She found that each student had a unique trajectory pattern that was often markedly different than another student. For example, student H decreased in complexity and increased in accuracy, while student L increased in complexity and decreased in accuracy. Larsen-Freeman (2006) suggested that the differences could be
based on whether the students were focusing on form (increasing accuracy) or expressiveness (increasing complexity).

Verspoor et al. (2008) investigated connected growers and cross-linguistic influence using correlations between variables. Their study analyzed 18 academic essays written by an advanced L1 Dutch learner of English over the period of three years. They compared variables within the construct of complexity, including frequency of words from the Academic Word List (Coxhead, 2000), average word length, TTR, average length of noun phrase, average words per finite verb, and average sentence length. By comparing two variables (sentence length and TTR) they demonstrated how raw data and standard correlations between two variables may overlook the complexity of the data. To examine the data for competitive growers, Verspoor et al. (2008) began with a standard correlation which showed a very weak negative correlation ($r = -0.03$). Then, they used linear regression to detrend the data, meaning that they statistically removed the general upward development of the data over time. After detrending the data, they tried the correlation again, with slightly better, but still not statistically significant, results ($r = -0.33, p = .087$). Lastly, they tried a moving correlation, which showed changes in the correlation over time. They found that at the beginning and end of their study, the data had a positive correlation, yet in the middle of the study the data was negatively correlated. Thus, the change in the correlation led to non-significant results. However, by looking at the moving correlation trajectory over time, they could see a strong negative correlation in the majority of the data (11 out of 18 points). They commented that “The
oscillations found in the correlation patterns are fascinating because they suggest that there is a complex relation between varied word use (i.e., TTR) and the length of the sentences (SL), which changes dynamically over time.” (p. 225).

Next, Verspoor et al. (2008) looked at the relationship between words per finite verb and sentence length. They hypothesized that these variables would be conditional growers since words per finite verb would generally result in longer sentences. However, they found that there was no correlation or pattern. Since Dutch uses more noun phrases to add complexity than English, they theorized that cross-linguistic influence was affecting their hypothesized outcome. Therefore, they analyzed the correlation between noun phrases and words per finite verbs. This time, there was a significant positive correlation \( (r = 0.516, \ p = 0.02) \) indicating that they were connected growers.

By using various tools unique to CDST and operating under the CDST theoretical framework, Polat and Kim (2014), Larsen-Freeman (2006), and Verspoor et al. (2008) pioneered our understanding of how variables progress in a non-linear trajectory and interact with each other as they develop over time. While Polat & Kim (2014) and Larsen-Freeman (2006) showed that variables develop non-linearly and in trajectories unique to the individual, Verspoor et al. (2008) demonstrated how interaction between variables can be analyzed to show dynamic relationships between them.
3.3 Gaps in the Current Research

Based on the work of Larsen-Freeman (2006), Polat and Kim (2014), and Verspoor et al. (2008), new research directions have emerged. Their work shows that there is still a vast gap in our knowledge about how the elements of language interact and develop as a system. The purpose of this study is to contribute to SLA knowledge in four areas in need of further investigation: non-linear development in the CAF variables, interactions between the CAF variables, the effects of a changing social ecosystem on development, and the effects of task differences in multiple modalities.

While a great deal of research exists on the development of the CAF constructs separately, most of that research is conducted on groups of learners using either experimental or cross-sectional design (e.g., Bulté & Housen, 2014; Kormos & Dénes, 2004; Vyatkina, 2012). This research is informative but leaves a gap in our understanding of how the variables develop over time and how they interact with each other. Studies incorporating CDST have shown that L2 learners’ development is non-linear and unpredictable (Larsen-Freeman, 2006; Polat & Kim, 2014; Verspoor et al, 2008). CDST also theorizes that changes in learners’ language may emerge slowly over time or suddenly appear in the form of a phase shift. Despite the efforts of Polat & Kim (2014), a completed phase shift has yet to be documented. Continued research in this area may show that certain constructs or variables are more likely to develop in a phase shift or that longer amounts of time are necessary to capture a complete phase shift.
Next, SLD studies rarely compare the development of all three constructs at once, with Larsen-Freeman (2006) being the exception. Since complexity and accuracy are indicators of L2 knowledge, whereas fluency primarily shows performance ability, comparisons between the development of the constructs can be complicated. Fortunately, task-based research has provided insight into CAF interaction during development by showing the effects of cognitive strain on the CAF variables and the effects of shifting focus between the variables (Skehan and Foster, 1999; Tavakoli & Foster, 2011). The tools utilized for labeling connected growers in complexity and accuracy could also be utilized to explore the relationship between performance and knowledge variables. While the label of “growers” may not be appropriate, such comparisons would show how fluency changes in relationship to developing accuracy and complexity. The results may contribute to our understanding of fluency development and possibly provide information regarding the levels of cognitive strain in relation to development in the knowledge variables.

In theory, CDST considers both the social and cognitive ecosystem, yet research rarely combines them. While Larsen-Freeman (2006) and Polat & Kim (2014) both mention the social ecosystem, that aspect of the learners’ system is presented as a possible explanation of the results rather than a focus of the study. To understand a complex dynamic system, this is an area in need of more exploration. While research has shown that motivation affects development (Gardner, 1968; Masgoret & Gardner, 2003; Dörnyei 2009), when, how, and on which parts of speech remains under-explored.
Lastly, researchers carefully choose their data collection stimuli in order to maximize the consistency and validity of their results. In longitudinal studies, the choice is more challenging due to practice effects and difficulties maintaining learners’ interest. Moreover, in studies such as Vyatkina (2012) which took place over four semesters, it was necessary to adjust the tasks for the learners’ increasing levels. While task-based learning explores subtle differences between similar tasks (Tavakoli & Foster, 2011), research between tasks with obvious differences (e.g. modalities, levels of difficulty, and levels of interaction) is lacking.

3.4 Research Questions

To address the gaps discussed above, this study focused on four research questions.

1) Second Language Development: Over 15 months in the US, what changes does Juan make in the areas of complexity, accuracy, and fluency?

2) Global Variable Interaction: Is there evidence of interconnected development between the global variables? Do they compete, support, or build on each other, or in CDST terms, are they connected growers? Does the interaction between knowledge variables and the performance variables show evidence of cognitive strain?
3) Environmental Factors: Is Juan’s CAF performance affected by native speaker exposure and interaction? Is Juan’s CAF performance affected by fluctuations in his motivation?

4) Task effects: Is Juan’s CAF performance affected by the type of task being performed?
CHAPTER 4

METHODOLOGY

4.1 Participants

4.1.1 Juan

Juan, a native speaker of Spanish from Colombia, arrived in the United States in August of 2012 when he was 33 years old. He began participating in this study on February 25, 2013 and continued until May 23, 2014. In total, the study lasted 15 months. Juan attended Temple University as a Fulbright Scholar studying music theory. To qualify for the Fulbright Program, he took the TOEFL iBT Examination on which he scored 80, placing him at intermediate low (Educational Testing Services, 2016a). Study abroad research suggests that this is the ideal level for learners to proceduralize the explicit knowledge they gained from formal instruction (DeKeyser, 2014; Pérez-Vidal, 2014). Juan indicated that he was aware of the advantages learning English would bring him. By earning his master’s degree in the United States, he would increase his pay and prestige at his home university, his ability to interact with American scholars who visit there, and his ability to use the English textbooks at his university. Juan decided to participate in the study in order to practice conversation. He reported being introverted and was looking for someone with whom he could practice his English. In sum, he appeared to have instrumental orientation and motivation.
4.1.2 Comparison Group

For the purpose of exploring colloquial speech acquisition, establishing speech community norms for the CAF variables, and separating second language effects from task differences, data were collected from a comparison group comprised of 20 L1 English undergraduates. Participants were between the ages of 18 and 52 ($Mdn = 20$) and included twelve males and eight females. The group contained speakers with regional accents and speakers of African-American English (AAE) thus making them representative of the language Juan might have heard while living in the Northeast. Students were recruited from three undergraduate general education classes and one undergraduate honors anthropology class. All but two participants had studied an L2, the most common being Spanish (31.6%). Participants ranked their L2 proficiency in speaking, listening, reading, and writing on a Likert scale of zero to seven, seven being completely fluent, and zero having no proficiency. The average L2 proficiency scores were: speaking $2.84$ ($SD = 1.25$), listening $2.63$ ($SD = 1.38$), reading $3.00$ ($SD = 1.79$), and writing $2.89$ ($SD = 1.69$). Additionally, 35% had studied more than one language. In these cases, their most proficient additional language was selected for the L2 analysis. Since the participants L2 scores were low, it is unlikely that their L2 influenced their L1 particularly because the majority studied in a classroom in the United States with little exposure to native-speakers.
4.2 Research Design

4.2.1 Data Elicitation

Juan’s data were collected at bi-weekly sessions held in my office over a period of 15 months, from February 25, 2013 – May 23, 2014. During that time, we had to suspend sessions four times from May 24 - June 21, 2013; August 1 – September 4, 2013; December 30, 2014 - January 17, 2015; and February, 24 - March 17, 2015 when one or both of us were unavailable. In addition, the original study design included tasks to measure cognitive restructuring, which were not included in this analysis. Therefore, some additional gaps are present in the data.

Since one of Juan’s goals was to increase his English conversation skills, many of our sessions were only conversational. From those sessions, four samples of conversation distributed evenly throughout the 15 months were chosen for analysis. Additionally, description tasks elicited from a series of pictures (photographs and cartoons) and narration tasks elicited from a children’s storybook and a short video were collected. From that data, four samples from the picture descriptions, five samples from the storybook narrations, and four samples from the video retellings were then selected. In addition, two frog stories were repeated towards the end of the study for comparison to Juan’s early samples. These tasks and their collection dates are described in detail in the next section on data elicitation stimuli. The tasks were chosen to provide Juan with a variety of difficulty levels, modalities, and topics of conversation. The diversity of the tasks and stimuli were also helpful in preventing practice effects and participant’s
disengagement due to boredom that often accompanies frequent data collection. All tasks were digitally recorded and transcribed in CHAT format (MacWhinney, 2000). To properly code for fluency and accuracy, the transcribers included filled pauses, stutters, false starts, repairs, and audible emotional responses such as laughter in the transcripts.

Oral narratives and description tasks were chosen to collect samples that approximated spontaneous speech, but could still be controlled for content, thus allowing comparable data over time (Pavlenko, 2008). Oral tasks were chosen over written tasks in order to analyze the effects of cognitive strain on oral fluency, and due to the additional cognitive load added by writing. Moreover, complexity has been shown to develop in oral speech before written text (Chan, Verspoor & Vahtrick, 2015) and task-based research has only focused on oral elicitation and would not be comparable to written data (Skehan & Foster, 1999; Skehan & Shum, 2014; Yuan & Ellis, 2003).

Each participant of the comparison group completed one story, one video, and one set of picture descriptions. The samples were collected during one session held in a university office. The specific book, video, and set of pictures were chosen to complete a corpus for another study. Each task represents all the tasks in that genre (book, video, or picture description) due to the impracticalities of having each volunteer participant complete 12 tasks in one session. Conversation samples were not collected from the comparison group due to issues in eliciting comparable data. Juan’s samples were collected from spontaneous conversations between us. In order to elicit comparable vocabulary and difficulty levels from the comparison group, it would be necessary to use
prompts, which would result in a monologic rather than dialogic response and therefore would not be a valid comparison.

4.2 Data Elicitation Stimuli

4.2.2.1 Frog Stories. Narratives were elicited from Juan from each of the following books, commonly referred to as frog stories: *A Boy, a Dog and a Frog* (Mayer, 1967), *Frog Where Are You?* (Mayer, 1969), *A Frog on His Own* (Mayer, 1973), *One Frog Too Many* (Mayer, 1975) and *A Frog Goes to Dinner* (Mayer, 1974). Two were repeated by Juan one year after their first elicitation, *A Boy, A Dog and a Frog* (Mayer, 1967) and *One Frog Too Many* (Mayer, 1975) (see Table 2 for dates). *A Frog Goes to Dinner* (Mayer, 1974) was used with the comparison group.

The frog stories are children’s picture books that show the adventures of a boy and his pets, one of which is a frog. They have been successfully used in studies regarding L1 development (Berman & Slobin, 1994), cross-linguistic differences in motion verbs (Pavlenko, 2010; Slobin, 2004), and lexical proficiency (Bulté et al., 2008), and many others. Of all the elicited oral tasks, the frog stories are the least cognitively complex in that the subject and vocabulary are childlike and the storyline is predictable. As Pavlenko (2008) points out, a cartoon book does not require participants “to imagine and thus reduces the cognitive load imposed by the task” (p. 312). Additionally, by incorporating a book series, the topic of each story is similar enough to compare data. While Juan may have experienced some practice effects, the differing plotlines and
minimum of two months between narrations should have reduced those effects. Juan and the comparison group participants were asked to describe the books as they were shown to them with no pre-planning. Additionally, they were encouraged to relax, speak naturally, and no time pressure was applied.

4.2.2.2 Mr. Bean videos. A second set of narratives was elicited with the following four videos starring a popular British comedian, known as Mr. Bean: *Mr. Bean - Dancing at a Nightclub* (Mr. Bean, 2010b), *Mr. Bean at the Swimming Pool* (Mr. Bean, 2009b), *Halloween with Mr. Bean* (Mr. Bean, 2010a), and *Mr. Bean at the Dentist* (Mr. Bean, 2009a) (see Table 2 for dates). *Mr. Bean at the Swimming Pool* (Mr. Bean, 2009b) was used with the comparison group.

In each of the videos, Mr. Bean finds trouble in an ordinary situation. Videos were chosen because they tend to elicit narratives that are more “adult-like, less artificial, and similar to spontaneous narratives” (Pavlenko, 2008, p. 312). The videos were muted to avoid any interference from language or background sounds. Juan and the comparison group participants watched the muted video first, then added their narrative as they watched it a second time with the sound still muted. Thus, this is the only task that allowed an aspect of pre-planning. In contrast to the frog stories, the participants were forced to add narration quickly in order to progress at the same rate as the events in the videos, increasing the task’s time pressure. Yuan & Ellis (2003) showed that on-line planning in the form of unlimited time, resulted in increased accuracy and complexity
among L2 learners. Since adding narration to continuous videos limits on-line planning, the task is arguably more difficult than the frog stories. As Skehan & Shum (2014) explain, “As the video is running the speaker is exposed to a considerable amount of input which has to be understood and then repackaged as production” (p. 190). Videos were chosen due to their previous success in studies of motion verbs (Pavlenko, 2010) and task-based learning research (Foster & Tavakoli, 2009; Skehan & Foster, 1999; Skehan & Shum, 2014).

4.2.2.3. Picture descriptions. The next task involved describing a set of photographs and cartoons in a PowerPoint presentation. Juan completed three sets (see Table 2 for dates). Set one contained seven photos and one drawing; set two contained six photos and two cartoons; and set three contained nine photos and one cartoon (see Appendix A for the photos and cartoons). As part of the original experiment design, some of the photos were repeated to analyze development over time. For this study, only the first attempt at the description was used. The control group completed set three.

The images, copied from photos and cartoons on the internet, show the result of an event (e.g., a broken glass) or a reciprocal action (e.g., hugging each other). The pictures were placed in a PowerPoint presentation which was advanced by the researcher. The participants were given an unlimited amount of time to complete the task. In addition to describing the event, they were asked to describe what might have caused the event. The images were chosen due to their causative or reciprocal nature. Since speakers of
Spanish may use the clitic *se* to express causatives and reciprocals (Toth & Guijarro-Fuentes, 2013; Whitley, 2002), this task may have increased the difficulty level for Juan in that *se* cannot be directly translated into English. Though not the focus of this study, it is noteworthy.

4.2.2.4 Spontaneous Conversation. At our first meeting, Juan mentioned that he would like to practice conversation. Therefore, conversation was a primary part of every session. We discussed a wide variety of topics including his studies, his home, his future plans, his struggles adjusting to the culture, his experiences learning the language, friends he made, and problems he was having in his daily life. The topics chosen for the eight analyses (two samples from each conversation) include the following: his country, his hometown, a problem he had with his roommate, his mother’s career, the idea of having a mentor, how he composes a piece of music, politics in Colombia, and thinking in English (see Table 2 for dates). The samples were chosen from four conversations spread evenly throughout the study (approximately 4-5 months apart). The exact selections were chosen from speech that had few interruptions and only discussed one topic. These selections should limit the introduction of new words (which would bias the lexical diversity measures) and limit dialogic interruptions.
Table 2
Task List

<table>
<thead>
<tr>
<th>Type of Task</th>
<th>Name of Task</th>
<th>Date Collected</th>
<th>Month</th>
<th>Month+task name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frog stories</td>
<td>A Boy, a Dog and a Frog (Mayer, 1967)</td>
<td>2/25/2013</td>
<td>1</td>
<td>1BDFBDF</td>
</tr>
<tr>
<td></td>
<td>One Frog Too Many (Mayer, 1975)</td>
<td>4/15/2013</td>
<td>2</td>
<td>2OFTM1OFTM2</td>
</tr>
<tr>
<td></td>
<td>Frog Where Are You? (Mayer, 1969)</td>
<td>7/10/2013</td>
<td>5</td>
<td>5FWAY1FWAY2</td>
</tr>
<tr>
<td></td>
<td>A Frog on His Own (Mayer, 1973)</td>
<td>10/04/2013</td>
<td>8</td>
<td>8FOHO1FOHO2</td>
</tr>
<tr>
<td></td>
<td>Frog Goes to Dinner (Mayer, 1974)</td>
<td>2/03/2014</td>
<td>12</td>
<td>12FGTD1FGTD2</td>
</tr>
<tr>
<td></td>
<td>A Boy, a Dog and a Frog (Mayer, 1967)</td>
<td>2/25/2014</td>
<td>13</td>
<td>13BDF3BDF4</td>
</tr>
<tr>
<td></td>
<td>One Frog Too Many (Mayer, 1975)</td>
<td>4/18/2014</td>
<td>14</td>
<td>14OFTM3OFTM4</td>
</tr>
<tr>
<td>Mr. Bean videos</td>
<td>Mr. Bean – Dancing at a Nightclub (Mr. Bean, 2010b)</td>
<td>3/11/2013</td>
<td>1</td>
<td>1DISCO1DISCO2</td>
</tr>
<tr>
<td></td>
<td>Mr. Bean at the Swimming Pool (Mr. Bean, 2009b)</td>
<td>9/04/2013</td>
<td>7</td>
<td>7SWP1SWP2</td>
</tr>
<tr>
<td></td>
<td>Halloween with Mr. Bean (Mr. Bean, 2010a)</td>
<td>11/06/2013</td>
<td>9</td>
<td>9HORROR1HORROR2</td>
</tr>
<tr>
<td></td>
<td>Mr. Bean at the Dentist (Mr. Bean, 2009a)</td>
<td>5/23/2014</td>
<td>15</td>
<td>15DENTIST1DENTIST2</td>
</tr>
<tr>
<td>Conversation</td>
<td>Bogota (his hometown)</td>
<td>2/25/2013</td>
<td>1</td>
<td>1BOGOTABOGOTA</td>
</tr>
<tr>
<td></td>
<td>Colombia (in general)</td>
<td>7/24/2013</td>
<td>5</td>
<td>5COLOMBIA5MOTHER</td>
</tr>
<tr>
<td></td>
<td>Roommate Issues</td>
<td></td>
<td></td>
<td>5ROOMATE5MOTHER</td>
</tr>
<tr>
<td></td>
<td>Mother’s career</td>
<td></td>
<td></td>
<td>5MOTHER5MOTHER</td>
</tr>
<tr>
<td></td>
<td>A mentor in their field of study</td>
<td>12/03/2013</td>
<td>10</td>
<td>10MENTOR10COMPOSE</td>
</tr>
<tr>
<td></td>
<td>His process of composing music</td>
<td>5/02/2014</td>
<td>15</td>
<td>15POLITICS15THINK</td>
</tr>
<tr>
<td></td>
<td>Colombian Politics</td>
<td></td>
<td></td>
<td>15POLITICS15THINK</td>
</tr>
<tr>
<td></td>
<td>Thinking in English</td>
<td></td>
<td></td>
<td>15POLITICS15THINK</td>
</tr>
<tr>
<td>Picture descriptions</td>
<td>Picture description 1</td>
<td>7/10/2013</td>
<td>5</td>
<td>5SET1SET2</td>
</tr>
<tr>
<td></td>
<td>Picture description 2</td>
<td>10/08/2013</td>
<td>8</td>
<td>8SET2SET2</td>
</tr>
<tr>
<td></td>
<td>Picture description 3</td>
<td>12/27/2013</td>
<td>11</td>
<td>11SET3SET2</td>
</tr>
</tbody>
</table>

Note: Juan’s first session was February 25th. Thus, months begin on the 25th of each month.
4.3 Data Coding and Calculations

4.3.1 Sample Preparation

The starting points for each sample were chosen from the middle of the samples since participants were more likely to speak naturally once they are comfortable with the task (Ellis & Barkhuizen, 2005; Polat & Kim, 2014). Since adding words to a sample tends to increase the tokens (the number of words per sample) but not the type (the number of unique words), 100 word samples were chosen to increase the accuracy of the lexical diversity measures. Even though the lexical diversity measure $D$ should account for text length differences, longer texts may include more topics, which would most likely affect the $D$ scores. Moreover, recent studies have shown that $D$ may be sensitive to text length as well (deBoer, 2014). Therefore, two samples of approximately 100 words from each task with disfluencies pruned and with contractions expanded were selected from each transcript. The choice for sample size and exclusion of disfluencies and contractions were based on previous studies (Chan et al., 2015; Polat & Kim, 2014; Spoelman & Verspoor, 2010). Two 100 word samples were taken from each task to increase the amount of data collected. Since the measures for subordination, elaboration and grammatical inaccuracy may be unclear and inaccurate if the full clause was not included, the sample size may vary slightly. As done in previous studies (Polat & Kim, 2014; Spoelman & Verspoor, 2010), 10% variation was allowed from the 100-word limit (95-105 words) in order to end the sample with a completed clause.
In the complexity analysis, disfluencies were deleted from the word count because the scores’ validity would be compromised by commonly repeated filler words and redundancies used as fillers (Duran, Malvern, Richards & Chipere, 2004). Though Duran et al. (2004) were referring to lexical diversity, the argument applies to the other measures of complexity. A text including many disfluencies will appear to be syntactically complex in the form of elaboration (words per AS-unit) in that filler words or redundancies are repeated two or three times. Thus, the text is not elaborate; it simply contains more words. Disfluencies were omitted from the analysis without changing the original transcript by coding each disfluency using Codes for the Human Analysis of Transcripts (CHAT) formatting and running a command (+r6) in the Computerized Language Analysis (CLAN) program (MacWhinney, 2000). However, the contractions had to be expanded by hand and saved as separate text files for the complexity analyses. The contraction expansion was necessary for CLAN to properly lemmatize the data and provide an accurate lexical diversity analysis. The original transcripts including the disfluencies were used in the fluency and accuracy analyses. Next, the transcripts were divided in AS-units.

4.3.1.1. Unit of Analysis. As children acquire language, they move from the one-word phase to the two-word phase, and slowly progress to increasingly complex sentences. In the 1970s, L1 acquisition researchers agreed on MLU (Mean Length of Utterance) as a measure of developing vocabulary and syntactic complexity (Brown, 1973). MLU, which
measures the number of words or syllables spoken between pauses, proved an efficient measure for developing children. However, it proved to be ineffective with adult L2 learners, who are more cognitively mature than L1 language learners and thus able to produce utterances of more than a few words in the beginning stages of L2 development (Larsen-Freeman & Strom, 1977). Moreover, pauses in L2 speech are not necessarily unit boundaries since L2 learners pause to search for lexicon, morphology, or grammatical structure more often than L1 speakers (de Bot, 1992; Segalowitz, 2010). Therefore, rather than only considering the amount of speech, L2 complexity studies shifted to considering the types of structures and diversity of vocabulary being produced. Thus, new units of analysis were developed.

At first, the units mainly fell into three areas: syntactic (grammar based), semantic (meaning based) or intonational (tonal based) units. Of these, syntactic units have proven to be the most reliable and objective, and consequently the most utilized (Ellis & Barkhuizen, 2005; Foster, Tonkyn & Wigglesworth, 2000). The unit is generally incorporated as the denominator of a ratio in a global measure (words/unit) or specific measure (perfect aspect/verb unit). T-units, or terminal units, were among the first to be developed (Larsen-Freeman & Strom, 1977). A t-unit includes an independent clause and its dependent clauses. In written data, t-units have been effective for meaningfully dividing text (Bardovi-Harlig, 1992; Bulté & Housen, 2014; Chan et al., 2015; Larsen-Freeman, 2006). However, oral data produces complications for t-units, such as interruptions, one word responses, confirmations, and incomplete thoughts.
An AS-unit, or analysis of speech unit, was developed to resolve these issues (Ellis & Barkhuizen, 2005; Foster et al., 2000; Norris & Ortega, 2009). AS-units contain an independent clause or a subclausal unit plus any subordinate clauses. Subclausal units are comprised of incomplete sentences, one or two word responses to questions, and sentences that can be elaborated into complete sentences. For example, the following sentence incorporates all three structures. “Why yes (subclausal unit), I would like to go (independent clause), if it’s not any trouble (subordinate clause)”.

Thus, where a t-unit might be lengthened by a subclausal unit such as agreement (why yes), confirmation (okay), and clarification phrases (right?), an AS-unit will be unaffected since these subclausal units create their own AS-unit. AS-units have been successfully incorporated in recent task-based learning and CDST research (Foster et al., 2000; Foster & Tavakoli, 2009; Polat & Kim, 2014). Since the present study focuses on oral speech, AS-units were chosen rather than t-units.

Lastly, each sample was coded for three measures of complexity, three measures of accuracy, and eleven measures of fluency (see Table 3 below). Each of these variables and their definitions will be explained in the next section.
Table 3
Analytical Variables

<table>
<thead>
<tr>
<th>Construct</th>
<th>Measure</th>
<th>Global or specific measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity (lexical diversity)</td>
<td>D</td>
<td>Global</td>
</tr>
<tr>
<td>Complexity (elaboration)</td>
<td>Mean words per AS-Unit</td>
<td>Global</td>
</tr>
<tr>
<td>Complexity (subordination)</td>
<td>Mean clauses per AS-Unit</td>
<td>Specific</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Mean errors per AS-Unit</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Errors with Past Tense per 100 words</td>
<td>Specific</td>
</tr>
<tr>
<td></td>
<td>Auxiliary Verb Omission per 100 words</td>
<td>Specific</td>
</tr>
<tr>
<td>Fluency (breakdown disfluencies)</td>
<td>Frequency of all pauses</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Frequency of silent pauses</td>
<td>Specific</td>
</tr>
<tr>
<td></td>
<td>Frequency of filled pauses</td>
<td>Specific</td>
</tr>
<tr>
<td>Fluency (repair disfluencies)</td>
<td>Frequency of all repair disfluencies</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Frequency of false starts (new train of thought)</td>
<td>Specific</td>
</tr>
<tr>
<td></td>
<td>Frequency of reformulations (self-corrections)</td>
<td>Specific</td>
</tr>
<tr>
<td></td>
<td>Frequency of redundancies (repeating sound, word, or phrase)</td>
<td>Specific</td>
</tr>
<tr>
<td>Fluency (speed)</td>
<td>Speech Rate (syllables per second)</td>
<td>Global</td>
</tr>
<tr>
<td></td>
<td>Mean length of silent pauses</td>
<td>Specific</td>
</tr>
<tr>
<td></td>
<td>Articulation Rate (syllables per second not including silent pauses)</td>
<td>Specific</td>
</tr>
<tr>
<td>Fluency (native-like speech)</td>
<td>Colloquial English forms</td>
<td>Specific</td>
</tr>
</tbody>
</table>

4.3.2 CAF Variable Coding

As indicated in Table 3, variables were classified as either global or specific. A global variable represents the construct as a whole, whereas a specific variable represents an aspect of the construct that is of particular interest within the global construct. Specific variables are often chosen due to their use in previous literature or links to development. SLA researchers often consider both global and specific accuracy variables in their
analysis of language development since the number of errors does not necessarily consider the developmental process of experimenting with grammar forms or lexicon. For example, Polat & Kim (2014) considered both the total number of errors (global) and errors related to the third person singular present tense (specific).

In this study, I evaluated the variables used in previous literature and selected the most common ones in order to provide useful comparisons. However, some of the common global variables were conflations of the specific variables. Therefore, in the development analysis, all variables were considered to compare specific measures to global measures. For the comparison between constructs, only global variables were considered due to Skehan and Foster’s (1999) argument that they are more effective for analyzing cognitive strain.

4.3.2.1 Complexity. Three complexity measures were chosen to provide a detailed description of Juan’s complexity changes: $D$ (lexical diversity), words per AS-unit (elaboration), and clauses per AS-unit (subordination). The coding for each will be described in detail in their corresponding sections below.

*Lexical Diversity.* Lexical diversity refers to the amount of unique vocabulary spoken in the sample. For this study, it will be measured with $D$ (McKee et al., 2000). Once the texts were pruned and contractions expanded, $D$ was calculated by CLAN.
D was chosen for its accuracy, ability to account for text length differences, and frequency in current literature (Jarvis, 2002; Malvern & Richards, 2002; McKee et al., 2000; Skehan, 2009). D was designed to correct TTR’s sensitivity to text-length (Malvern & Richards, 2002; McKee et al., 2000). Since D is calculated with random words from the entire text, it is not as sensitive to topic shifts. Jarvis (2002) found D to be one of the most accurate lexical diversity measures in that D closely models the TTR curves of the original text. Moreover, Skehan (2009) successfully used D in task-based learning studies and Bulté et al., (2008) and Bulté and Housen (2014) showed that D increased as language developed in their participants. In this study, CLAN was used to calculate D for all of Juan’s samples and the control group samples. The process for calculating TTR curves and D is as follows:

1. A curve is plotted using the TTR of the first word (N=1), then the TTR of first two words, and so on sequentially until the TTRs for the entire text has been plotted as a TTR curve. Thus, a 500-word text, would have 500 separate TTR scores plotted onto a graph. In Figure 2 below, a TTR curve for the following sentence is plotted: “The girl took the candy from the store”. Thus, the first point represents 1 unique word divided by 1 total word, which equals 1. Once a word is repeated (the), the score drops. For the fourth data point, there are three unique words divided by four total words since “the” is repeated. Thus, the score is .75. The score drops again for the seventh data point when “the” is repeated again. Now there are five unique words divided by seven total words equaling .71.
2. Once the TTR curve is complete, the program *vocd* (available in CLAN) creates a hypothetical curve that is compared to the original TTR. This hypothetical curve is created by choosing a random sampling of words for that point on the graph (35 random words rather than words 1 – 35). In the example above, the line could vary considerable depending on the random selection of “the”. If “the” were selected as the first three words, the scores would drop from 1 to .5 to .3 for the first three points. On the other hand, if *the* is not selected until the last three points, the TTR curve would remain at 1 until the 6th word. To account for a repeated word being selected over and over, the random draw for each point is repeated 100 times and the average is plotted on the graph. This process is only conducted on points 35-50. Thus, the text must have at least 50 words in order to calculate the hypothetical curves with *vocd*. 

Figure 2. Example of a TTR curve for a single sentence.
3. Next, the random selection process and procedures listed in Step 2 are repeated two more times resulting in *vocd* creating three hypothetical lines that will be used to calculate *D* (see Figure 3). The hypothetical curves represent the highest and lowest possible lexical diversity for the given text (deBoer, 2014). The line with the highest TTR scores (*vocd TTR1* on Figure 3) represents the most lexically diverse random sample. The line with the lowest TTR scores (*vocd TTR 3* on Figure 3) represents the least lexically diverse random sample.

![Figure 3. Representation of the original TTR Curve Chart and the theoretical TTR Curves produced by *vocd*](image)

4. Next, *D* is calculated for each hypothetical curve using the following equation:

\[ TTR = \frac{D}{N} \left[ 1 + 2\left( \frac{N}{D} \right) - 1 \right] \]

In this equation, *N* equals the number of words per text;
TTR represents the TTR for the hypothetical curve; and $D$ represents a parameter that is adjusted until the predicted TTR is as close as possible to the actual TTR. To calculate $D$, an algorithm is run that replaces the $D$ value in the equation until the hypothetical TTR is as close as possible to the actual TTR.

5. Lastly, the three $D$ scores from the hypothetical curves are averaged together producing $D_{\text{optimum}}$. $D_{\text{optimum}}$ is commonly referred to as $D$ in the literature (DeBoer, 2014) and is the final score in the CLAN output for $\text{vocd}$. Thus, $D$ is “defined by the position of the curve relative to ideal curves between the theoretical minimum and maximum” (DeBoer, 2014).

*Elaboration.* As the next measure of complexity, elaboration, was chosen as an indicator of increasingly complex grammatical forms within a sentence. Increases in the measure suggest that speakers or writers are using more detailed descriptions including adjectives, adverbs, prepositional phrases, nominal phrases, and subordinate clauses (Norris & Ortega, 2009; Pallotti, 2009). Previous studies have measured elaboration with words per t-units in writing (Bulté & Housen, 2014; Chan et al., 2015) and words per AS-units with speech (Polat & Kim, 2014). Since this study analyzes oral data, elaboration was calculated by measuring the number of words per AS-unit.

To calculate elaboration, each AS-unit was separated by a carriage return in the transcripts. By doing so, the MLU (mean length of utterance) command in CLAN was able to compute the words per AS-unit and then calculate the average words per AS-unit
for the sample, which measures elaboration. As often happens when manipulating a program to do something other than it was designed to do, an issue arose that required the transcripts to be altered slightly. In order to maintain the AS-units, interruptions in the middle of an AS-unit (generally “uhhuh” or “mhmm”) were moved until after the AS-unit was complete. If I had not done so, the elaboration analysis (words per AS-unit) would be inaccurate. Notes were added to the bottom of each transcript when changes were made in order to maintain their integrity.

Subordination. Subordination is measured by the mean clauses per AS-unit. Increases in subordination indicate the learners’ ability to produce complex sentences including relative and subordinate clauses. Subordination has been shown to increase with proficiency (Chan et al., 2015) and with narrations describing simultaneous tasks (Foster & Tavakoli, 2009).

Even though subordination is included in elaboration (words per AS-unit), subordination was chosen as a specific measure for this study due to its prevalence in current literature (Bulté & Housen, 2014; Chan et al. 2015; Foster & Tavakoli, 2009) and the emergence of subordinate clauses noted in the initial analyses of Juan’s transcripts. Subordination was calculated by hand using the CLAN output from the elaboration analysis. The output contains the pruned version of the sample and carriage returns after each AS-unit. The clauses (independent, subclausal, and subordinate) were then separated
by a “::” as indicated by Foster et al. (2000). They were then counted for each AS-unit and the average clauses per AS-unit was calculated.

Since this analysis is subject to researcher error, a second coder confirmed the AS-unit separation and clause markers in all of Juan’s data and in 10% of the comparison group’s data. The amount of 10% was selected based on Foster and Tavakoli’s (2009) inter-rater reliability standards. If the raters disagreed, all differences were discussed until agreement was reached as was done in Larsen-Freeman (2006) and Polat and Kim (2014).

4.3.2.2 Accuracy. For the accuracy analysis, three quantitative variables (errors per AS-unit, past tense errors per 100 words, and auxiliary verb omissions per 100 words) were coded. Errors in the transcripts were identified and coded by the researcher and two additional raters who completed their analysis separately from each other and the researcher. Selections were based on native speaker intuition. In order to determine how an error should be counted, the researcher and the two inter-raters rewrote the sentence containing the error while maintaining as much of the original structure as possible and each change was counted as an error. However, there was often more than one option to correct a sentence. Differences were discussed until a consensus was reached. This methodology was suggested by Ellis & Barkhuizen (2005) and implemented in Polat & Kim (2014).
Once all raters agreed, the errors were marked in CHAT format on the transcript so that the frequency of errors for each transcript could be calculated by CLAN (MacWhinney, 2000). CLAN’s analysis provided a frequency count for each error type and a total of errors per transcript. By using CHAT’s codes, consistent and established error types were selected, which adds validity to the error labels. The categories for this study include: morphological errors (e.g., improper formation of verbs or pluralization), semantic errors (e.g., improper noun or preposition choice), omissions (subdivided by part of speech), errors related to formal lexical devices (articles), and sentence structure. For a full list of the sub-category errors found in the participants’ samples, see Appendix B. The total number of errors per transcript was then divided by the total number of AS-units per transcript to provide the global measure errors per AS-unit.

It should be noted that one change could involve two errors (e.g., semantic and grammatical) and therefore such a change was counted as two errors. For example, in the sentence “he expulse (. ) from the swimming pool” (Juan, 7SWP1, September 4, 2013), was rewritten as “he kicks him out from the swimming pool”. Thus “expulse” was coded for omission of the 3rd person singular “s” and a semantic error.

Global accuracy has been questioned as a valid measure in that learners at different levels produce different types of errors (Lambert & Kormos, 2014; Pallotti, 2009). Since learners are constantly expanding their repertoire of forms by experimenting with new forms (and making errors), a simple frequency error count does not truly account for development. As Ellis and Barkhuizen (2005) point out, error production is
“not just a matter of using a deviant form or a target-language form but also selecting from a number of different deviant forms” (p. 55). Therefore, specific accuracy measures were also included. Once the data had been coded, CLAN produced an output of errors by type, which was transferred to an Excel sheet (see Appendix C). From that Excel sheet, two specific accuracy measures, errors related to the past tense and omission of auxiliary verbs, were chosen due to increasing fluctuation followed by decreased errors which may indicate a phase shift. They were then analyzed qualitatively by referring to the original transcripts.

4.3.2.3. Fluency. Due to the diversity of definitions, fluency studies differ considerably on the variables they include. This study chose to emphasize fluency by including 11 quantitative variables for three reasons. First, oral fluency appears to be an under-researched variable in CAF comparisons and CDST studies, most likely due to its status as a variable of performance rather than L2 knowledge (Segalowitz, 2010; Skehan, 2003). Second, Juan was a study abroad student, and research has shown that fluency is affected more than the other CAF constructs in the study abroad context (DeKeyser, 2014; Segalowitz & Freed, 2004). Lastly, even though all of the variables included have been utilized in other research, there is no consensus concerning which are the best measures of fluency. Therefore, this study contributes to the literature in CAF, CDST, and study abroad. The 11 variables were separated into four types of fluency based on Skehan (2003): speed fluency (speech rate, length of silent pauses, articulation rate);
breakdown fluency (frequency of silent pauses, frequency of filled pauses, and total pauses); repair disfluency (redundancies, self-corrections, false starts, and total redundancies); and native-like speech (colloquial English forms). The last variable was chosen based on study abroad research which indicates that increased use of colloquial forms may result in formulaic constructions and memorized exemplars (Kinginger, 2008; Towel et al., 1996). Such forms were shown to increase fluency by allowing students to increase the amount of speech they produce and the speed in which they produce it (Towell et al., 1996).

*Speed Fluency.* For speed fluency, three variables were calculated: speech rate (syllables per second), length of silent pauses, and articulation rate (syllables per second minus silent pauses). To calculate the silent pauses, speech rate, and articulation rate, two audio analysis programs were used: Audacity Audio Editor (Mazzoni, 2016) and Praat version 6.0.15 (Boersma & Weenink, 2016).

After the audio files were clipped to match the 100 word transcripts in Audacity Audio Editor, and then the silent pauses were marked using Audacity’s Silence Finder command (Mazzoni, 2016). In this study, silent pauses are defined as anything longer than .4 seconds and less than 30 decibels in volume. The number of seconds was based on previous studies (Tavakoli & Foster, 2011; Towell et al., 1996), whereas the decibel level was chosen based on trial and error in pilot analyses. It was found that 30 decibels might include background noise, but generally did not include the speaker’s voice. Since
Silence Finder only marked the ends of the pauses, it was necessary to listen to each sample and mark the pause length manually. Once marked, Audacity outputs a report containing the start and end time for each silent pause, which was then exported to an Excel sheet and the length of each silent pause was calculated. The pause lengths were then averaged to produce the mean length of silent pause variable.

Next, the seconds per samples were recorded in order to calculate the speech rate and articulation rate. Since the interlocutor should not interfere with the participants’ scores, the interlocutor’s speech and the time to turn the pages in the frog stories were labeled and deleted from the sample before recording length of the sample. Two sample lengths were recorded, the length of the pruned sample (as described above) and the length of the pruned sample minus the combined length of the silent pauses. The former was used to calculated speech rate, while the latter was used to calculate articulation rate.

Next, the pruned samples (minus interlocutor and page turns) were imported into Praat (Boersma & Weenink, 2016). Praat then calculated a syllable per sample analysis using a programming script produced by de Jong & Wempe (2009). The number of syllables per transcript was calculated by counting phonetic peaks in the transcript. The peaks are based on the phonetic intensity. The script calculates the peaks (generally vowels) and valleys (generally consonants) and counts the peaks. Since vowels are generally the base of a syllable and have more intensity, the count is able to estimate the number of syllables in a sound file. However, de Jong and Wempe (2009) note that some slurred or unaccented syllables may be missed by this analysis. They propose that
“missing such unprominent syllables is not problematic for researchers who want to compare speech rate between speakers and tasks, since the underestimation of speech rate is consistent” (p. 389). As a precaution, five samples were coded by hand and found to have no more than three syllables difference per transcript between the Praat script output and the hand count in samples of 100 – 200 syllables.

For the first variable, speech rate, the syllables per sample were divided by the second per sample (including silent pauses). Speech rate provides a global measure in that it includes both articulation rate (speaking time) and silent pauses. Speech rate has been shown to impact perceptions of fluency (Bosker et al., 2014b; Kormos & Dénes, 2004) and to increase with development (Towell et al., 1996).

The next measure, articulation rate, was calculated in order to distinguish between speech slowed by lengthy pauses and speech slowly produced by the speaker. Though both speech rate and articulation rate indicate strain on the cognitive resources, articulation rate reduces the conflation between breakdown and speed fluency (Bosker et al., 2012). In order to avoid such conflation, Rossiter (2009) removed all disfluencies (repairs and filled pauses) to correlate speed with native-speaker perception. A similar pruning was considered for this study. However, since most of Juan’s transcripts contained many disfluencies ($m = 8.14$ per 100 words), his samples would have been manipulated considerably, and therefore the analysis would not be representative of his speech. Moreover, repair disfluencies tend to be noticed less by native-speakers (Kormos & Dénes, 2004), which may indicate that they are not produced at a different speed than
regular speech. In order to calculate the articulation rate, the total seconds per silent pauses was subtracted from the seconds per sample. Then, the number of syllables were divided by the seconds per sample (minus silent pauses).

**Breakdown Disfluency.** The breakdown disfluency measures for this study include silent pauses, filled pauses, and a global variable of the total pauses. Of all the fluency constructs, breakdown fluency is included the most often, including all of the studies reviewed in this dissertation. In both L1 and L2 speakers, both filled and silent pauses may indicate that the speaker is experiencing cognitive strain (Foster & Tavakoli, 2009; Skehan, 1996; Segalowitz, 2010) or pausing between grammatical phrases (Kormos & Dénes, 2004). In relation to L2 fluency, the speakers may also be searching for a word, the correct morphological or syntactic formation, a translation of an untranslatable word or phrase, or they may be trying to apply declarative knowledge to their speech.

Silent pauses, as defined above, were coded into the transcripts by adding the standard CHAT symbol “(. )” in the locations determined by the Audacity analysis. Filled pauses were defined as audible sounds that that have no lexical definition, such as “um” or “uh” (Bosker, Pinget, Quené, Sanders & de Jong, 2014a; Kormos & Dénes, 2004). These were included in the original transcription, then marked with the standard CHAT symbols “&” (MacWhinney, 2000). After the transcripts were coded, CLAN (MacWhinney, 2000) calculated the frequency of each. The totals for each sample were
copied and pasted into an Excel spreadsheet in which they were totaled for the variable of total pauses.

*Repair Disfluency.* Repair disfluency includes the frequency of three specific variables: redundancies, reformulations, and false starts (Ellis & Barkhuizen, 2005) and a global variable, total repairs, in which the frequencies of all three were added together. Redundancies, referred to as repetitions in some literature, include repeated words, phrases, or phonemes (Bosker et al. 2012; 2014a; 2014b; Kormos & Dénes. 2004). Reformulations, or self-corrections, were defined as stopping in the middle of a word or phrase, then correcting it (Ellis & Barkhuizen, 2005; Lennon, 1990; Skehan & Foster, 1999). For example, “I think I thought that …” would be a reformulation whereas “I think I think that …” was recorded as a redundancy. Lastly, false starts are defined as utterances or sentences that are not complete (Ellis & Barkhuizen, 2005; Skehan & Foster, 1999). In contrast to fragments, which were defined as a phrase not including a finite verb, false starts indicate that the current idea was abandoned. For example, “I am leaving for … I need to pack” was coded as false start whereas “leaving for Canada tomorrow” was coded as a subject and auxiliary verb omission. The transcripts were coded for each type of repair disfluency in standard CHAT format and therefore the specific variables’ frequencies could be calculated by CLAN (MacWhinney, 2000). These numbers were transferred to an Excel sheet where the total disfluencies variable was calculated.
Colloquial Speech. Colloquial speech was defined as vocabulary, grammatical formations, or sentence structures that are not prescriptively correct, but were found in the comparison group’s samples and Juan’s later samples. Kinginger (2008) explains the value of colloquial speech in that “they serve to enhance the learner’s communicative repertoire” (p. 16) thus reducing the cognitive strain of searching for lexicon. Similar conclusions were made by Towell et al., (1996) regarding memorized exemplars of colloquial speech. As such, they were coded as a fluency variable. For this study, three types were chosen to be analyzed. They include “gonna” or “wanna” (grammar reductions), “cuz” (shortening of because) and a direct quote preceded by “like” as in “He was like I don’t think so”.

4.4 Data Analysis

The analyses described in this section were conducted after transcripts of the data listed in Table 2 were coded as indicated in Table 3. The section below describes how the data were analyzed for second language development, global variables interaction, environmental factors, and task effects.
4.4.1 Second Language Development Analysis

To investigate Juan’s language development, three types of analysis were conducted: emerging CAF changes, periods of fluctuation and phase shifts, and specific error analyses. This section will explain each in detail.

As is often done in CDST studies, graphs of developmental trajectories were created from each of Juan’s CAF variables in a similar manner to the graphs in Larsen-Freeman (2006), Polat and Kim (2014), and van Geert and van Dijk (2002). Since significant differences were found between the four types of tasks, the CAF scores for each task had to be converted to a measure that would make them comparable. For the first analysis, 0 – 1 values were chosen. Verspoor et al., 2011) explain that “values can be recalculated to values from 0 – 1 so that originally different kinds of values and interactions of different constructs can be visualized and compared to each other more easily” (p. 161). For example, Chan et al (2015) used 0 – 1 values to compare mean length of unit, degree or subordination, and degree of coordination. 0 – 1 values were calculated in Excel with the following formula: (score – minimum score) / (maximum score – minimum score) (Verspoor et al., 2011). The results place the high score at one, the low score at zero and the remaining variables on a scale between zero and one. Once the scores for each task were converted separately, they were combined and placed onto a graph in Excel. Thus, each task had four scores of one and four scores of zero, lessening the effects of task differences.
In order to investigate gradually emerging trends, polynomial trendlines were added to trajectory graphs as was done by Polat & Kim (2014) and Chan et al. (2008). Trendlines mathematically smooth the data into straight or curved lines allowing trends to be seen more easily. Polynomial rather than linear trendlines were chosen to show curves which capture changes in the direction of the trend.

The 0 – 1 value conversions were also completed for the comparison group. From that data, 95% confidence interval ranges were calculated and placed on the graphs of Juan’s data. The comparison group data were used to evaluate if Juan moved toward his speech communities’ standards. Since speech rates, colloquial speech, and non-standard grammar forms vary throughout the United States (Lippi-Green, 2012), establishing Juan’s community’s standards was helpful understanding how immersion might have affected his scores. Since Juan did not receive explicit instruction, his speech community provided his primary source of examples and interaction. 0 – 1 values were selected over z-scores due to their usage in previous studies (Chan et al., 2015; Spoelman & Verspoor, 2010; Verspoor et al., 2011) and due to issues calculating confidence intervals using z-scores. Since the comparison group did very well in some measures, their z-scores were fairly low. For example, in grammatical accuracy they ranged from -1.19 to 1.94 errors per AS-unit with one outlier of 4.105. Since there was such little variation, the average z-score was -6.61. However, their average 0 – 1 values, which contained no negative numbers, was .22584. Since the average score is part of the confidence interval formula
(critical value multiplied by the standard deviation plus or minus the mean score), the 0–1 values produced more precise confidence intervals.

Next, the graphs were visually inspected for periods of fluctuations that may indicate phase shifts. When periods of fluctuation were observed, min-max lines were added to the graphs. Min-max graphs show periods of stability and fluctuations in scores by focusing on the range of the scores. When a moving average of the maximum scores and minimum scores are added to the trajectory graphs, the lines frame the data. As such, periods of stability and fluctuation are easily defined. When scores have a large range, the learner is unstable in their usage; when the lines are close together, the learner’s language is stable. Theoretically, fluctuation indicates a period of restructuring. If that period is followed by stability that is different than before, the learner has experienced a phase shift (Larsen-Freeman, 2006; Polat & Kim, 2014).

The min-max lines were calculated in Excel using the “min” and “max” formulas. To see changes in fluctuation, the minimum and maximum score for the first four data points were calculated and placed on the graph, then the second through fifth point, third through sixth and so on. Next, the data were examined for periods of stable production before and after periods of fluctuation. If the before and after periods appear to be different, there is an indication that Juan moved to a new attractor. If periods before and after appear to be similar, then he most likely returned to the previous attractor state and the restructuring was incomplete.
The next developmental analysis examined certain types of errors in Juan’s data that showed periods of fluctuation. First, the frequency count of each type of error was calculated using CLAN. Next, the results were placed on an Excel sheet (see Appendix C) and two specific types of errors were selected due to a period of fluctuation in their frequency over time: morphological errors related to the past tense and auxiliary verb omission. Both error types contain a period with few errors, followed by a period of increased errors, and finally a period with no errors. To investigate the possibility of avoidance, the transcripts were examined for overall usage of past tense and auxiliary verbs for each transcript. Then a ratio of errors by usage was calculated for the three periods. The data were then examined for phase shifts.

4.4.2 Global Variable Interaction

In the CDST framework, systems are in a constant state of change and are affected by those systems around them (Larsen-Freeman & Cameron, 2008). As part of that interaction, the CAF variables may be connected growers meaning that they may interact in competition or cooperation with each other or that one variable may be dependent on another to develop (Verspoor et al, 2011). For this study, the global variables of elaboration, lexical diversity, errors per AS-unit, total pauses, speech rate, and total disfluencies were compared to examine such relationships. Due to their limited scope, specific variables were not compared to specific variables in another construct. For example, a comparison of past tense errors and silent pauses provides less information
than a comparison between total errors and total pauses. While both may indicate a relationship between accuracy and breakdown fluency, the second indicates breakdowns related to all types of errors, which is more indicative of the learner’s overall pattern.

Despite the call from Larsen-Freeman (2006) to continue comparisons between all of the CAF variables, most CDST and CAF research has limited comparisons between the CAF variables to complexity and accuracy (Mulder & Hulstijn, 2011; Polat & Kim, 2014; Spoelman & Verspoor, 2010). Since complexity and accuracy reflect the learners’ L2 knowledge whereas fluency reflects their proceduralization of that knowledge, the choice has theoretical validity. However, task-based learning research has shown that comparisons between all three variables contribute to a more complete understanding of the learner (Ellis & Yuan, 2004; Skehan, 1996, 2003; Skehan & Foster, 1999; Tavakoli & Foster, 2011). To make comparisons with theoretical validity, comparisons began between the global variables that indicate the learner’s language knowledge (lexical diversity, elaboration, and accuracy). Then, each of these variables were compared to the global fluency variables (total pauses, speech rate, and total repairs), which indicated the proceduralization of the knowledge variables and performance interference caused by cognitive strain. Accordingly, the first set of comparisons indicate language development and connected growers whereas the second set of comparisons indicate levels of proceduralization and cognitive strain.

In order to make these comparisons, the variables need to be compatible both meaningfully and numerically. While the 0-1 values equated them numerically, increases
in some variables indicated a positive change while increases in others indicated a negative change. For example, increasing lexical diversity represents a positive change while increasing errors per AS-unit represents a negative change. For these analyses, the scores indicating a negative change were converted to a positive scale. To accomplish the conversion, the raw data were converted to z-scores. Z-scores include both positive and negative scores with most scores clustering between 1 and -1. Therefore, the conversion was accomplished by changing the positive scores to negative and the negative scores to positive. The difference from zero is the same, just in the opposite direction. The following variables were converted in order for an increase to be positive and comparable to the other variables: errors per AS-unit, frequency of all pauses, and frequency of all repair disfluencies.

In addition, z-scores account for more variation within the variables than 0-1 values. While 0-1 values are based on minimum and maximum scores, z-scores are based on standard deviations from the mean. In instances where there is a great deal of variation, z-scores represent the variation with finer measurements than 0-1 values. Moreover, other scholars, such as Larsen-Freeman (2006), chose z-scores to make inter-construct comparisons. As an additional validity measure, graphs with z-scores and graphs with 0-1 values were compared. While the poles of the scales were different, nearly identical patterns were found.

In order to compare the variables, two types of comparisons were made: periods of fluctuation and moving correlations. Periods of fluctuation were analyzed using min-
max graphs to see if certain variables entered periods of fluctuation together or separately, which may indicate a type of cooperative or conditional type of relationship. While min-max graphs have been used to look at fluctuation (Polat & Kim, 2014), comparisons between different variable’s periods of fluctuation do not appear in the literature. Thus, this analysis is experimental in nature. In addition to min-max lines, I looked at differences in the range between areas that seemed stable and areas that appeared to fluctuate. The period of time was determined by visually inspecting the min-max graph.

Moving correlations were first calculated between lexical diversity, elaboration, and errors per AS-unit. If the majority of the data points were correlated, the relationship was labeled as a connected grower. Positive correlations were labeled as cooperative growers whereas negative correlations were labeled as competitive. Next, moving correlations between each of the knowledge variables and the fluency variables were calculated and plotted on a graph. The moving correlations were calculated following the guidelines in Verspoor et al., (2011). To begin, the first four data points for each variable were selected and correlated using the “correl” formula in Excel. Then, another set of four pairs was calculated from data point two through five (overlapping the previous group), three through six, and so on. These correlations were then plotted on a graph. The purpose of this analysis is to see how dynamic relationships develop in correlations over time.
If stable moving correlations were found in Juan’s data, Spearman’s rank correlation were calculated for both Juan and the control group then compared. The purpose of this comparison was to see if the variables interacted differently, which may indicate differences between first and second language speakers.

4.4.3 Environmental Factors

The last analysis was conducted to investigate if events in Juan’s social ecosystem affected his CAF scores. Five sets of factors were compared: a timeline of instrumental and integrative factors mentioned by Juan during our sessions; changes in standardized proficiency exams over time; inclusion of colloquial speech in Juan’s samples that was also found in his speech community; changes in Juan’s CAF scores; and Juan’s own comments regarding his learning and environment. All of the data were then compared qualitatively based on instrumental or integrative factors for learning the language (Masgoret & Gardner, 2003), and Juan’s ideal L2 self or ought-to self (Dörnyei, 2009; Dörnyei & Chan, 2013).

The timeline of motivational factors was created from information taken from session conversation transcripts, the researcher’s notes, general proficiency scores, and essays Juan wrote as part of the study (see Appendix D). The data were separated into three categories. The first category, integrative factors, included events and activities, requests for language assistance with forms he encountered, and comments regarding his attitudes towards native speakers that indicate the quantity of and interest in native
speaker interaction. The second category, instrumental motivation factors, included activities and events that would lead to or prevent achieving his professional goals, such as upcoming exams, active steps toward applying to doctoral programs, and the decisions regarding his future. Lastly, general proficiency scores were also placed on the timeline. The proficiency measures included the Oxford Quick Placement exam (UCLES, 2001a, 2001b) scores (administered four times), TOEFL iBT (Educational Testing Services, 2016a) scores, and the Nation Vocabulary Test (Nation, 2001). These were included as a general measure of proficiency and as a motivational factor since one of his goals was to improve his English.

For colloquial speech, quotes from all of Juan’s analyzed samples were placed on a chart chronologically (see Table 5 in section 5.3). Three types of colloquial speech were chosen, gonna and wanna (grammar reductions), cuz (shortened words), and like followed by a direct quote. These were chosen because they exist in the comparison’s groups speech and appear in Juan’s speech after month 5, which implies he did not already know them. The data were then analyzed for increases and decreases in the form’s usage and compared to Juan’s mentions of native-speaker contact in the motivational factors timeline.

Next, quotes from Juan regarding the above factors and dates were included in the analysis to provide triangulation of the data. Quotes were referred to whenever an event might have affected his motivation, but Juan’s intentions were unclear. While causative claims cannot be made from these analyses, to ignore his social ecosystem altogether
would not present a complete picture of his learning experience. Lastly, all numerical data were analyzed for changes over time which was then compared to the environmental factors timeline to explore possible connections between them.

### 4.4.4 Task Effects Analysis

In task-based learning, variables are analyzed to investigate how tasks affect the cognitive resources available to produce accurate, complex, and fluent speech (Robinson, 2011; Skehan, 2014). The purpose of this analysis was to investigate if the different tasks had a substantial enough effect on Juan’s CAF scores to merit separate comparisons and which tasks encourage or discourage higher complexity, accuracy, and fluency scores.

To conduct this analysis, Juan’s CAF scores were input into SPSS and coded by task modality (frog stories, Mr. Bean videos, picture descriptions, and conversation). Next, a Kruskal-Wallis analysis was conducted comparing each of the variables by task. This same analysis was conducted on the comparison group’s scores for the three tasks they performed and compared to Juan’s data. For ease of analysis, if the majority of the variables in either the comparison group or Juan’s data showed significant differences between tasks, all of the data was separated by task, converted to 0-1 and z-scores, then placed on the trajectory graphs. The Kruskal-Wallis analysis, a non-parametric procedure, was chosen since the data for both groups are not normally distributed. Furthermore, the sample size of Juan’s data was small, containing less than 15 samples
for each task. As an additional measure, task differences were analyzed through descriptive statistics.

In sum, the data will be coded for 17 CAF variables (including two specific errors and colloquial speech). That data will then be separated by task, converted to 0-1 values, recombined, placed on trajectory graphs, and analyzed for emerging development and periods of fluctuation. The second analysis will compare the knowledge variables’ development to each other using moving correlations to look for connected growers. Next, the knowledge variables will be compared to the fluency variables by calculating moving correlations. The fourth analysis will examine motivational factors by placing them on a timeline which will then be compared to the CAF variable’s development. Lastly, task differences will be analyzed using Kruskal-Wallis tests to investigate for statistically significant differences between the frog books, Mr. Bean videos, picture descriptions and conversations.
CHAPTER 5

RESULTS

By analyzing multiple variables in Juan’s data over time, this study contributes to several areas of second language development. First, it follows the development of a second language learner longitudinally and shows that the CAF variables develop non-linearly with some increasing and decreasing together while others change independently. Secondly, the results show interconnected patterns between variables by correlating global CAF variables. Third, the data indicates that motivation and native-speaker exposure are related to Juan’s CAF development. Lastly, it shows the effects of task differences on the CAF variables. The following chapter presents the results from each research question.

5.1 Research Question 1: Development of CAF:

Over 15 months in the US, what changes does Juan make in the areas of complexity, accuracy, and fluency?

Over the 15 months of Juan’s immersion, his ability to express himself in English clearly increased. While the following section will show numerically that most of his complexity, accuracy, and fluency scores improved, the following two speech samples provide an overview of how his language developed. The first sample was taken during our first session when I asked him to tell me about his hometown. The second sample
was taken during his last month while we were discussing his disapproval of a political leader in Colombia (see Appendix E for the complete transcripts).

a) so yeah ten years befo ago was like (. &eh . &eh .) many problems like (. &eh (. &eh &em violence or something. (. but now everything changing and (. we have a lot of people there from everywhere in the world so. (1Bogota, February 25, 2013, Month 1)

b) and the president signed to (. to be the (. the hero of this kind of peace (. and he's trying to do that. I mean &ah come on. (. he want to be nominated for the (. peace Nobel prize. (15Politics, May 2, 2014, Month 15)

In the first sample, he was struggling to find words to express his thoughts as is evidenced by resorting to expressions such as “or something” and by using both filled pauses (&eh) and silent pauses (indicated with (.)) to hesitate while he searched for the correct word or grammar form. Fifteen months later, his vocabulary became more precise, using words such as “hero” and “nominated”. Moreover, he could express a complex opinion more smoothly, even including some colloquial expressions such as “ah, come on”. He also used fewer coordinating conjunctions and began correctly using infinitive clauses, the progressive aspect, and past tenses. Lastly, over the course of our conversations, Juan became increasingly able to express more complex ideas. By December (month 10) he was able to explain the process by which he composes music, a task many people would have difficulty completing in their native language.

For the quantitative evaluation of Juan’s speech, three topics for analysis were chosen: emerging changes, fluctuation, and specific errors analyses. The first examined
gradual increases and decreases in the 14 numerical CAF variables. The second used min-max graphs to analyze fluctuations in the variables that visually appeared to have fluctuating scores. Finally, a grammatical error analysis was performed, which looked at two specific areas in which Juan’s data showed periods of fluctuation: the past tense and auxiliary verbs.

5.1.1 Emerging Changes Analysis

To explore emerging changes in Juan’s data, his raw scores were converted to 0 – 1 values, which is often done in CDST analyses (Chan et al., 2015; Verspoor et al., 2008). 0 – 1 values were chosen over z-scores to provide the ability to calculate the comparison group’s 95% confidence interval. As is often found in CDST studies, Juan’s scores are rarely linear (see Figures 4 – 17). His scores often varied from session to session and even within the same session. Therefore, the following analyses are based on polynomial trendlines which smoothed the data and revealed emerging patterns despite fluctuations in the data. Finally, Juan’s scores were compared to the 95% confidence interval of the comparison group in order to see the extent to which he adapted to his speech community’s standards.

Overall, Juan improved in 8 of 14 variables. He improved and began scoring above the comparison group in both lexical diversity (starting during month 1) and subordination (starting after month 8), though his scores were not consistent (see Figures
4 and 5). His lexical diversity scores peaked during month 10 and then regressed slightly. However, the trendline never moved below the comparison group.

Figure 4. Lexical diversity development 0 – 1 values.
*LD = Lexical Diversity, CGMin/CGMax = 95% Comparison Group Confidence Interval, Poly = Polynomial Trendline*

Figure 5. Subordination 0-1 development 0 – 1 values
*Subord = Subordination scores, CGMin/CGMax = 95% Comparison Group Confidence Interval, Poly = Polynomial Trendline*
In the fluency constructs, the trendlines show gradual improvements in six of the ten variables. The number of silent, filled, and total pauses, seconds per silent pause, and false starts reduced (see Figure 6, 7, 8, 9 and 10) and speech rate increased (Figure 11), all of which indicate improvement.

The remaining fluency variables, redundancies and total repair disfluencies, worsened and then returned to their previous levels (see Figures 12 and 13).

Figure 6. Silent pauses development 0 – 1 values
*S Pauses* = Silent Pauses, CGMin/CGMax = 95% Comparison Group Confidence Interval, Poly = Polynomial Trendline
Figure 7. Filled pauses development 0 – 1 Values
*F Pauses* = Filled Pauses, *CGMin/CGMax* = 95% Comparison Group Confidence Interval, *Poly* = Polynomial Trendline

Figure 8: Total pauses development 0 – 1 Values
*Total* = Total Pauses, *CGMin/CGMax* = 95% Comparison Group Confidence Interval, *Poly* = Polynomial Trendline
Figure 9. Seconds per silent pauses development 0 – 1 values
F Pauses = Filled Pauses, CGMin/CGMax = 95% Comparison Group Confidence Interval, Poly = Polynomial Trendline

Figure 10. False starts development 0 – 1 values
F Start = False Starts, CGMin/CGMax = 95% Comparison Group Confidence Interval, Poly = Polynomial Trendline
Figure 11. Speech Rate development 0 – 1 values

$SpRate = Speech Rate, CGMin/CGMax = 95\% Comparison Group Confidence Interval, Poly = Polynomial Trendline$

Figure 12. Redundancies development 0 – 1 values

$Redund = Redundancies, CGMin/CGMax = 95\% Comparison Group Confidence Interval, Poly = Polynomial Trendline$
Though benchmarks for fluency are difficult to establish due to the complexity of defining fluency (Lambert & Kormos, 2014), it is noteworthy that the improved variables moved closer to the 95% confidence interval of the comparison group (see Figures 6 – 11). However, the trend was not always linear. While the trendlines of silent, filled, and total pauses show a linear decrease over time, seconds per silent pause and speech rate both peaked around month 8, then moved back towards their previous levels. Since they did not return to or go below their previous levels the changes still indicate improvement. Lastly, false starts showed a general decrease, but due to fluctuating scores towards the end, Juan never reached the speech community’s standard (see Figure 10).

The last two variables, redundancies and total repair disfluencies, appeared to temporarily increase (worsen) peaking at month 8, then returned to their previous levels, which were almost within the speech community’s norms (see Figure 12 and 13). Since
redundancies were more common than the other repair disfluencies in Juan’s data, it is not surprising that total repair disfluencies (the global variable) follows redundancies’ trajectory pattern.

Lastly, four variables were in a state of fluctuation at the end of the study: errors per AS-unit, elaboration, reformulations, and articulation rate (see Figures 14 – 17). Despite improving trendlines in elaboration, reformulations, and articulation rate, Juan’s bimodal scores suggest instability rather than emerging trends. Elaboration and reformulations continually move from high to low starting in month 5 (see Figures 15 and 16). Throughout the study Juan’s reformulation scores are generally higher than the comparison group, only reaching or going below the 95% confidence interval 10 out of 36 times. In the case of articulation rate, the trendline is flat and remains within the comparison group’s 95% confidence interval, even though the scores continually move from 0 – 1 (see Figure 17).

Fluctuations in errors per AS-unit and elaboration (see Figures 14 and 15) may indicate that Juan was restructuring his language with new grammar forms, function words, or descriptive words. Since reformulations and articulation rate are measures of performance rather than L2 knowledge, that fluctuation needs to be considered in connection to another variable to be meaningful. Section 5.2 will make such a comparison.
Figure 14. Errors per AS-unit development 0 – 1 values
0-1 Err/ASU = Errors per AS-unit, CGMin/CGMax = 95% Comparison Group Confidence Interval, Poly = Polynomial Trendline

Figure 15. Elaboration development 0 – 1 values
Elab = Elaboration, CGMin/CGMax = 95% Comparison Group Confidence Interval, Poly = Polynomial Trendline
Figure 16. Reformulations development 0 – 1 values
Reform = Reformulations, CGMin/CGMax = 95% Comparison Group Confidence Interval, Poly = Polynomial Trendline

Figure 17. Articulation Rate development 0 – 1 values
ArtRate = Articulation Rate, CGMin/CGMax = 95% Comparison Group Confidence Interval, Poly = Polynomial Trendline
As a last point, even though some conflated variables performed equally, others did not. Total pauses (combination of silent and filled pauses) and speech rate (conflation of seconds per silent pauses and articulation rate), moved in the same direction as their non-conflated variables. However, subordination (part of elaboration) and false starts (part of total repair disfluencies) moved in different directions than their global counterpart. Thus, even though subordination and elaboration are both measuring complexity, and false starts and repair disfluencies are both measuring repair disfluency, they are measuring different aspects of the construct.

5.1.2 Fluctuation Analysis

The next set of analyses examined changes in the fluctuation levels in the data, which are indicated by adding moving minimum and maximum lines to the graphs. Increased fluctuation may indicate that Juan was restructuring his language patterns and was therefore inconsistent in his usage (Polat & Kim, 2014; Verspoor et al., 2011). In order to complete a phase shift, learners need to move to a new attractor state (Larsen-Freeman, 2006; Polat & Kim, 2014), which would be indicated by the learner’s scores stabilizing at either a higher or lower score than before the period of fluctuation. This section will describe four periods where Juan entered a period of fluctuation. However, a complete phase shift was not captured. Two variables did not change to a different level after fluctuation and two variables were in fluctuation at the end of the study.
Consequently, no conclusive arguments can be made for a complete phase shift as described in the literature.

Juan’s scores for lexical diversity and filled pauses entered a period of fluctuation during month 5, both ranging from zero to one (R = 1) (see Figure 18 and 19). Both appear to stabilize, lexical diversity during month 10 with scores varying from .87 to .61 (R = .26) and filled pauses during month 8 varying from .26 to .35 (R = .09). Though more stable than before, Juan’s lexical diversity’s scores decrease from month 10 until the third transcription in month 15. Consequently, it is not likely that this change indicates a phase shift, which would be stable rather than decreasing. In contrast, filled pauses appear to stabilize at a lower rate than the period of fluctuation with a brief period of fluctuation at month 11. However, fluency variables are thought to indicate proceduralization rather than the learner’s proficiency (Segalowitz, 2010; Skehan, 2003). Thus, Juan may have moved to an attractor state of less filled pauses or he may just be pausing less due to the reduced strain on his cognitive resources.
At the beginning of the study, subordination and errors per AS-unit appeared to be generally stable, but then entered periods of fluctuation as the study progressed (see Figures 20 and 21). For the first five months, errors per AS-unit ranged from .08 to .77 ($R = .69$). During month 6, errors per AS-unit entered a short period of fluctuation with
ranges from .17 to 1 (R = .83), followed by a brief stable period with scores between .10 to .57 (R = .47) until month 10. Then, from the month 10 until the end, errors per AS-unit changed to the largest period of fluctuation between 0 and 1 (R = 1). Since the study ended during a period of fluctuation in errors per AS-unit, it is impossible to know when or if the scores returned to a stable state and whether the fluctuation resulted in a phase shift.

Figure 20. Min-max graph for errors per AS-unit 0 – 1 values
Subordination scores varied from zero to 0.5 (R = .5) from halfway through month 1 until month 8 (see Figure 21). Beginning at month 8, subordination entered a period of fluctuation with scores ranging from .16 to 1.00 (R = .84), then returned to more stable scores beginning at the end of month 12 with scores ranging from .29 to .92 (R = .63). Even though the scores stabilized, they returned to their previous levels which indicates a return to the previous attractor state rather than a phase shift. In addition, the in the last sample, subordination returns to a score of 1, which may indicate a new period of fluctuation.

5.1.3 Error Analysis

To complete the development analysis, the raw data of coded errors from the CLAN output were entered into an Excel spreadsheet organized by date. The data were then analyzed for patterns of interest. The most frequent errors included semantically
related words choices (incorrect prepositions, pronoun gender, or conjunctions) and omission of the “s” on the 3rd person singular. Both errors remained stable throughout the study showing no improvement or worsening. Two variables appeared to enter a period of fluctuation: incorrect past tense formation, and auxiliary verb omission. Before month 5, Juan had only one error related to the past tense (see Figure 22, 1Columbia). However, he only used past tense an average of .625 times per transcript, which could indicate an inability to form the past tense or avoidance of the structure. From months 5 – 8, he had an average of 1.17 past tense related errors per transcript and used the past tense an average of 2.16 times per transcript. Lastly, during months 9 – 15, he had only two transcripts with past tense errors (see Figure 22, 112Set3 and 15Dentist), yet used it an average of .93 times per transcript. Thus, it appears that his past tense usage went through a period where he rarely used the past tense, followed by increased usage and errors, and then increased accuracy, but lesser usage. In this case, it appears that Juan went through a period of experimentation, but did not restructure his language. Rather he returned to a state of lesser usage (with fewer errors). It should also be mentioned that his past tense usage was not affected by task as he completed one of each task between months 5 – 8.
A similar pattern was found with Juan’s auxiliary verb omission (see Figure 23). With two exceptions of present perfect on the first and last transcript, all of Juan’s auxiliary verb usage was related to progressive aspect. From month 1 to midway through month 5, Juan omitted auxiliary verbs from present progressive constructions three times out of 28 usages (11%). Moreover, he did not use past or future progressive. From halfway through month 5 to month 8, he omitted the auxiliary verb six times out of 26 usages (23%). Additionally, the past and future progressive appeared three times in these transcripts without errors. After month 8, he no longer omitted auxiliary verbs. However, he also reverted to using only the present progressive until the last month when three instances of the past progressive and one instance of the past perfect appeared. In sum, Juan appears to have experienced a phase shift related to the progressive aspect. His initial attractor state included only the present progressive with auxiliary verbs omitted occasionally. Then he entered a period fluctuation where he was experimenting with new forms (past and future...
progressive), but made more errors related to the present progressive. Finally, after month 8, he entered a new attractor state with correct auxiliary verb usage, but limited his usage to only the present progressive. During month 15, he may have entered a new period of fluctuation where he was experimenting with past progressive and present perfect aspect again.

![Graph showing present progressive usage and errors](image)

**Figure 23.** Present progressive usage and errors

The following excerpts from Juan’s transcripts illustrate qualitatively how Juan’s verb usage improved. Example c and d are describing the same picture in a *One Frog Too Many* (Mayer, 1975) approximately one year apart.

**c)** Ah so (.) they’re in the water | and finally this (.) small frog (.) jumps (.) into the (.) the raft (2OFTM2, April 15, 2013, month 2)

**d)** she’s watching* ahead | (. ) then the old frog just (. ) jumped <into the> (/) into the boat (14OFTM4, April 18, 2014, month 14)
In the example c, Juan only used simple present verbs, whereas in example d he used the present progressive and the past tense. Although both examples use the verbs correctly, the second is more complex in the types of verbs chosen. Moreover, in sample d, disfluencies surrounded *jumped*, (a silent pause beforehand and a redundancy after), which suggests he might have been struggling to form the past tense of this verb. Thus, his pattern of usage became more complex and he appears to be experiencing cognitive strain resulting in disfluencies. While a global variable analysis of errors per AS-unit did not capture his advancing complexity development, a more qualitative and specific analysis did.

In summary, Juan showed improvement in 8 out of the 14 variables. Of the remaining six, two worsened and then returned to their previous state, and four were in a period of fluctuation at the end of the study. The specific error analysis showed that Juan’s usage of the progressive aspect improved, his past tense usage improved temporarily, and he remained at the same error level in semantically related word choices and the third person singular.
5.2 Research Question 2: Global Variable Interaction:

*Is there evidence of interconnected development between the global variables?*

*Do they compete, support, or build on each other, or in CDST terms, are they connected growers? Does the interaction between knowledge variables and the performance variables show evidence of cognitive strain?*

To answer this research question, three types of analysis were conducted on six global variables: comparisons between periods of fluctuation, Spearman’s correlations, and moving correlations. The purpose of moving correlations is to show if and how the correlations between variables changed over time (for discussion about moving correlations see section 4.5.2). As explained in section 4.5.2, the scores in this section were converted to z-scores so that the scores in which an increase represents a negative change (e.g., increasing errors), could be inverted to a comparable score where an increase is positive. Inverted variables included errors per AS-unit (accuracy), all repair disfluencies, silent and filled pauses, and seconds per silent pauses. Moreover, z-scores better represent variation within the score in that the score is based on standard deviations rather than minimum and maximum values.

In order to compare Juan to his speech community, Spearman’s correlations were also calculated from the z-scores of the comparison group. This comparison serves the purpose of exploring whether there is a connected relationship between the CAF variables in native speakers and if there is, seeing if Juan moved toward the speech
community standards. Such comparisons could add insight towards possible differences between second language learners and native-speakers.

Since complexity and accuracy reflect the learners’ understanding of L2 forms, whereas fluency reflects learners’ performance (Derwing et al, 2007; Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009; Skehan, 1996), the comparison between variables in this study was designed to reflect that difference. Therefore, correlations between the complexity and accuracy variables were calculated first. Then, in order to consider disfluencies that may have been caused by changes in lexical diversity, accuracy, or elaboration, the fluency variables were correlated with each of the knowledge variables. By doing so, Juan’s ability to produce language fluently was compared to his developing language abilities rather than treating the variables as equal.

5.2.1 Comparisons of Variables with Periods of Fluctuation

As mentioned in the previous section, four variables (lexical diversity, errors per AS-unit, subordination, and filled pauses) were found to have periods of fluctuation followed or proceeded by periods of greater stability. Since this study limits the comparison to global variables, the specific variables of subordination and filled pauses will not be included in this analysis.

Once the scores were converted to z-scores, a fluctuation pattern that was not visible in the 0-1 values appeared in elaboration due to the finer detail captured by a Z-score. While the 0-1 values show four scores of one and four scores of zero (each
representing a minimum or maximum score for each task), the z-scores show standard deviations from the average scores allowing for a more detailed analysis. Thus, the high z-scores were \( z = 1.06, 1.45, 1.87, \) and 3.12 and the low scores were \( z = -.87, -.18, -1.08 \) and -1.94 (see Figure 24). As such, a pattern of fluctuation appeared from month 11 – 15 which was not apparent in the 0-1 values and as such elaboration will be included in this analysis.

Figure 24. Min-max graph for elaboration z-scores

Over the course of 15 months, three knowledge variables went through periods of fluctuation (lexical diversity, elaboration, and error per AS-unit) indicated by larger distances between the min-max lines. From the beginning of the study until the middle, lexical diversity appears to be in a period of fluctuation (see Figure 25). From month 1 – 9, Juan’s scores ranged from \( Z = -1.5 - 2.27 \) (\( R = 3.77 \)). After month 9 the range reduced to \( Z = -.9 - 1.8 \) (\( R = 2.7 \)).
Figure 25. Min-max graph for lexical diversity $z$-scores

From the middle to the end, fluctuations increased in both elaboration and errors per AS-unit (see Figures 24 and 26). From month 1 – 10, elaboration varied from $Z = -1.88 - 1.45$ ($R = 2.35$), whereas after month 10, scores varied from $Z = -1.04 - 3.11$ ($R = 5.05$). Error per AS-unit varied from $-1.02 - 1.81$ ($R = 2.83$) until month 7. After that point, the scores varied from $-1.41 - 2.08$ ($R = 3.49$) (see Figures 24 and 26).

In terms of interaction between variables, lexical diversity appears to be developing differently than accuracy and elaboration. While lexical diversity is fluctuating at the beginning of the study, elaboration begins fluctuating in month 11 and accuracy begins fluctuating in month 7. This may be due to Juan choosing to focus on specific variables, the environmental demands for new vocabulary, and/or the increased ability to experiment with elaboration and accuracy after a certain level of lexical
diversity is obtained. Therefore, these results may indicate a conditional grower relationship between lexical diversity and both accuracy and elaboration.

Figure 26. Min-max graph for errors per AS-unit z-scores

5.2.2 Correlations Between Complexity and Accuracy Variables

For this analysis, moving correlations and Spearman’s correlations were run between the global knowledge variables of elaboration (words per AS-unit), accuracy (error per AS-unit), and lexical diversity (D). Moving correlations were chosen to account for changes in the correlations over time (for discussion on how to calculate moving correlations, see section 4.5.2). As the results will explain in detail, the Spearman’s correlations concurred with the moving correlations but were not significant in two out of the three of the comparisons. Similar issues were found in Spoelman & Verspoor (2010), who found that changes in moving correlations explained the lack of
correlation in their findings with word complexity and accuracy rates. Since this study emphasizes the complex relationships that change over time, the moving correlations were given greater importance. Cohen’s strength of correlations was used to interpret the data.

The results of Juan’s correlations were also compared to the comparison group’s Spearman’s correlations. Although only weak correlations were found in the comparison group, they were statistically significant. Since the comparison group is only a reference, differences in the correlation will be noted, but not used for definitive conclusions.

5.2.2.1 Moving correlations between lexical diversity and elaboration. Elaboration and lexical diversity produced a consistent medium to strong negative moving correlation in 31 out of 36 data pairs (see Figure 27). As such, these variables were labeled as competitive growers. In other words, when elaboration increases, lexical diversity decreases and vice-versa. This point is more clearly illustrated in Figure 28, which shows the raw data for elaboration and lexical diversity on the same graph. An inverse relationship can be seen in the majority of the data.
Figure 27. Moving correlation between lexical diversity and elaboration

Figure 28. Lexical diversity and elaboration trajectories

Since the developmental trajectories of lexical diversity and elaboration are in opposition to each other (see Figures 4 and 15), they may have a conditional relationship (one must develop before the other appears) in addition to a competitive relationship. In the first half of the study, lexical diversity increased whereas elaboration decreased. In
the second half of the study, elaboration increased while lexical diversity decreased.
Moreover, fluctuation periods between them occurred at opposite ends of the study (see Figures 24 and 25).

One explanation for this relationship relates to content words. During the first half of the study, Juan’s sentences were mainly content words, which tend to be more lexically diverse. As his syntax abilities improved, his AS-unit lengthened due to an increase in function words. Function words were repeated frequently and consequently resulted in decreases in his lexical diversity scores.

The following excerpts from Juan’s first and second narration of *A Boy, A Dog and a Frog* (Mayer, 1967) describing the same event and taken approximately one year apart exemplify the changes in speech’s elaboration.

e) and (. ) the kid eh (. ) trap* the dog with the net (1BDF1, February 25, 2013, Month 1)

f) and then in this scene the <the> kid is (. ) with the dog in the net (13BDF3, February 24, 2014, Month 12)

In sample e, Juan includes five function words, whereas in sample f, he includes 10 function words (not counting the redundant “the”). Furthermore, the sample contains the same number of content words. Sample e contains four content words (kid, trap, dog, net) and sample f contains four content words (scene, kid, dog, net), yet f contains 5 more words than e.
The comparison group correlations showed the opposite pattern, with a weak positive yet significant correlation between lexical diversity and elaboration

\( r(188) = .244, p = .007 \). The data indicates that the speakers in this community use longer AS-units with more diverse vocabulary. Since native speakers are already using a complete set of function words, the length of AS-unit would only be affected by additional descriptors such as adverbs, adjectives, and dependent clauses rather than increased function words which would result in a higher lexical diversity score.

5.2.2.2. Moving correlations between elaboration and accuracy. The results indicate that elaboration and accuracy were competitive growers. In this case, 28 of the 36 data points were negatively correlated (see Figure 29). Thus, Juan’s shorter AS-units were more accurate than his longer AS-units. Since longer AS-units would increase the demand for accurate syntax and morphology, more opportunities for errors would be available. Similar to elaboration and lexical diversity, elaboration and accuracy developed in opposing patterns in that errors per AS-unit improved through the first half of the study, then worsened (see Figure 14) whereas elaboration worsened during the first half of the study, then improved (see Figure 15). However, since both variables were in a state of fluctuation together, it is unlikely that they are conditional growers. Juan appeared to be concentrating on one variable or the other while both were developing simultaneously. This tendency becomes more obvious during the period of fluctuation as the scores become more extreme (see Figure 30).
As would be expected, no correlation was found between elaboration and accuracy in the comparison group, whose accuracy scores were high throughout the study.

Figure 29. Moving correlation between elaboration and accuracy

Figure 30. Elaboration and errors per AS-unit trajectories
5.2.2.3. *Moving correlations between lexical diversity and accuracy*. The moving correlation for lexical diversity and accuracy showed mostly weak correlations or no correlations in 19 of the 36 data points (see Figure 31). However, it was the only variable to have a significant Spearman’s correlation \( r(34) = .416, p = .012 \). Even though the Spearman’s correlation is significant, the moving correlation shifts from positive to negative. Therefore, the correlations are fluctuating rather than maintaining a consistent pattern over time as was found with elaboration and accuracy and elaboration and errors per AS-unit. In the correlated data points, 7 data pairs are positively correlated, and 10 are negatively correlated. Thus, a connected grower relationship cannot be established. In the fluctuation analysis, the periods of fluctuation were sequential with lexical diversity fluctuating in the first half of the study and accuracy fluctuating at the end. In contrast, the developmental patterns were the same (both improved during the first half then began worsening). Since there were no consistent correlations or patterns in their growth, their development and fluctuation patterns appeared to have changed independently of each other.
Figure 31. Moving correlation between lexical diversity and accuracy

The comparison group had a weak yet statistically significant negative correlation between lexical diversity and accuracy ($r(118) = -0.309$, $p = .001$). Thus, in the comparison group, lexical diversity decreases as accuracy increases, but the relationship is relatively weak.

5.2.3 Comparisons Between Proceduralization and Knowledge Variables

The purpose of this analysis is to explore how fluency is affected by changes in accuracy and complexity, an area that remains underexplored in the CAF and CDST literature. To examine this topic, moving correlations were calculated between the global variables of accuracy, elaboration, and lexical diversity, with the global fluency variables of total pauses, total repairs, and speech rate.

In both accuracy and elaboration, the relationship with fluency remained mostly consistent over time. As Table 4 shows, accuracy was positively correlated with fluency
for the majority of the data points. Elaboration was negatively correlated with pauses; positively correlated with speech rate; and repairs were divided with slightly more data points being negatively correlated than positively correlated. Lastly, lexical diversity and the majority of all three fluency variables were weakly or non-correlated. As an additional measure, Spearman’s correlations were calculated to measure the general trends of the data. With the exception of the correlation between elaboration and total repairs, the Spearman’s correlations concur with the moving correlation data (see Table 4).

Table 4

Frequency of correlated data pairs and Spearman’s correlations between state of knowledge (complexity and accuracy) and proceduralization variables (fluency)

<table>
<thead>
<tr>
<th>Moving correlations</th>
<th>Negatively correlated</th>
<th>Positively correlated</th>
<th>Weak and non-correlated</th>
<th>r(34)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy and total pauses</td>
<td>9</td>
<td>18</td>
<td>9</td>
<td>.130</td>
<td>.448</td>
</tr>
<tr>
<td>Accuracy and total repairs</td>
<td>3</td>
<td>21</td>
<td>12</td>
<td>.330</td>
<td>.049*</td>
</tr>
<tr>
<td>Accuracy and speech rate</td>
<td>9</td>
<td>18</td>
<td>9</td>
<td>.065</td>
<td>.707</td>
</tr>
<tr>
<td>Elaboration and total pauses</td>
<td>27</td>
<td>2</td>
<td>7</td>
<td>-.353</td>
<td>.002*</td>
</tr>
<tr>
<td>Elaboration and total repairs</td>
<td>15</td>
<td>12</td>
<td>9</td>
<td>.217</td>
<td>.205</td>
</tr>
<tr>
<td>Elaboration and speech rate</td>
<td>3</td>
<td>24</td>
<td>9</td>
<td>.145</td>
<td>.400</td>
</tr>
<tr>
<td>Lexical diversity and total pauses</td>
<td>2</td>
<td>13</td>
<td>21</td>
<td>.282</td>
<td>.096</td>
</tr>
<tr>
<td>Lexical diversity and total repairs</td>
<td>17</td>
<td>2</td>
<td>18</td>
<td>.073</td>
<td>.674</td>
</tr>
<tr>
<td>Lexical diversity and speech rate</td>
<td>3</td>
<td>15</td>
<td>18</td>
<td>.171</td>
<td>.320</td>
</tr>
</tbody>
</table>

* Correlation is significant at the .05 level
In sum, fluency appears to be correlated with accuracy and elaboration, but more strongly correlated with accuracy. Fluency is only weakly correlated with lexical diversity. It is also noteworthy that the comparison group only had one significant correlation. Speech rate and elaboration had a medium positive correlation $r(118) = .316, p = .000$, meaning that as their clauses became longer, they spoke faster.

Despite the stability of the majority of the relationships in Table 4, only two sets of variables maintained a consistent relationship over the 15 months of the study: accuracy and speech rate remained in a period of fluctuation between positive and negative correlations (see figure 32); and elaboration remained negatively correlated with pauses (see figure 33). Thus, Juan’s levels of accuracy did not appear to affect how quickly he spoke, and as Juan’s clauses increased in length, he paused less and vice versa.

![Figure 32. Moving correlations between accuracy and speech rate](image-url)
As fitting with CDST, each of the knowledge variables (accuracy, elaboration, and lexical diversity) have a unique relationship with fluency yet are interconnected as a system. The remaining seven moving correlations reflect three periods in which the variables changed, months 1 – 2, months 5 – 11, and months 11 – 15. During the middle period, the knowledge variables were entering and exiting periods of fluctuation (see Figures 24 – 26). This analysis indicates that periods of fluctuation may be related to fluency.

Before month five, as Juan’s accuracy improved, his fluency generally improved as well (see Figures 34 and 35). As his elaboration and lexical diversity improved, his fluency improved in some variables but worsened in others (see Figures 36 – 40). More specifically:
a) accuracy was mostly positively correlated with pauses and repairs meaning he paused and repaired less when his accuracy increased (see Figures 34 and 35);

b) elaboration was positively correlated with pauses, yet negatively correlated with repairs meaning that as his clauses increased in length he paused more and repaired less (see Figures 33 and 37);

c) lexical diversity was mostly negatively correlated with pauses and positively correlated with repairs meaning that as he used more diverse vocabulary, he paused more and repaired less (see Figures 38 and 39).

d) Speech rate was positively correlated with lexical diversity, negatively correlated with accuracy and gradually changed from positive to negative with elaboration (see Figures 32, 36 and 40) meaning that as he used more diverse vocabulary he spoke faster; when he made errors, he spoke slower; and as he made longer clauses in the beginning he spoke slower, but gradually was able to produce longer clauses at a higher rate of speed.

In sum, as Juan’s accuracy improved, he produced fewer disfluencies and spoke slower; as his sentences became more elaborate, he paused more and repaired less; and as he increased his lexical diversity, he paused less, repaired more, and spoke faster.

As a final point, it should be noted that Juan’s first two samples (1BDF1, 1BDF2) were often different from the rest of Month 1 even though the first four samples (1BDF1, 1BDF2, 1Bogota and 1Colombia) were collected on the same day. His scores may have been affected by nervousness as they were the first sample.
Figure 34. Moving correlation between accuracy and total pauses

Figure 35. Moving correlation between accuracy and total repairs
Figure 36. Moving correlation between elaboration and speech rate

Figure 37. Moving correlations between elaboration and total repairs
Figure 38. Moving correlation between lexical diversity and total pauses

Figure 39. Moving correlation between lexical diversity and total repairs
From month 5 – 9, Juan experimented with grammar, as is evidenced by errors in past tense formation and auxiliary verbs as well as his usage of more complex verb forms such as the past and future progressive (see figure 22 and 23). The moving correlations reflect Juan’s increased cognitive strain. The correlations between accuracy and repairs, and between accuracy and pauses, switched from positive to negative meaning that his fluency is decreasing while he produces correct grammatical forms (see Figures 34 and 35). He appears to be pausing and repairing his speech in order to produce more accurate grammar, whereas before he paused and repaired his speech while producing less accurate grammar. During month 9, those correlations switch back to positive. This occurs as he enters the period of fluctuation in accuracy (see Figure 26).

Elaboration entered a period of fluctuation in month 11 (see Figure 24). During month 9, the correlation between elaboration and repairs switches from positive to negative (see Figure 37). Moreover, during months 9 – 10, the correlations between
elaboration and speech rate switch from positively correlated to negatively correlated (see Figure 36) and begin a downward trend. In other words, Juan is slowing down as his clauses lengthen. Once again, the period of fluctuation is preceded by changes in the correlations between fluency and the variable entering fluctuation.

By month 10, all three knowledge variables have either finished or started a period of fluctuation. Out of the nine correlations, five of them changed then returned to their previous states. When a correlation changed in elaboration, it did not return to its previous state. Since the correlations related to elaboration changed later than the others (month 10), they may have switched back after the study ended. Only two correlations did not change during months 5 - 10, the correlations between accuracy and speech rate, which remained in fluctuation, and the correlation between elaboration and pauses, which remained negatively correlated.

In summary, the relationships between the knowledge variables (accuracy, lexical diversity, and elaboration) and the performance variables (pauses, repairs, and speech rate) are non-linear, complex, and appear to be interconnected. Before each variable entered or exited a period of fluctuation, at least two of the correlations between fluency variables changed from positive to negative (accuracy and elaboration), or correlated to uncorrelated (lexical diversity), with the exception of elaboration and speech rate. CDST theorizes that periods of fluctuation indicate that the learner is alternating between two attractor states (the old and newly learned form) (Polat & Kim, 2014). The results of these analyses appear to show that changes in fluency occurred while Juan is learning
new vocabulary or grammar which may indicate periods of increased cognitive strain. The learning seems to occur immediately before periods of fluctuation (accuracy and elaboration) or stabilizing after alternating between forms (lexical diversity). While this is not surprising, it does provide a new tool for analyzing development.

5.3 Research Question 3: Environmental Factors

*Is Juan’s CAF performance affected by native speaker exposure and interaction?*

*Is it affected by fluctuations in his motivation?*

This section will discuss changes in Juan’s motivation and interaction with native English speakers based on the triangulation of five types of data: a) a timeline of instrumental and integrative factors that Juan mentioned in our sessions b) changes in standardized proficiency exams over time c) Juan’s own comments regarding his learning and environment, d) inclusion of colloquial speech in his samples, and e) changes in his CAF scores. The purpose of this analysis is to investigate connections between Juan’s motivation levels and changes in his CAF scores.

The first step for this analysis was to create a motivational factors timeline including integrative motivational factors, instrumental motivational factors, and standardized proficiency measures (see Appendix D). Integrative factors included events, activities, requests for language assistance, and comments regarding his attitudes toward native speakers that indicate the quantity of and interest in native speaker interaction. Instrumental factors included activities and events that would lead to or prevent achieving
his professional goals. Lastly, standardized proficiency scores were included as measures of his level, but also as a motivational factor in that one of his goals was to improve his English, specifically his TOEFL score.

Next, his speech samples were analyzed for colloquial speech. Three types of colloquial speech were chosen due to their appearance during Juan’s immersion (see Table 5) and because they appear in the control group’s samples. The three types include grammar reductions (gonna, wanna), *like* followed by a direct quote, and shortened words (*cuz* rather than because).

### Table 5

**Colloquial speech in Juan’s transcripts**

<table>
<thead>
<tr>
<th>Month and Date</th>
<th>Quote</th>
<th>Transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 (7/24/13)</td>
<td>“It’s like ‘oh my god’”’</td>
<td>5Mother</td>
</tr>
<tr>
<td>5 (7/24/13)</td>
<td>“We’re gonna get eh a regular pension or something”</td>
<td>5Roommate</td>
</tr>
<tr>
<td>5 (7/24/13)</td>
<td>“You’re gonna have to pay that” “So, I’m gonna do this”</td>
<td></td>
</tr>
<tr>
<td>7 (09/04/13)</td>
<td>“Now he’s gonna did something different”</td>
<td>7SWP2</td>
</tr>
<tr>
<td>8 (10/04/13)</td>
<td>“It’s more like ‘I have a frog on my arm’”</td>
<td>8FOHO</td>
</tr>
<tr>
<td>8 (10/04/13)</td>
<td>“He’s gonna do something”</td>
<td>8FOHO2</td>
</tr>
<tr>
<td>8 (10/08/13)</td>
<td>“He’s fine cuz something has to happen”</td>
<td>8SET 2</td>
</tr>
<tr>
<td>8 (10/08/13)</td>
<td>“It’s owner said like ‘hey you don’t have to do that’”</td>
<td>82SET2</td>
</tr>
<tr>
<td>9 (11/16/13)</td>
<td>“Something is gonna happen very very scary”</td>
<td>9HORROR</td>
</tr>
<tr>
<td>9 (11/16/13)</td>
<td>“It’s gonna be nice”</td>
<td>9HORROR2</td>
</tr>
<tr>
<td>10 (12/03/13)</td>
<td>“I like that cuz you are in a kin”</td>
<td>10Mentor</td>
</tr>
<tr>
<td>13 (02/24/14)</td>
<td>“he’s just screaming saying something like ‘Hey, I’m gonna catch you.’”</td>
<td>13BDF3</td>
</tr>
<tr>
<td>14 (04/18/14)</td>
<td>“and the kid is saying like ‘you have to stay there’”</td>
<td>14OFTM3</td>
</tr>
</tbody>
</table>
These two types of data were then compared to changes in Juan’s CAF scores to see if changes in his CAF scores coincided with anything unusual in his life events. Lastly, Juan’s own comments regarding his motivation were considered to ensure that the motivational factors were affecting his motivation. These comments included emotional commentary about struggles and successes, and comments about his attitudes toward his life in the United States, his changing goals, the English language, and speakers of English.

Based on Dörnyei’s (2005) Process Model of L2 Motivational, the data was then divided in the Preactional, Actional, and Postactional Stages of motivation. During months 1 – 2, the events in Juan’s life were typical of a study abroad student. He took the Oxford Quick Placement Exam Version 1 (UCLES, 2001a) in which he scored 35/60 (intermediate). One month later he took the Oxford Quick Placement Exam Version 2 (UCLES, 2001b) in which he improved to 40/60 (intermediate high). At this point he was living on campus and attending classes in English. During months 3 and 4, he returned to Colombia and reported speaking only Spanish during his visit.

During months 5 – 10, Juan’s instrumental and integrative motivation appear to be high in that he was taking active steps toward integrating in his community and setting long term goals such as entering a doctoral program in the United States. On the Model of L2 Motivation (Dörnyei, 2005), Juan appeared to have moved from the Preactional stage (setting goals) to the Actional stage (carrying out subtasks). Thus, he was moving
towards what Dörnyei (2005) would define as the ideal L2 self or “specific representations of one’s self in future states” (p. 99). In order to achieve Juan’s ideal L2 self, he needed to improve his TOEFL score to meet doctoral programs’ minimum standards and become fluent enough to integrate into daily US life and academia. From months 7 – 9, he applied to three prestigious universities’ doctoral programs, registered to take the TOEFL exam, and asked for help preparing for it.

Juan also showed evidence of integrating into the community, which was a struggle for him due to his introverted nature. First, he moved to an off-campus apartment and mentioned interactions in English with the doorman. He also mentioned that he was the only Spanish speaker in most of his classes, had befriended a native English speaker with whom he spoke “Spanglish”, and a Puerto Rican with whom he spoke some English. Moreover, his language showed evidence of integration. Towards the end of month 5, colloquial speech began to appear in his transcripts (see Table 5 above). He began bringing samples of vocabulary and phrases from people he spoke to with him to our sessions that he was struggling to understand, including phrasal verbs used by his adviser. He also took the TOEFL exam during month 9. Although he was disappointed with the results, he improved five points and his highest score was in speaking. Lastly, during month 10, two musical compositions that he wrote were featured in a concert. In the following excerpt, he expresses how pleased and surprised he was that people talked to him about his compositions after the performance (in English).
two guys approached me … they were really honest that they like it, the piece. … sometime you say OK it's consolation just a formal a polite way to say ok, but they were really... I mean honestly. (Conversation, December, 3, 2013, Month 10, see Appendix F).

During the last four months of his stay, Juan’s experienced a motivational transformational event, which Shoaib & Dörnyei (2005) define as an event that will “result in the profound restructuring of the individual's motivational disposition” (p. 31).

In month 12, Juan reported he and his wife were expecting a baby. While he expressed his continued interest learning English, he was conflicted about remaining in the US. For example, he said “She’s a little sick and she’s having a problem and she’s alone (.) uh so now I feel like I need to go there” (session transcript, February 24, 2014, Month 13, see Appendix F). Even though he said that he would stay if he was accepted at one of the universities where he applied, his attitude about doctoral programs seemed to change. In the same session, he said:

So when I think about the Ph.D. (.) I think oh my god (.) I don’t want it … I mean I really don’t want to do it and I think what should I do that if I don’t want? … I mean, I know it’s a kind of an important thing for academic focus (.) but in fact I I I don’t think so I mean … I think that that the academic world create those ideas (.) I mean they are not really important. (session transcript, February 24, 2014, Month 13, see Appendix F)

In addition, he reported that his L1 English friends had begun speaking to him in Spanish so that they could practice. Despite these changes, he still scored 39/60 (intermediate high) on the Oxford Quick Placement Examination Version 1 (UCLES, 2001a) during this month. At this point, Juan appeared to be in the Postactional Stage of
the Model of L2 Motivation (Dörnyei, 2005). He was evaluating his situation, finding causes for his successes and failures, and creating new goals.

By month 13, he made his decision to return home. Nevertheless, he said he wanted to push himself to learn as much English as he could before he left. Beginning in month 10, his usage of colloquial speech reduced considerably in his transcripts and disappeared altogether by month 15. When I asked about colloquial speech, Juan explained that “I just tried to use what I heard by that time; sometimes I did it consciously and others I didn’t. What I think is that sometimes I just tried to speak freely, trying to think not much” (personal communication, June 7, 2016, Appendix F). If he was in fact still using what he heard, his reduction in colloquial speech probably indicates that his native speaker exposure had diminished. During month 13, he took the Oxford Quick Placement Examination Version 2 (UCLES, 2001b) and received a lower score 38/60 (intermediate level). He also reported less contact with native English speakers and mentioned feeling uncomfortable here. Specifically, he said, “I don't feel comfortable because I feel like a stranger everywhere.” (session transcript, April 25, 2014, Month 15, Appendix F). Finally, he reported that he was not accepted to the universities of his choice. However, he mentioned that he may return to Temple for his doctorate eventually or attend a doctoral program in France.

Even though Juan expressed his desire to keep learning English during this period, his reasons for improving appear to be strictly instrumental. For example, he was writing his master’s thesis and was scheduled to give an oral thesis presentation in
English. He also mentioned that when he returned to teaching in Colombia, he would only need English to read textbooks and materials to prepare to teach his classes. He added that “probably I would have to do something with English I think 10 percentages” (session transcript, April 25, 2014, month 15, Appendix F).

Regarding Juan’s ought-to and L2 ideal self, he seemed to have shifted to an ought-to self, motivated by instrumental tasks, his family and colleague’s expectations, and his own disappointments. When he explained why he came to Temple he mentioned pressure from his professors in Colombia. He said, “I never did what I want to do I usually follow my maestros [teachers] … they say you’re good at this but and they okay you say I’m good in that okay and they are right absolutely they are right.” (session transcript, February 24, 2014, month 13, Appendix F). He also mentioned that his wife and family encouraged him to come to the United States and that they were all proud of him.

In terms of his own expectations, Juan expressed his disappointment in his speaking skills. He said, “my speaking has improved but not what I expected … I wanted to speak really (. ) really really fluent … but I have to think a lot … it's like the words goes go out from my mouth but I don't control them” (session transcript, February 24, 2014, month 13). Later in the session he added, “I’m worried because I’m leaving here … I would and I think why (. ) I should I should be better. I’ve been for two years.” (session transcript, February 24, 2014, month 13, Appendix F).
To compare Juan’s motivational factors to his CAF variables, Juan’s motivational timeline was compared to the periods that the correlations between the knowledge and performance variables underwent changes. At the beginning of the study, Juan’s moving correlation data settled into a fairly stable pattern, which coincides with Juan’s Preactional Stage. From months 5 - 11 seven of the moving correlations between knowledge and performance variables shifted into a new pattern which seems to indicate increased cognitive strain, particularly in the area of accuracy. At the same time, Juan was applying to graduate schools, taking the TOEFL exam, and colloquial speech was appearing in his speech samples. He appeared to be in the Actional Stage, which coincides with indications of increased cognitive strain in his CAF variables. Lastly, from months 11-15, he decided to return home and was evaluating his progress. He appeared to be in the Postactional Stage, which coincided with a shift back to his original fluency patterns.

Thus, both the qualitative and quantitative data indicated that Juan increased his efforts to improve his English from months 5 – 11. The qualitative data indicated that instrumental and integrative purposes motivated by his L2 ideal self influenced his efforts. By month 12, the effects of the changes to his future plans appeared to affect both his quantitative and qualitative data. While most of his scores continued to improve, particularly in the construct of fluency (reformulations, redundancies, filled pauses and articulation rate), others began to worsen (lexical diversity, silent pauses, speech rate), and accuracy and elaboration entered a period of fluctuation. Furthermore, his colloquial
speech reduced and completely disappeared during month 14. As Juan himself pointed out, interacting with native speakers became a burden:

    Now, I feel that is really difficult to improve if we do not speak a lot with others, but I am not that person who want to speak to others or to have friends. So I keep improving my writing, reading and listening skills, but my speaking is everyday worst and worst. (Essay, March 27, 2014, month 14, Appendix F)

    Moreover, deadlines for completing his thesis added additional pressure to refocus his language efforts. He mentioned on various occasions during this period that he felt that his writing skills were improving. For example, during a session in month 15, when asked where thought he improved the most, he responded, “I did a good job with writing … if I focus on writing I I can do a very good job, I know that” (session transcript, April 25, 2014, month 15, Appendix F).

    In sum, changes associated with Juan’s CAF scores and colloquial speech usage appear to have been affected by changes in Juan’s life events. The triangulation of the qualitative data, quantitative data, and comments from Juan, indicate that his decision to return home rather than continuing his education moved him into the Postactional stage of Dörnyei’s (2005) L2 Motivational System and may have influenced his CAF scores as well.
5.4 Research Question 4: Task effects

Is Juan’s CAF performance affected by the type of task being performed?

This analysis contributes to task-based learning by analyzing if differences between tasks affected Juan or the comparison group’s CAF scores. The analysis was also completed in order to decide whether or not 0-1 values and z-scores should be calculated separately for each task before combining them on the timeline. Since differences were found, the scores were calculated separately and then combined. The data was analyzed using non-parametric Kruskal-Wallis tests since the majority of the data for both Juan and the comparison group was not normally distributed.

In Juan’s data, only three out of 15 variables showed significant differences: lexical diversity ($\chi^2(3) = 12.19, p = .007$), articulation rate ($\chi^2 (3) = 12.37, p = .006$), and seconds per silent pause ($\chi^2 = 13.07, p = .004$) (see Table 6).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Chi-Square Score</th>
<th>df</th>
<th>med</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Diversity</td>
<td>12.19</td>
<td>3</td>
<td>92.27</td>
<td>.007</td>
</tr>
<tr>
<td>Articulation Rate</td>
<td>12.37</td>
<td>3</td>
<td>3.97</td>
<td>.006</td>
</tr>
<tr>
<td>Seconds per Silent pause</td>
<td>13.07</td>
<td>3</td>
<td>0.98</td>
<td>.004</td>
</tr>
</tbody>
</table>

However, the control group varied significantly on most tasks (see Table 7), indicating that task differences may still need to be considered. The comparison group’s results showed significant differences between the three types of stimuli including A Frog goes to Dinner (Mayer, 1974), Mr. Bean at the Swimming Pool (Mr. Bean, 2009b), and
Set 3 of the picture description tasks on all but three variables: redundancies ($\chi^2(2) = 3.622, p = .163$), reformulations ($\chi^2(2) = 2.437, p = .296$), number of silent pauses ($\chi^2(2) = 3.917, p = .141$) (see Table 7).

**Table 7**

*Task differences between comparison group members*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Chi-Square</th>
<th>df</th>
<th>med</th>
<th>p</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Diversity</td>
<td>31.07</td>
<td>2</td>
<td>94.58</td>
<td>.000*</td>
<td>43.73</td>
<td>259.26</td>
</tr>
<tr>
<td>Elaboration</td>
<td>16.96</td>
<td>2</td>
<td>9.68</td>
<td>.000*</td>
<td>6.38</td>
<td>21.40</td>
</tr>
<tr>
<td>Subordination</td>
<td>7.11</td>
<td>2</td>
<td>1.50</td>
<td>.029*</td>
<td>0.22</td>
<td>3.20</td>
</tr>
<tr>
<td>Errors/AS-unit</td>
<td>6.48</td>
<td>2</td>
<td>0.18</td>
<td>.039*</td>
<td>0.00</td>
<td>18.00</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.62</td>
<td>2</td>
<td>0.00</td>
<td>.163</td>
<td>0.00</td>
<td>18.00</td>
</tr>
<tr>
<td>False Start</td>
<td>10.47</td>
<td>2</td>
<td>0.00</td>
<td>.005*</td>
<td>0.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Reformulations</td>
<td>2.44</td>
<td>2</td>
<td>0.00</td>
<td>.296</td>
<td>0.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Total Disfluencies</td>
<td>7.99</td>
<td>2</td>
<td>1.00</td>
<td>.018*</td>
<td>0.00</td>
<td>9.00</td>
</tr>
<tr>
<td>Number of Silent Pauses</td>
<td>3.92</td>
<td>2</td>
<td>13.00</td>
<td>.141</td>
<td>1.00</td>
<td>28.00</td>
</tr>
<tr>
<td>Number of Filled Pauses</td>
<td>39.47</td>
<td>2</td>
<td>2.00</td>
<td>.000*</td>
<td>0.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Total Pauses</td>
<td>12.99</td>
<td>2</td>
<td>15.00</td>
<td>.002*</td>
<td>1.00</td>
<td>33.00</td>
</tr>
<tr>
<td>Speech Rate</td>
<td>29.25</td>
<td>2</td>
<td>3.00</td>
<td>.000*</td>
<td>1.27</td>
<td>4.69</td>
</tr>
<tr>
<td>Articulation Rate</td>
<td>16.93</td>
<td>2</td>
<td>4.66</td>
<td>.000*</td>
<td>2.26</td>
<td>7.08</td>
</tr>
<tr>
<td>Seconds / Silent Pause</td>
<td>42.67</td>
<td>2</td>
<td>1.10</td>
<td>.000*</td>
<td>0.49</td>
<td>4.93</td>
</tr>
</tbody>
</table>

*Statistically significant at $p < .05$*

Although beyond the scope of this study, the results add validity to Foster and Tavakoli’s (2009) point that a “native-speaker baseline is a neglected dimension of studies into second language performance” (p. 866). While most studies assume that native-speakers would not be as affected as second language learners by task differences, these results indicate that they were in fact affected more than Juan.
Despite the lack of statistical significance in Juan’s data, his descriptive data is compatible with the comparison group’s statistical and descriptive data in several areas (see Tables 7, 8 and 9). For both Juan and the comparison group, lexical diversity and errors per AS-unit are lower in the frog stories indicating that this task is the least cognitively complex. The Mr. Bean videos produced the slowest speech rate in both groups, yet for different reasons. While Juan had significantly longer silent pauses (see Table 6) the control group had a significantly slower articulation rate (see Table 7). Lastly, the picture descriptions produced significantly higher scores in lexical diversity
(see Tables 6 and 7) and the most accuracy errors (see Tables 8 and 9) in both groups, which indicates it was the most complex task. Moreover, the first time Juan completed the picture description task, he commented on how difficult it was. He said “I really felt uncomfortable with this exercise … Actually because I can’t find the word” (session transcript, July 10, 2013, Month 5 Appendix F).

### Table 9
Mean and Standard Deviation for comparison group’s tasks

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frog stories</th>
<th>Mr. Bean video</th>
<th>Picture descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$ ($SD$)</td>
<td>$M$ ($SD$)</td>
<td>$M$ ($SD$)</td>
</tr>
<tr>
<td>Lexical Diversity</td>
<td>33.54 (33.54)</td>
<td>99.04 (28.99)</td>
<td>126.06 (39.52)</td>
</tr>
<tr>
<td>Elaboration</td>
<td>11.19 (2.80)</td>
<td>8.98 (1.77)</td>
<td>9.72 (2.03)</td>
</tr>
<tr>
<td>Subordination</td>
<td>1.68 (0.43)</td>
<td>1.54 (0.28)</td>
<td>1.43 (0.30)</td>
</tr>
<tr>
<td>Errors/AS-unit</td>
<td>0.18 (0.30)</td>
<td>0.27 (0.23)</td>
<td>0.26 (0.27)</td>
</tr>
<tr>
<td>Silent Pauses</td>
<td>11.50 (4.42)</td>
<td>13.25 (3.65)</td>
<td>13.53 (4.30)</td>
</tr>
<tr>
<td>Filled Pauses</td>
<td>2.25 (2.27)</td>
<td>1.03 (1.17)</td>
<td>4.27 (2.22)</td>
</tr>
<tr>
<td>Total Pauses</td>
<td>13.73 (5.52)</td>
<td>14.27 (3.87)</td>
<td>17.80 (5.36)</td>
</tr>
<tr>
<td>Redundancies</td>
<td>0.78 (1.23)</td>
<td>0.85 (1.73)</td>
<td>1.05 (1.50)</td>
</tr>
<tr>
<td>False Starts</td>
<td>0.28 (0.51)</td>
<td>0.08 (0.26)</td>
<td>0.48 (0.72)</td>
</tr>
<tr>
<td>Reformulations</td>
<td>0.73 (0.88)</td>
<td>0.43 (0.64)</td>
<td>0.55 (0.75)</td>
</tr>
<tr>
<td>Total Repair</td>
<td>1.78 (1.78)</td>
<td>1.35 (2.13)</td>
<td>2.08 (1.79)</td>
</tr>
<tr>
<td>Disfluencies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech Rate</td>
<td>3.33 (0.41)</td>
<td>2.6 (0.66)</td>
<td>3.00 (0.43)</td>
</tr>
<tr>
<td>Articulation Rate</td>
<td>4.33 (0.53)</td>
<td>4.88 (0.74)</td>
<td>4.77 (0.73)</td>
</tr>
<tr>
<td>Length of Silent</td>
<td>0.85 (0.17)</td>
<td>1.98 (1.18)</td>
<td>1.22 (0.35)</td>
</tr>
<tr>
<td>Pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In addition to the similarities, there was a noteworthy difference between Juan and the comparison group. In the control group, subordination was lowest in the picture description task and highest in the frog stories, whereas the opposite was true with Juan.
Moreover, Juan produced the least errors in this task. It appears that he had chosen to focus on accuracy rather than complexity while narrating this task.

Lastly, the conversation task (which was only performed by Juan) produced more fluent and more complex scores than most of the other tasks with the fewest pauses and repair disfluencies as well as the highest elaboration and subordination scores.

To answer this research question, both the comparison group and Juan’s CAF scores were considered. Since significant differences were found on most variables in the control group and Juan’s descriptive results mimicked many of the control groups significant results, it was decided that Juan’s CAF scores were most likely affected by task differences. In order to generalize these conclusions for second language learners, additional studies with larger groups of L2 English learners would need to be conducted.
CHAPTER 6
DISCUSSION AND CONCLUSION

The aim of this study was to explore the longitudinal development of the CAF variables over time using CDST as a theoretical framework. As such, the contributions of this work relate to both SLA and CDST. In addition to the analysis of Juan’s language system as a whole, the results are also relevant to the various pieces including each of the CAF variables, motivation to learn an L2, and task-based differences. In what follows, I will discuss the answers to the four research questions, their theoretical and methodological implications, the limitations of this study, and directions for future research.

6.1 Second Language Development in a Naturalistic Environment

6.1.1 Research Questions 1: Development of CAF

Juan’s usage of the CAF variables changed in non-linear patterns over the 15 months of this study. Since the trajectories were non-linear, trendlines were used to smooth the data. Based on those trendlines, eight variables showed improvement; two variables worsened temporarily and then returned to their previous state; and in the remaining four variables, improvement or worsening could not be determined because they ended in a period of fluctuation. The eight variables that showed improvement included lexical diversity, subordination, seconds per silent pause, filled pauses, silent pauses, total pauses, speech rate, and false starts. Five of those showed emerging
improvement, whereas three peaked at month 8, then regressed slightly (see Figures 4–11). Although Juan’s scores did not ultimately worsen in any of the variables, six variables did not show improvement (see Figures 12-17). Redundancies and total repair disfluencies, temporarily worsened and then returned to their previous states. Lastly, the remaining four variables, errors per AS-unit, elaboration, reformulations, and articulation rate, were in a state of fluctuation at the end of the study and consequently could not be labeled as improved or worsened.

CDST posits that factors within a language system interact, which creates changes within the system (Larsen-Freeman & Cameron, 2008). Based on a learner’s initial state and the subsequent interaction between variables, the results for each learner may take very different paths, resulting in unpredictable outcomes. Juan’s scores changed considerably from session to session and often within the same session or transcript.

From session to session, outside factors such as life events, fatigue, excitement from the day’s activities, frustration from failures, or elation from successes may have affected his scores. As was found by Waninge, de Bot, and Dörnyei (2014), learners’ scores may also be affected within one session by fatigue, boredom, increased interest, and periods of increased concentration.

Juan’s non-linear development appeared to self-organize and adapt within his cognitive and social ecosystem just as CDST would predict. Thus, goals to improve his English a need to socially integrate, and his ideal L2-self, motivated him, while his developing knowledge and cognitive limits created parameters for his development.
Based on Juan’s TOEFL and Oxford Quick Placement Exam, Juan entered the study at an intermediate low level. In order to complete his master’s degree and successfully maneuver through his daily life, he needed to interact with professors, classmates, and other interlocutors involved in daily activities. The resulting interactions provided enough language information and examples to move toward various attractor states (increased lexical diversity, subordination, colloquial speech, and fluency). Consequently, even though Juan’s scores are non-linear, his language system self-organized in ways that led to improvement.

Juan entered or exited periods of fluctuation in three knowledge variables during the study. Theoretically, periods of fluctuation reflect a transition period that may result in restructuring learners’ language system, or in CDST terms, a phase shift (Polat & Kim, 2014). Juan entered a period of fluctuation in errors per AS-unit in month 5 and in elaboration in month 11 (see Figures 24 and 26). Additionally, he exited a period of fluctuation in lexical diversity during month 10 (see Figure 25). Although lexical diversity, elaboration, and errors per AS-unit fluctuated, none of the global variables resulted in a phase shift (restructuring). After fluctuating, lexical diversity gradually decreased rather than stabilizing. Elaboration and errors per AS-unit were fluctuating at the end of study, which makes it impossible to know if restructuring occurred. However, Juan’s use of the present progressive from the specific error analysis may indicate a phase shift. Before entering the period of restructuring in errors per AS-unit, Juan was experimenting with the past tense and progressive aspect resulting in increased errors.
while his usage fluctuated (see Figures 22 and 23). Then, the present progressive shifted into a period of consistent usage without errors, a clearly different pattern than before the fluctuation period. While phase shifts have traditionally been analyzed with singular variables, the data indicates that to define a pattern, more than one specific (rather than global) variable may be effective. In this case, the combination of usage and errors were both necessary to establish a change in pattern. While past tense entered a period of fluctuation it returned to the same pattern as before the fluctuation, indicating an unsuccessful restructuring attempt.

Given that Polat and Kim (2014) were also unable to show phase shifts in global variables, these results may indicate that future phase shift research may want to focus on specific errors. Global variables may be measuring too much at once to see the phase shifts happening within the global measure. While Polat and Kim (2014) also conducted a specific error analysis, they chose the 3rd person singular present -s, which is notoriously under-developed among ESL learners (Ellis, 2008; DeKeyser, 2005). Moreover, Juan also remained in a state of fluctuation with third person singular present -s throughout the study.

6.1.2 Research Question 2: CAF global variable interaction

The results of the second research question regarding CAF interaction was divided into two parts: correlations between L2 knowledge variables (accuracy, elaboration, and lexical diversity) and between L2 knowledge and proceduralization
variables (total pauses, total repairs, and speech rate). Rather than the statistical significance of the combined scores, the relationships were based on consistent moving correlations over time and the strength of Cohen’s effect size.

Between the knowledge variables, a medium to strong negative correlation, or a competitive grower relationship, was found between elaboration and lexical diversity and between elaboration and accuracy. As Juan’s clauses became longer, he generally made more errors and used less lexically diverse vocabulary. In this case, the grower relationships most likely reflect Juan’s increased use of function words. Over the 15 months, Juan appeared to be experimenting with adding function words and verbs forms with auxiliary verbs (see examples a – f in sections 5.1 and 5.2). The increased function words resulted in reduced lexical diversity (due to their repetitious nature) and errors in new verb forms requiring auxiliary verbs. Thus, the decreasing lexical diversity between months 11-15 (see figure 4) and fluctuating errors from months 5 – 15 (see figure 14) are marking development, though not necessarily the development the CAF variables were intended to measure. Therefore, future research of intermediate students using the measures of lexical diversity and elaboration, may want to conduct a qualitative analysis specifically looking at function words or removing them from these analyses to reveal the actual gains in these measures.

The second purpose of the moving correlations was to explore how focusing on complexity, accuracy, or lexical diversity would affect Juan’s levels of cognitive strain (measured by disfluencies and slower speech). The results showed that each of the
knowledge variables had a unique relationship with fluency. In the majority of the data, accuracy was positively correlated with total pauses, total repairs, and speech rate; lexical diversity was weakly or not correlated with total pauses, total repairs, and speech rate; and elaboration was negatively correlated with total repairs, positively correlated with speech rate, and weakly or non-correlated with total pauses. Thus, accuracy and fluency generally improved together; lexical diversity did not have a strong relationship with fluency; and the correlations between elaboration and fluency varied depending on the fluency variable.

While the knowledge variables generally maintained consistent correlations throughout the study, seven out of nine of the correlations between fluency and knowledge variables shifted into different correlations between months 5 -11, which became one of the main foci of this study. Even though task-based learning suggests that focusing on fluency could affect accuracy and complexity (Foster & Tavakoli 2009; Skehan, 1996), moving correlations between the fluency and knowledge variables in Juan’s data seemed to suggest that his disfluencies were also strongly related to cognitive strain. Juan’s development appears to begin with increased cognitive strain related to lexical diversity (increased diversity with slower speech with more repairs) followed by increased cognitive strain related to accuracy during months 5 - 10 (increased accuracy with more pauses and repairs), and lastly, increased cognitive strain related to elaboration in months 10 through 15 (longer phrases with more repairs and slower speech). As Segalowitz (2010) and Levelt (1989, 1999) hypothesized, language forms that have not
been proceduralized strained the learner’s cognitive resources, thus creating disfluencies. While Juan’s results may simply reflect the easiest development path for Juan, the environmental data indicates his decision to focus on certain areas may have been more deliberate. The next section discusses this possibility.

6.1.3 Research Question 3: Environmental Factors

In a complex system, agents within the system and factors in the system’s environment may lead to profound changes to the system (Holland, 2006). Thus, elements from the social and cognitive ecosystem may affect a language learner. As researchers are finding, a CDST framework that allows them to analyze motivational changes over time is a powerful element in understanding a learner’s development. For example, Campbell and Storch’s (2011) study shows that the same factor that motivates one student may discourage another. Since CDST highlights individual differences, studies utilizing its framework can delve deeper into individual differences between motivational factors.

Juan’s data indicates that he began moving toward an ideal L2 self, then during month 12, shifted toward the ought-to-self due to a motivational transformational episode. In the beginning of the study, his ideal L2 self was that of a doctoral student studying in the US. However, when he discovered he and his wife were expecting a baby, he decided to return to Colombia, and his L2 self was reoriented towards becoming a professor in Columbia who only uses English occasionally. At this point, his language
system reorganized due to the shift to an ought-to self (Dörnyei, 2009), which was inspired by family, colleagues, his “maestros”, and his own expectations telling him what he ought-to do such as completing his master’s thesis. Thus, his goals became more instrumental and less integrative.

During the study, Juan appeared to go through the stages of the Process Model of L2 Motivation (Dörnyei, 2005), which aligned with changes in his CAF data. He began by setting goals such as attending a doctoral program in the US (Preactional Stage). During the first two months, Juan showed improvements on the Oxford Quick Placement Examination (UCLES, 2001a, 2001b). He was living on campus and reported being the only Spanish speaker in most of his classes. Based on his CAF data, he was concentrating on vocabulary, and most likely learning English as necessary to function in his environment. At this point, Juan was entering his second semester at Temple.

During month 5, Juan mentioned that after receiving his master’s degree from Temple, he wanted to pursue his doctoral degree, which would continue his stay in the US for at least three more years. Though he may have had a goal of completing his doctoral degree in the US from the beginning the study, during Month 5 he began taking active steps to achieve it. He applied to three graduate schools and scheduled his TOEFL Examination. These actions indicate that he had moved from the Preactional Stage to the Actional Stage of Dörnyei’s (2005) Process Model of L2 Motivation. Moreover, his CAF scores also reflected a change as discussed in the previous section. During this period, he began to focus on grammatical accuracy. While Juan did not explicitly study vocabulary
in the previous months, he made active efforts to focus on grammar. He asked for help preparing for the TOEFL Examination, brought examples of speech from his community that he was struggling to understand (phrasal verbs), began using colloquial speech, and adding past, past progressive, and future progressive verbs to his samples. Thus, he appeared to be paying attention, actively trying to adapt to his speech community, and increasing his efforts to improve his grammar.

Beginning in month 10, Juan seemed to decrease his efforts toward grammatical accuracy and integration. The correlations between accuracy and fluency returned to their previous state (a positive correlation), and accuracy began a period of fluctuation. Month 10 occurred in December, thus he was concentrating on his end of semester tasks and preparing to return home for a month of holidays. Upon his return in month 12, he told me that he and his wife were expecting a child. He also commented on how disappointed he was that he still could not understand people outside of the university. By month 13, he had decided not to pursue his doctorate. After month 12, he commented that he was disappointed in himself that he had not improved his speaking more. He also mentioned that his L1 English friends were speaking Spanish to him and that he was not enjoying interacting with people in English. In contrast, he mentioned that his writing had improved. Thus, he had re-evaluated his situation and began looking for causal reasons for failures and successes, placing him in the Postactional Stage of Dörnyei’s (2005) Process Model of L2 Motivation. While he stated emphatically that he wanted to learn as
much English as possible before he left, he had lost the reasons for his integrative motivation and his Ideal L2 Self no longer included a doctoral program in the US.

It also appears that Juan’s ought-to self had replaced the ideal L2-self. In addition to the pressure and blame he placed on himself, he mentioned that he wanted to make his family proud and that his wife had high expectations of him. During his last month, he stated the following:

so yeah they they [parents] have these kind of feeling of proudness? … and my wife is like (. ) you have to do this … she she was one of the first (. ) yeah the main person to say you have to do it even if you don't want to ( . ) you have to do it you have time there are many people that want to do that and they don't have this opportunity. (Session transcript, April 25, 2014, Month 15, Appendix F).

He also shifted his focus to academic writing and instrumental goals including his master’s thesis. He appeared to have entered a new Preactional Stage, but this one focused on academic English and writing. When asked about his motivation after the study, Juan agreed that his focus most likely changed from speaking to writing. As such, he concentrated less on speaking and less on experimenting with new forms in his speech. Consequently, the correlations between the CAF variables reflected decreased effort while producing them.

These results indicate Juan’s CAF variables were affected by changes in his motivation, and thus that his language system reorganized based on changes in his social ecosystem. The results also indicated that native-speaker exposure and interaction appeared to affect Juan’s CAF scores and were dependent on motivation. Even though Van Dijk, Verspoor, and Lowie (2011) point out that learners experience cyclical periods
of attrition and progress, which may explain the changes in Juan’s data without considering motivation, the coincidence of these motivational factors occurring so closely to the changes in his scores is difficult to ignore.

Based on the data, it appears that the Juan’s Ideal L2 self triggered a period of progress, whereas his decision not to pursue his doctorate triggered the ought-to self, thus changing his focus. Moreover, van Dijk et al. (2011) suggest that effects of the environment are intensified on a learner moving between stages of learning. They stated:

> different environmental factors may affect the scores, but it is reasonable to assume that these different environmental factors have a greater effect on someone who is in the process of moving from one stage to the other and less on the individual who has reached a more stable state. (p. 61)

Thus, Juan’s developing language may have made his language more susceptible to the effects of his environment.

Campbell & Storch (2011) and Shoaib & Dörnyei (2005) suggest that factors mentioned by the learner may have more saliency than standardized surveys such as the AMTB (Masgoret & Gardner, 2003). Moreover, case studies are able to focus the individual learner’s circumstances and timing, whereas a survey given to large groups of learners may not capture the same microgenetic details. Although this study’s results cannot be generalized to a group of learners, the results suggest that motivational factors may change the course of a learner’s CAF trajectory. Juan’s results may be particularly interesting to study abroad researchers investigating the effects of motivation on approaching return dates, which appear to have influenced Juan’s integrative motivation and his focus on grammar.
6.1.4 Research Question 4: Task differences

Only 3 out of 14 variables had significant differences between Juan’s speech samples of the frog stories, Mr. Bean videos, picture descriptions, and conversation samples: lexical diversity, seconds per silent pause, and articulation rate. However, comparison group’s results and the descriptive statistics from Juan’s data show patterns of differences between the tasks. This section will discuss the results related to each task.

The frog stories resulted in the least errors and lowest lexical diversity in both sets of data. However, subordination was the lowest in Juan’s data and the highest in the control group. These results support Robinson’s (2011) argument that easing the cognitive load of a task does not necessarily result in increasing the complexity of the speech. The frog stories’ plots are simple, but do contain simultaneous events, which have been shown to increase subordination (Foster & Tavakoli, 2009; Tavakoli & Foster 2011). This is likely why the control group’s scores were high. However, if the task did not challenge Juan, he may have chosen to concentrate on simple grammatically correct sentences rather than detailed description.

In Juan’s data, Mr. Bean videos produced the slowest speech production (articulation rate) and the most disfluencies. However, articulation rate may be a problematic measure in that these variables could be considered conflated. Redundancies, false starts, and reformulations were not removed from the syllables per second. Therefore, instead of producing speech at a slower rate, he may have been producing repair disfluencies at a slower rate. Contributing to this point, the comparison group
produced the fastest articulation rate and the fewest repair disfluencies with this task.

Contrary to Yuan and Ellis’ (2003) findings, time pressure that should have been added with the Mr. Bean videos did not appear to have an effect on Juan’s accuracy. His accuracy scores were not considerably different than in the picture descriptions where no time pressure was applied or in the spontaneous conversation where time pressure was applied in the form of dialogic speech.

As Juan mentioned, the picture descriptions were difficult for him and his scores reflect that difficulty. He was speaking quickly (highest articulation rate), yet produced less elaborate speech, and paused silently more often and for longer periods. Both Juan and the comparison group produced the most lexically diverse speech in this task indicating a more diverse vocabulary was necessary to complete it. Moreover, the comparison group had significantly more pauses and disfluencies in this task than the others, which indicated it was more difficult for them as well.

Finally, during spontaneous conversation, Juan scores were better than the other tasks in seconds per silent pause, repair disfluencies, elaboration, and subordination. Spanish relies more heavily on subordinate clauses than English (Whitley, 2002), this may indicate that while he was speaking freely, he was translating more directly from his L1. However, he did not appear to be overusing subordinate clauses because he remained close to the 95% confidence interval of the comparison group. With regards to pauses, the dialogic nature of the conversation most likely shortened the length of silent pauses in that I probably said something if the period seemed too long.
It is noteworthy that articulation rate is the only variable where task differences were evident after converting the raw data to 0 – 1 values and z-scores. The frog stories and conversations were consistently spoken faster than the Mr. Bean videos and picture descriptions despite converting the data to standardized scales before combining them. Even though lexical diversity and seconds per silent pauses also showed significant differences in Juan’s task difference analysis, only articulation rate appears to have maintained that effect during the development analyses. This may indicate that this variable is more sensitive than the others to task effects and its use should be considered carefully when using multiple tasks in research.

6.2 Theoretical and Methodological Implications

6.2.1 CDST Tools for Analysis

CDST research follows the premise that language learners’ systems change and self-organize. Consequently, a change in one variable could affect the entire system. (Beckner et al., 2009, Larsen-Freeman, 1997; Verspoor et al., 2011). CDST researchers have developed tools to analyze non-linear changes and relationships between variables, specifically min-max graphs and moving correlations. This section will discuss how both were used and expanded in this study.

Min-max graphs show periods of stability and fluctuations in scores by focusing on the range of the scores, or window created by the min-max lines. Theoretically, fluctuations indicate a period of restructuring. If the period of fluctuation is followed by a
higher or lower stable score, it is referred to as a phase shift. The results of this study show greater success finding phase shifts with the specific variables over the global variables. While both this study and Polat & Kim (2014) set out to find phase shifts, neither study was successful with global CAF variables. However, the analyses of specific errors (past tense and progressive aspect) revealed patterns that resembled a phase shift (present progressive) and resembled an unsuccessful phase shift (past tense). Moreover, the results of Juan’s accuracy analyses indicate that periods of fluctuation in specific variables and global variables occurred at different times. While Juan’s specific accuracy variables (past tense and progressive aspect) fluctuated during months 5 – 8, global accuracy did not fluctuate until month 7. Juan’s fluency and accuracy correlation changed from months 5 – 10, or right before accuracy began fluctuating. Hence, it appears that specific variables may produce better results while exploring phase shifts and different results with periods of fluctuation. Future research on global and specific variables may provide further understanding of how phase shifts and fluctuation apply to language development.

In this study, moving correlations were used to provide insights into the relationships between variables and the changes in those relationships. Verspoor & van Dijk (2011) defined three types of relationships between variables, referred to as connected growers, two of which were found in this study. Negative correlations (indicating competitive growers) were found between lexical diversity and elaboration and between elaboration and accuracy. Therefore, as one of these variables increased the
other decreased and vice versa. Competitive growers indicated that Juan was switching his concentration between the two variables. No correlation was found between lexical diversity and accuracy, indicating that these variables developed independently in Juan.

The term connected growers was created for first language acquisition analysis. Thus, scholars assume the language is “growing” at a rapid rate. However, this study found that the concept can be applied to adult language as well, which may be changing but is not developing at a rapid rate. In the native-speaking comparison group a weak positive correlation (cooperative) was found between lexical diversity and elaboration (Juan’s was negative); and a weak negative correlation (competitive) was found between lexical diversity and accuracy (Juan’s had no correlation). In the case of native speakers, the variables are not growing together, but have a relationship related to their usage. It may be useful to consider that difference, when studying advanced language learners in future research.

This study was unique in that moving correlations were also used to investigate changes in the relationships between variables over time. While correlations between the language knowledge variables (accuracy, lexical diversity, and elaboration) remained mostly consistent, the correlations between the knowledge variables and the proceduralization variables (pauses, repairs, and speech rate) changed in patterns related to Juan’s increased efforts. Further evidence of increased effort was explained by the relationships between the periods of fluctuation. Immediately before or immediately after a period of fluctuation, the correlation changed. As theorized by CDST, these fluctuations
indicate an attempt to restructure. Thus, Juan appears to be making an effort to experiment with new forms immediately before or after attempting to restructure.

Lastly, the correlations coincide with specific periods in Juan’s motivation. As he was in the Preactional Stage of Dörnyei’s (2005) Process Model of L2 Motivation, he was focusing on vocabulary. As he entered the Actional Stage (taking active steps toward achieving his goals), his focus shifted to accuracy. As he finished the cycle in the Postactional Stage (evaluating the situation and his goals), his focus shifted to elaboration. The system develops logically in that Juan needed vocabulary to function in his initial situation. When he began to actively work toward his goal, his focus shifted to the most salient aspect of learning, grammatical accuracy. Just as he was taught in school, he could learn a grammar form and move to the next form. During the last few months, he shifted to elaboration, or created longer clauses, which is a skill that merges vocabulary and grammar.

As mentioned previously, the measures of moving correlations and min-max graphs were successful in showing fluctuations in the developing variables and changing relationships between variables. CDST studies often include analyses that smooth the data (moving averages), detrend the data (mathematically removing trends that develop over time), and check for statistical significance (the Monte Carlo Analysis) (Verspoor et al., 2011). The purpose of these analyses is to test for significant variance or to establish stable relationships between the variables. The purpose of this study was to look at changes over time and changing relationships between the variables. Therefore, based on
the research questions and the exploratory nature of the study, statistical significance was not the primary focus. While inferential statistics might have shown that the fluctuations and/or the correlations were not due to chance, detrending the data would have removed information regarding changes over time. As Spoelman and Verspoor (2010) found, a general correlation between accuracy and complexity may not show significant results, yet a moving correlation of the same data can show periods of strong correlations alternating over time. Holland (2006) reminds us that “exploratory computer-executable models define possibilities not actualities” (p. 3). Rather than focusing on statistical significance, this study used micro-analyses of Juan’s language in the transcripts and descriptive statistics (such as the range of the fluctuations) to establish the validity of the results. Similar analyses using descriptive rather than inferential statistics were also done by Polat & Kim (2014) and Vyatkin (2012).

6.2.2 Comparison group

Including comparison or control groups in second language development studies has become a controversial topic. The arguments are generally based on two assumptions: a) the control group represents the goal, which is often unachievable, and b) that variation within the control group is minimal. This study assumed neither. In this case, the comparison group represented Juan’s speech community, which represented the language he was exposed to every day. The original premise of this study was to include the range as well as the 95% confidence interval, thus including a broader scope of
influences with which Juan may have interacted. However, within a month, Juan was near or within the 95% confidence interval on most variables. In some cases, his scores moved beyond the comparison group’s 95% confidence interval and began regressing toward the comparison group, but not below it. As such he may have been drawn to a new attractor state then fluctuated around that attractor state as he stabilized. Without the comparison group’s example representing a possible new attractor state, his worsening scores could have been interpreted as attrition.

The idea of using a speech community rather than a general standard for all English speakers holds particularly true regarding fluency, which varies considerably even within the United States (Lippi-Green, 2012). For example, if Juan had studied in the South rather than the Northeast, his speed fluency examples would have been slower. By limiting comparisons to his speech community rather than a generalized norm for all English speakers, the standards for the community can vary, as they should. Juan’s data showed that his language was affected by his speech community in that colloquial speech appeared in his transcripts. Moreover, colloquial speech appeared while Juan was actively increasing his efforts to integrate and improve his grammar. However, after his goals became less integrative and more instrumental, the colloquial speech disappeared. Since his ideal L2 self no longer included living and working in the US, integrating into the community was not as important or of interest to him. Including comparison group data strengthened this argument.
6.3 CAF Methodological Considerations

Current literature in the areas of CAF expresses concerns with the definitions, functionality, and validity of the CAF variables (Biber et al, 2011; Housen & Kuiken, 2009; Norris & Ortega, 2009). Since definitions of these constructs vary, current researchers are working towards operationalizing the CAF variables as well as experimenting with new variables (Bulté, & Housen, 2014; Skehan, 2014). This study contributes to that research by including multiple variables from previous research and thus experimenting with their success in measuring development over time, an under-researched topic in CAF. This section will discuss this study’s successes and limitations related to each of the CAF constructs.

6.3.1 Complexity

For this study, subordination (clause per AS-unit), elaboration (words per AS-unit), and lexical diversity (D) were used to measure complexity. This section discusses issues related to measuring elaboration, subordination, and D.

The first issue relates to the unit of measurement, AS-units. Since this study divided the speech into AS-units, coordinating conjunctions were not analyzed. However, their presence created an issue regarding the division of AS-units. Foster et al. (2000) stated that a new AS-unit should begin after a coordinating conjunction that is between complete thoughts. They also suggested that if the speaker pauses for more than .4 seconds between compound verbs or nouns and after or before a coordinating
conjunction, a new AS-unit should be started since the speaker has most likely started a new thought. In this study, that division often resulted in AS-units without a subject and/or auxiliary verbs resulting from pauses between a list of actions. As such, they were counted as grammatical errors. Omission of subjects and auxiliary verbs were the most common errors in the comparison group, resulting in an average of 1.23 omitted subjects and .43 omitted auxiliary verbs per 100-word transcript. For example, one participant said, “and his parents don’t look too happy <that> um (.) that some I guess that he’s pointing (pause longer than .4 seconds) and making a scene about it” (CG97FGTD100W2). Based on Foster et al. (2000), “and making a scene about it” is a new AS-unit and was coded with two errors, omission of the subject and omission of the auxiliary verb error. While long pauses after a coordinating conjunction could indicate a new train of thought and merit a new AS-unit, confounding factors that cause strain on cognitive resources also resulted in long pauses while completing a thought. An area for future research might include a method for distinguishing pauses created by cognitive strain and pauses indicating a new theme.

The next consideration relates to elaboration. Several scholars argue that nominalization overtakes subordination in intermediate to advanced learners and therefore subordination may not be an accurate measure for complexity with advanced learners (Lambert & Kormos, 2014; Norris & Ortega, 2009; Ortega, 2012). However, those findings were related to written data, whereas Chan et al. (2015) found subordination to increase in oral production of low to intermediate level learners. Since
little data has been collected on oral samples with intermediate to advanced learners, both subordination and elaboration were analyzed. The findings correspond with Chan et al. (2015) in that Juan’s trendlines showed clear linear improvements in subordination (see Figure 5). Whether developing subordination in this study relates to differences in oral and written data, Juan’s proficiency level, L1 transfer, or Juan’s individual speech patterns is indeterminable. Future studies of small groups including both the oral and written data of learners at different levels and language backgrounds may be able to answer this question.

While subordination conflates with the measure of elaboration, Juan’s elaboration scores developed in a different pattern than his subordination scores. As the trendlines on Figure 15 indicate, his elaboration scores decreased until month 8, then began increasing whereas subordination maintained a steady increase throughout the entire study (see Figure 5). Norris & Ortega (2009) argued that elaboration increases in complexity in the form of prepositional phrases, adjectives, adverbs, and non-finite verb phrases. Juan’s data concurs with their argument. After one year of immersion, Juan added two prepositional phrases and one temporal connector to his description of the same event (see examples e and f, section 5.2.2.1). However, the results also suggest that elaboration included more than descriptors and clauses. Juan also doubled the number of function words from example e to example f. While function words are an integral part of development, the complexity literature generally emphasizes description words and phrases or advanced verb forms. Consequently, researchers using elaboration as a
measure of development may need to include a micro-analysis of their data for specific measures of function words, descriptors, and clauses to investigate why their participant’s elaboration scores are increasing.

Lastly, although $D$ was chosen due to its success in other research, Juan’s increasing use of function words also seemed to distort the findings. While Juan’s lexical diversity decreased at the end, he may have just been increasing the number of function words. There are two options that could have eliminated this issue. The first would be to use CLAN’s output listing of the frequency of each word to eliminate the function words before calculating $D$. Secondly, Jarvis (2002) found that TTR curves created with Uber U worked successfully with content words. While this measure is not readily accessible in CLAN, research specifically dedicated to lexical diversity may find it worth the effort to calculate lexical diversity using the formula for Uber U.

6.3.2 Accuracy

The analysis of the global accuracy variable (error per AS-unit) was useful in showing periods of increased focus on accuracy and the competitive relationships between accuracy and elaboration. However, as Norris and Ortega (2009) point out, global variables such as errors per AS-unit, fail to capture developmental advances. For example, Juan began using the simple past tense, future progressive aspect, and past progressive verb aspect. As he was experimenting, he committed more errors. Without a specific error analysis, the data only shows that he committed more errors which would
only be interpreted as worsening scores. Moreover, the specific error analysis seemed to capture a phase shift among the present progressive aspect. Therefore, scholars may need to carefully consider what they are trying to capture before choosing between global and specific accuracy variables. If they are looking for relationships between variables or cognitive strain, global variables seem to serve the purpose. However, if they are analyzing development, specific variables may provide more informative results.

Part of the effectiveness of the specific error analyses in this study can be attributed to the consistency and validity of the CHAT codes (see Appendix B). Once coded in CHAT, CLAN provided easily interpreted output that indicated where closer analysis was needed. However, there was one exception. The code [+gram] seems limited in that it includes all sentence structure errors. Additional subdivisions in CHAT such as conditional structure, word order, extra prepositions or connectors, and question inversion errors would have increased the effectiveness of this analysis.

6.3.3 Fluency

The purpose of this study’s fluency analyses was to contribute to understanding how L2 knowledge development (accuracy and complexity) interacts with proceduralization development (fluency). This section discusses the additional contributions to the understanding of the subdivisions of fluency and some possible areas of concern.
The trendlines of filled pauses, silent pauses and total pauses showed linear improvement, indicating consistency within the breakdown disfluency construct. However, the constructs of repair disfluencies and speed fluency were inconsistent. While false starts (repair) showed a linear improvement, redundancies (repair) and speech rate (speed) improved, and then regressed. Meanwhile, reformulations (repair) and articulation rate (speed) fluctuated too much to establish a pattern. These mixed results could indicate that Juan had an individual preference for which disfluencies to commit while his cognitive resources were strained. Towards the middle of the study, Juan developed a noticeable habit of stuttering (redundancies) which was not present when he spoke Spanish. This seems to indicate that he had an unconscious preference for this form of disfluency since it diminished when he entered the Postactional Stage of motivation and was no longer striving to become his Ideal L2 Self. However, it may also indicate that the variables within the construct are not as strongly related as researchers originally theorized or are too conflated with other fluency variables to be considered separately. For example, speech rate contains pauses and repairs, yet is considered part of the speed fluency construct. These results indicate a need for increased research into individual differences related to fluency. Perhaps learners select a certain disfluency (most likely subconsciously) that becomes their habit while under cognitive strain and perhaps it changes as they gain proficiency. Continued longitudinal case studies may provide answers to these questions.
6.4. Limitations of the Study

As all case studies, the results of this study reflect on individual’s development and are therefore limited in generalizability. Juan was older than many language students (33 – 34 years old), well educated, male, and introverted. In addition to individual differences, the study is limited to L1 Spanish and L2 English, which are both Indo-European languages using Latin-based scripts. In studies involving different languages (both L1 and L2), the students may have crosslinguistic differences that result in different learning trajectories. Those differences might include grammatical forms (noun declensions, classifiers, tenses), written scripts (Cyrillic, Arabic, Chinese) or phonetic structures (consonant clusters, tones, clicks). They may also may have cross-cultural differences that affect development such as gender roles and forms of politeness. As CDST studies increase, research should address these differences. However, the data is generalizable in that it shows validity to the theoretical framework. Juan’s development was non-linear, interconnected, and influenced by the social ecosystem. As such, it adds evidence to a growing set of data analyzed with CDST on those generalizable principles.

Unlike traditional experimental research, longitudinal research requires repeated data collection, which may be interrupted by events in the participant and researcher’s lives. The second limitation relates to scheduling. Juan and I met once a week as often as was feasible for both of us, but due to life issues, semester breaks, and scheduling conflicts, even meeting every two weeks was not always possible. Consequently, there were three gaps of one month and inconsistent intervals in the data collection. Moreover,
the original research design included tasks created to measure cognitive restructuring. Since cognitive restructuring was not included in this study, days in which only cognitive restructuring data were collected caused additional gaps in data collection.

Nevertheless, CDST posits that the learner’s entire ecosystem affects language development. Events that lead to gaps in data collection may also lead to additional factors of interest. For example, during month 10 (December) sessions were limited due to end of semester demands. Juan’s level of integration also began declining during this period. While his interest in integrating and his changing CAF scores could have lessened due to other reasons, the coincidence between end of semester stress and integration was a logical factor to investigate.

Lastly, two issues arose related to task differences. First, the tasks were not distributed evenly throughout the study with the exception of the four conversation samples. For example, the picture description task was added in month 5 of the study. The second issue relates to pauses. In order to lessen the researcher’s influence on the data, the pauses caused by page turns in frog stories were deleted from the data. Those pauses could have provided the speaker additional time to predict what would happen next and plan his turn, which may have resulted in increased speed compared to the other tasks. While the adjustments made with 0 -1 values and z-scores lessened the differences between tasks, the potential for influence remains.
6.5 Conclusions and Future Directions for Research

This study contributes insight to the field of second language development by following a native speaker of Spanish learning English while living in the United States for 15 months. By analyzing the CAF variables under the lens of CDST, the results reveal that Juan’s development was non-linear, interconnected, self-organizing, and influenced by both the cognitive and social ecosystem. Yet, as discussed in the previous section, many questions remain, particularly in the area of methodology. As a summary, this section highlights three problematic areas of analysis, evaluates two CDST concepts in need of further research, discusses the benefits of including comparison groups, and emphasizes the need for continued longitudinal second language development studies.

Three types of speech phenomena complicated these analyses, which may indicate the need for further investigation. The first, function words, affected elaboration and $D$ scores. Juan’s increasing ability to use function words automatically led to increased elaboration scores and decreased $D$ scores. Unfortunately, neither measure was designed to account for this phenomenon. Methods that accurately and efficiently remove functions words before the analysis, or control for them before calculating the scores would increase these measures’ effectiveness. Next, coordinating conjunctions were problematic while dividing AS-units. The most obvious solution is to qualitatively look at the sentence. However, doing so decreases the objectivity of the measure, which Malvern and Richards (2002) were trying to avoid. Experiments that include carefully defined subjective measures may provide a compromise. Lastly, Juan’s results in the area of
subordination contradicted current research’s finding that intermediate to advanced learners will switch from subordination to nominalization. The difference seems to be related to two possible causes. Since Spanish speakers use more subordinating clauses than English, Juan may have been able translate more literally as his language skills improved. Another option could be the differences between oral samples and written text. This study analyzed oral samples, whereas most complexity research analyzes written texts. While some research suggests nominalization is more common in written texts, a definitive conclusion has not been made. Continued research into cross-linguistic influence or the differences between oral and written language may provide answers.

Since CDST studies are still in the process of applying methods to theory, it is not surprising that a few issues arose in applying CDST’s tools to CDST’s concepts. For example, a complete phase shift was not found with the global variables despite their clear periods of fluctuations. The specific variables led to greater success by utilizing both usage and errors to examine the fluctuations. Moreover, I chose errors based on the CLAN output of all errors rather than investigating variables that I suspected may have experienced a phase shift. In the global variables, fluency decreases were found immediately before periods of fluctuation which appears to indicate experimenting while learning followed by restructuring. At this point, research regarding phase shifts make sense in theory, but seem limited in our understanding of how they work. Continued longitudinal work and work with denser and more consistent data may result in capturing a phase shift or at least increasing our understanding of how phase shifts happen.
The data from this study also contributes to the understanding of connected growers and may suggest a need to expand the term, or create a new term. While consistent correlations were found between fluency and the knowledge variables, the changing relationships found with moving correlations does not fit the concept of connected growers. Since the changes in correlations appear to be the effect of cognitive strain, research may benefit from separate label such as cognitive strain relationships. Additionally, correlations between the knowledge variables were found in native-speakers, whose speech may be growing, but at a much lesser rate than a non-native speaker. Rather than a grower relationship, it may be more useful to think of correlations among native-speakers as a speech community tendency.

The next point also relates to the comparison group. While the concept of control groups has been shown to be problematic, native-speakers affect L2 learners that interact with them. Rather than the comparison group representing a goal, CDST allows for native-speakers to be part of the social ecosystem and therefore an influence rather than a goal. While L2 students may never achieve native-like speech (or want to), native-speakers represent a significant amount of immersed students’ language examples and implicit (and sometimes explicit) instruction. As such, incorporating them into CDST studies of immersed learners may add insight to the learners’ development. For example, Juan’s development appeared to show attrition after the 8th month in a few variables. However, he had surpassed the 95% confidence interval of native-speakers and was
moving back towards it. Thus, he might have been fluctuating above and below the native speakers similarly to what might be expected from an attractor state.

Although CDST argues that each learner’s trajectory is unique and dependent on their initial state, continued research will most likely lead to some emerging trends across learners. To find such generalizable data, future studies will benefit from including a variety of initial states such as various first and second languages (and third and fourth), ages, genders, socio-economic levels, education levels, proficiency levels, and learning situations. Although each learner’s trajectory will be unique, we may find trends in how phase shifts develop, how the CAF variables interact, how knowledge and proceduralization develop together, and how changes in the social and cognitive ecosystems interact with language development. In addition, researchers working on operationalizing the CAF measures would benefit from continued longitudinal studies to provide contrast to the current experimental and cross-sectional research. Lastly, any research utilizing narrative stimuli, would benefit from continued research of task differences including more types of stimuli. Such analyses could provide information to increase the comparability of future studies. Moreover, longitudinal studies, which are suited to using a variety of stimuli to prevent practice effects, would benefit from knowing the expected differences of multiple stimuli.

In sum, future CDST, CAF, and task-based research would benefit from additional longitudinal studies in order to understand second language development and the variables that affect it. By understanding how second language develops as a system,
we may be able to provide learning materials and methods suggestions to teach learners with different cognitive strengths and learning circumstances, who develop in one area, yet lack in another.
REFERENCES CITED


Bradbury, R. (1952). A Sound of Thunder in *R is for Rocket*. New York: Doubleday. Retrieved from: [https://docs.google.com/document/d/1XFtrc-PgR8XPbKtU5j-HnzYNydHbub-Q9EEnomNO8CI/edit](https://docs.google.com/document/d/1XFtrc-PgR8XPbKtU5j-HnzYNydHbub-Q9EEnomNO8CI/edit)


Mr. Bean (2009a). Mr. Bean at the dentist. From Official Mr. Bean. Retrieved from: [https://www.youtube.com/watch?v=lW6R9kSGV2Q](https://www.youtube.com/watch?v=lW6R9kSGV2Q)

Mr. Bean (2009b). *Mr. Bean at the swimming pool*. From Official Mr. Bean. Retrieved from: [http://www.youtube.com/watch?v=gZujYUcY5xc](http://www.youtube.com/watch?v=gZujYUcY5xc)


Mr. Bean (2010b). Mr. Bean – Dancing at a Nightclub. From *Mr. Bean Goes to Town*. Retrieved from: [http://www.youtube.com/watch?v=iT_IiHsZtWg](http://www.youtube.com/watch?v=iT_IiHsZtWg)


APPENDIX A

MATERIALS FROM PICTURE DESCRIPTION SETS 1-3

Note: Permission for using these photos is granted through Title 17 US Code Section 107: Limitations on exclusive rights. Fair use (Cornell, 2012)

*Set 1:*

[Images of various photos from Set 1]
Appendix A: Set 2
Appendix A: Set 3
#### APPENDIX B

**CHAT CODES FROM MACWHINNEY (2000)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Error</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic Error</td>
<td>Related word</td>
<td>[* s:r]</td>
</tr>
<tr>
<td></td>
<td>Unrelated Word</td>
<td>[* s:u]</td>
</tr>
<tr>
<td>Morphological</td>
<td>Agreement</td>
<td>[* m:a]</td>
</tr>
<tr>
<td></td>
<td>Agreement specific:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Omission 3rd person singular s</td>
<td>[* m:a:0es]</td>
</tr>
<tr>
<td></td>
<td>Omission s on plural noun</td>
<td>[* m:a:s]</td>
</tr>
<tr>
<td></td>
<td>Other morphological errors:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error with progressive</td>
<td>[* m:a:-ing]</td>
</tr>
<tr>
<td></td>
<td>Error with past</td>
<td>[* m:a:ed]</td>
</tr>
<tr>
<td></td>
<td>Error with perfective</td>
<td>[* m:a:en]</td>
</tr>
<tr>
<td></td>
<td>Error with possessive</td>
<td>[* m:a:’s]</td>
</tr>
<tr>
<td>Formal Lexical Device</td>
<td>Zero article should be the</td>
<td>[* f:a:0:d]</td>
</tr>
<tr>
<td></td>
<td>Zero article should be a/an</td>
<td>[* f:a:0:i]</td>
</tr>
<tr>
<td></td>
<td>Definite article should be</td>
<td>[* f:a:d]</td>
</tr>
<tr>
<td></td>
<td>indefinite</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indefinite article should</td>
<td>[* f:a:i]</td>
</tr>
<tr>
<td></td>
<td>be definite</td>
<td></td>
</tr>
<tr>
<td>Missing word</td>
<td>Coded by part of speech</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subject</td>
<td>0subj</td>
</tr>
<tr>
<td></td>
<td>Auxiliary verb</td>
<td>0aux</td>
</tr>
<tr>
<td></td>
<td>Verb</td>
<td>0v</td>
</tr>
<tr>
<td></td>
<td>Article</td>
<td>0art</td>
</tr>
<tr>
<td></td>
<td>Preposition</td>
<td>0prep</td>
</tr>
<tr>
<td></td>
<td>Object of transitive verb</td>
<td>0obj</td>
</tr>
<tr>
<td></td>
<td>Object of preposition</td>
<td>0p0bj</td>
</tr>
<tr>
<td>Sentence Structure</td>
<td>Errors related to sentence</td>
<td>All coded</td>
</tr>
<tr>
<td></td>
<td>structure</td>
<td>as [+gram].</td>
</tr>
</tbody>
</table>
### APPENDIX C

#### RESULTS FROM ERROR ANALYSIS

<table>
<thead>
<tr>
<th>Task</th>
<th>Semantic related</th>
<th>Semantic Unrelated</th>
<th>Morph Agreement</th>
<th>Omit 3rd Person Singular</th>
<th>Omit “s” on Plural</th>
<th>Past Tense</th>
<th>Perfect Aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1BDF1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1BDF2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1Bogota</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1Colombia</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1Dico1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1Dico2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2OFTM1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2OFTM2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5FWAY1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5FWAY2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5Set11</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5Set12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5Mother</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5Roommate</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7SWP1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>7SWP2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8FOHO1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8FOHO2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8SET21</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8SET22</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>9Horror1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9Horror2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10Compose</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10Mentor</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11Set31</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11Set32</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>12FGTD1</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12FGTD2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13BDF3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13BDF4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14OFTM3</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14OFTM4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15Politics</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15Think</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15Dentist1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15Dentist2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
## Appendix C: Results from error analysis continued

<table>
<thead>
<tr>
<th>Task</th>
<th>Omit Article</th>
<th>Omit Subject</th>
<th>Omit Aux Verb</th>
<th>Omit Verb</th>
<th>Omit Preposition</th>
<th>Omit Object</th>
<th>Omit Particle</th>
<th>Sent Struct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dentist1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dentist2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BDF1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BDF2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Bogota</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Colombia</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Disco1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Disco2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OFTM1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OFTM2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FWAY1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FWAY2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Set11</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Set12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Mother</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Roommate</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>SWP1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SWP2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FOHO1</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FOHO2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SET21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SET22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Horror1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Horror2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Compose</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Mentor</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Set31</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Set32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FGTD1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>FGTD2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>BDF3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BDF4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>OFTM3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>OFTM4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Appendix C: Results from error analysis continued

<table>
<thead>
<tr>
<th></th>
<th>Politics</th>
<th>Think</th>
<th>Dentist1</th>
<th>Dentist2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politics</td>
<td>1 0 0 0 0 0 1 0 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Think</td>
<td>0 1 0 0 0 0 0 0 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dentist1</td>
<td>1 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dentist2</td>
<td>1 1 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D

MOTIVATIONAL FACTORS TIMELINE

<table>
<thead>
<tr>
<th>Date</th>
<th>Proficiency Measures</th>
<th>Integrative Factors</th>
<th>Instrumental Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td></td>
<td>Studied in Boston for 3 months</td>
<td>Juan’s university offers Fulbright scholarship. In order to qualify, he must have 80 on the iBT TOEFL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Won scholarship for study in Boston</td>
</tr>
<tr>
<td>2012</td>
<td>Studies at the Berlitz Institute in Colombia – says he learned a lot.</td>
<td>Attended 12-week program in Tempe, AZ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>80/120 TOEFL iBT (equal to B1 on CEFR)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>August 2013</td>
<td>Arrives in Philadelphia as master’s student</td>
<td>Living on campus Attending classes in English</td>
<td>Awarded Fulbright Scholarship</td>
</tr>
<tr>
<td></td>
<td>Reports only speaking Spanish once a day with wife on Skype</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Month 1 - sessions with researcher begin</td>
<td>Oxford Placement Examination B1- 35/60 (intermediate)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix D: Motivational Factors Timeline Continued

<table>
<thead>
<tr>
<th>Date</th>
<th>Proficiency Measures</th>
<th>Integrative Factors</th>
<th>Instrumental Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month 2</td>
<td>Oxford Placement Examination</td>
<td>Returned to Columbia for visit, spoke only</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B2 - 40/60 (intermediate high)</td>
<td>Spanish</td>
<td></td>
</tr>
<tr>
<td>Month 4</td>
<td></td>
<td>Parents visit.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Translated for them.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Felt the transition back was easy</td>
<td></td>
</tr>
<tr>
<td>Month 5</td>
<td>Nation (2001) Vocabulary Test:</td>
<td>Reports that English has affected his Spanish</td>
<td>Mentions he is considering applying to doctoral programs.</td>
</tr>
<tr>
<td></td>
<td>2000-word level</td>
<td>writing skills.</td>
<td>Plans to be here 3 more years.</td>
</tr>
<tr>
<td></td>
<td>– 100%</td>
<td>Colloquial speech appears in his transcripts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3000-word level</td>
<td>(see Table 8).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26/30</td>
<td>Requests help with pronunciation</td>
<td>Struggling with phone calls.</td>
</tr>
<tr>
<td></td>
<td>5000-word level</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28/30</td>
<td>Having problems adjusting to lack of</td>
<td>Working on vocabulary during our sessions.</td>
</tr>
<tr>
<td></td>
<td>Academic</td>
<td>formality in English.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
<td>Thinks his professors are angry because they</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28/30</td>
<td>are so direct.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10,000 word level</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>level 20/30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix D: Motivational Factors Timeline Continued

<table>
<thead>
<tr>
<th>Date</th>
<th>Proficiency Measures</th>
<th>Integrative Factors</th>
<th>Instrumental Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month 7</td>
<td></td>
<td>Moves to off-campus apartment, mostly English speakers and friendly doorman</td>
<td>Applies to three doctoral program (Princeton, Columbia and Brandeis)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Only Spanish speaker in most of his classes, but befriends a Puerto Rican in one class and an English speaker who knows some Spanish</td>
<td>Struggling with phone calls to utilities companies – requests practice</td>
</tr>
<tr>
<td>Month 8</td>
<td></td>
<td>Requests help with phrasal verbs, brings samples from his adviser.</td>
<td>Registers for TOEFL exam. Requests help preparing.</td>
</tr>
<tr>
<td>Month 9</td>
<td>TOEFL iBT scores 85/120 (B1 CEFR)</td>
<td></td>
<td>Frustrated with TOEFL score. Needed 100 for Temple’s doctoral program.</td>
</tr>
<tr>
<td>Month 10</td>
<td></td>
<td>Discusses compositions with people at the concert</td>
<td>Performance of two of his compositions</td>
</tr>
<tr>
<td>Month 11</td>
<td></td>
<td>Returns to Colombia. Wife returns with him for a week</td>
<td>Found out that he and wife are expecting baby. (Motivational transformational episode)</td>
</tr>
</tbody>
</table>
Appendix D: Motivational Factors Timeline Continued

<table>
<thead>
<tr>
<th>Month</th>
<th>Proficiency Measures</th>
<th>Integrative Factors</th>
<th>Instrumental Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month 12</td>
<td>Oxford Placement Exam B2- 39/60 (Intermediate high)</td>
<td>Reports that L1 English friends are speaking to him in Spanish.</td>
<td>Says he will continue if accepted into a doctoral program.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conflicted about continuing in doctoral program</td>
<td>Year evaluation – proud that he speaks more in our sessions. Struggling with complex math</td>
</tr>
<tr>
<td>Date</td>
<td>Proficiency Measures</td>
<td>Integrative Factors</td>
<td>Instrumental Factors</td>
</tr>
<tr>
<td>Month 13</td>
<td>Returned to Colombia</td>
<td>Says his writing style has changed due to English</td>
<td>Decision to return to Colombia is made</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Says he wants to learn as much as he can before he leaves</td>
<td></td>
</tr>
<tr>
<td>Month 14</td>
<td>Oxford Placement Exam B1 – 38/60 (intermediate)</td>
<td>Reports thinking in English for basic things, but still translates for academia.</td>
<td>Finds out he was not accepted in doctoral program</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Colloquial speech has reduced (see Table 8)</td>
<td>Writes in essay that his writing is improving, but speaking is getting worse.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reports not enjoying speaking with native speakers.</td>
<td>Struggling with gerund/infinitive difference.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Disappointed that his English did not improve more.</td>
</tr>
<tr>
<td>Month 15</td>
<td>No colloquial speech in his transcripts</td>
<td></td>
<td>Practicing for master’s thesis presentation</td>
</tr>
</tbody>
</table>
APPENDIX E

SAMPLE TRANSCRIPTS FROM ANALYSIS

The following transcripts are selected from the analyzed data in this study. Therefore, they are coded in CHAT format. JUA represent Juan while ANN represents the researcher.

A Boy, a Dog and a Frog (1BDF1) February 25, 2013 (Month 1)

*JUA: but (.) when they almost (.) have it.

*JUA: the frog (.) fall [* m:a:0es] down.

*JUA: and (.) the kid &eh (.) trap [* m:a:0es] the dog with the net.

*ANN: okay.

*JUA: &uh (.) on [* s:r] this everything is funny.

*JUA: (.) and (laughter) <the> [/] the frog (.) has swum (.) to a rock (.).

*JUA: but it is angry (.) (be)cause he was happy on the (.) trunk or (.) whatever.

*JUA: and (.) in [* s:r] the other side the dog is just [/][/].

*JUA: now this is 0art funny (.) situation.

*JUA: so (.) I don't think (.) they are angry or happy.

*JUA: just they're happy.

*ANN: Okay.

*JUA: Now they are angry.

*JUA: (. ) and the kid is (. ) <sss> [/] screaming (.) at the <do> [/] frog (.) <just> [/][/] (.).

*JUA: again it’s like (.) &ah (.) this is the kind of situation that <you> [/] (.) you are (.) frustrated.
*JUA: and then in this scene the kid is with the dog in the net.

*JUA: it's a real funny moment.

*JUA: and the frog is far away trying to climbing [*mːa] in [*sːr] a rock.

*JUA: (.) and it is very disappointing <beca> [/] again.

*JUA: he was really calm.

*JUA: and they're trying to annoy him.

*JUA: &uhh the kid is angry.

*JUA: he's just screaming saying something like hey I'm gonna catch you.

*JUA: I can catch you.

*JUA: you know and the frog is just looking at them.

*JUA: the dog is &eh just still under the net.

*JUA: and they are really disappointed.

*JUA: so they decide just go away.
One Frog Too Many (2OFTM2) April 15th, 2013 (Month 2)

*JUA: so (.). the kid decided (.). <t> [/] to leave (.). 0art small frog <in the> [/] (.). <in> [/] (.). &eh &uh in [* s:r] the land.

*JUA: and they're trying <to> [/] (.). to go into the lake.

*JUA: and so they put the little animals in a small (.). &uh (.). I don't know in a (.). round <boat> [/] boat.

*ANN: raft.

*JUA: raft.

*ANN: that's what we'd call it.

*JUA: oh okay.

*ANN: for a small boat like that (.). especially made of wood.

*JUA: (.). Ah so (.). they're in the water.

*JUA: and finally this (.). small frog (.). jumps into <the> [/] (.). the raft.

*JUA: (.). and (.). yeah.

*JUA: (.). but &uh the kid (.). doesn't know that (.).

*JUA: and then again <small one is> [/] <the> [/] the big one is (.). saying something about to the small one. [+gram]

*JUA: (.). I don't know.

*JUA: he doesn't like it.

*ANN: okay.

*JUA: and again (.). he kick [* m:a:0es] (.). the small one off (.). of <the> [/] the raft.
*JUA: so (.) when he go [* m:a:0es] to the lake [/-] (.) when <they> [/] they are in the lake (.) and (.) they have a kind of boat (.) made by [* s:r] wood.

*JUA: it's <a> [/] a flat boat.

*JUA: (. <and> [/] (. and they are (. riding 0prep (. the boat.

*JUA: but they leave alone the old frog. [+gram]

*JUA: so (. he is (. disappointed about that.

*JUA: (. and the kid is saying like (. you have to stay there.

*JUA: (. okay just wait for us.

*JUA: but then <when the kid> [/] when the kid is [///] (. &uh (. &hm.

*JUA: (. so they depart.

*JUA: (. and (. the kid is watching [* s:r] (. &hay (. forward.


*ANN: yeah, ahead good.

*JUA: (. she’s [* s:r] watching [* s:r] ahead.

*JUA: (. then the old frog just (. jumped <into the> [/] into the boat.
Conversation Sample: IBOGOTA, Feb. 25th, 2013 (Month 1)

*JUA: (. ) so yeah ten years <befo> [/ -] ago 0subj was [* m:a] like (. ) &eh (. ) &eh (. ) many problems like (. ) &eh (. ) &eh &em violence or something.

*JUA: (. ) but now everything 0aux changing.

*JUA: and (. ) we have a lot of people there from everywhere in the world so.

*ANN: really?

*JUA: (. ) yeah.

*JUA: so <it’s> [/] it’s growing (. ) really really fast.

*JUA: and (. ) <it mean> [/ -] it means like we have (. ) everything.

*JUA: 0subj 0v more problems (be)cause we have more people.

*ANN: yeah.

*JUA: we have more problems.

*JUA: (. ) and (. ) the [* f:a:0:d] Bogota because Columbia has different place [* m:0s].

*ANN: mhmm.

*JUA: so they <are> [/] are different <to> [/] (. ) to the country or [* s:r] to the coast (. ).

[+gram]

*JUA: but &uh <I> [/] I miss Bogota because it's my city so.

*ANN: yeah.

*JUA: So.

*ANN: I understand.

*JUA: &um (. ) &um.

*ANN: how long how long have you been away?

*JUA: (. ) Sorry say [///].
*ANN: When was the last time you were in Bogota?

*JUA: Oh July.

**Conversation Sample: 15POLITICS, May 2, 2014 (Month 15)**

*JUA: ah no <I> [/] (.)(laughter) (.). I'm very crazy with [* s:r] that so.

*ANN: (.). yeah.

*JUA: (.). I don't know.

*JUA: (.). and the president signed 0obj <to> [/] (.). to be <the> [/] (.). the hero of this kind of peace (.).

*JUA: and he's trying to do that.

*JUA: I mean &ah come on.

*JUA: (.). he want [* m:a:0es] to be nominated for the (.). peace nobel prize. [+gram]

*ANN: (.). &ahh.

*JUA: (.). oh come on.

*JUA: (.). he wants that.

*JUA: and (.). I hate that.

*ANN: (.). yeah.

*JUA: <that's ridiculous> [/] I mean (.). that's really ridiculous.

*JUA: like Obama like (.). he won this <pre> [/] prize.

*ANN: mhmm.

*JUA: (.). yes that’s the word, prize?

*ANN: yeah, the nobel peace prize.

*JUA: the nobel peace prize.
*JUA: (. ) you know <he> [/ - ] (. ) he's maybe been (. ) the person that has (. ) averted most [* s:r] in war [* m:a:0s] than 0art others. [+gram]

Picture Description Set 1 (5SET1), July 10, 2013 (Month 5)

*JUA: was she dying?

*JUA: (. ) it looks like 0subj is dead.

*JUA: and <someone > [/] just (. ) &mm (. ) &uh (. ) someone [///] &mm she was (. ) upstairs.

*JUA: and then someone (. ) kill [* m:a:-ed] her.

*JUA: (. ) no (. ) 0subj put [* s:r] 0obj.

*ANN:(.) push.

*JUA: okay &ah <push> [/] yes 0subj push [* m:a:-ed] her.

*JUA: (. ) &um but (. ) she was dead (. ) before too.

*ANN:(.) How do you think she died?

*JUA: (. ) I think (. ) because [* s:r] (. ) her eyes (. ) her arms (. ) <she’s not> [/] she's not like trying <to> [/] (. ) to help herself. [+gram]

*ANN:(.) uhhuh.

*JUA: 0subj 0v like the body is stretched out (. ) <retched> [/ - ] no retching [* s:u].

*ANN:(.) mmhmm.

*JUA: Yes it's totally [///].

*ANN:(.) limp.

*JUA: (. ) limp?
*ANN: limp, yes.

*JUA: (. ) Yeah (laughter).

*JUA: so there's <no> [/] no life in her face.

*JUA: <it> [///] (. ) She isn't trying to do something because (. ) &eh otherwise <she> [/] (. ) she would have a different position (. ) in her body yeah.

*ANN: (. ) okay.

*JUA: 0subj 0v something like this.

*JUA: (. ) &um <So> [/] but it's not clear (. ) where she's falling.

*JUA: (. ) It looks like (. ) a little &um [///].

*JUA: (. ) I don't know.

*JUA: (. ) it could be she's not falling.

*JUA: <maybe> [/] (. ) maybe the perspective is different.

*JUA: and (. ) she has (. ) the [* s:u] feet (. ) on the floor.

Mr. Bean at the Swimming Pool (7SWP1) September 4, 2013 (Month 8)

*JUA: oh the guy &uh xxx <in> [/] in the swimming pool.

*JUA: (. ) and (. ) <his> [/] <he's> [/] (laughter) (. ) <he> [/] <he> [/] he walks really funny [* m:a].

*JUA: (. ) I don't know how to call that <the> [/] the underwear (. ) for the swimming pool .

*JUA: (. ) and now he's trying <to> [/] (. ) to use (. ) something for kids.

*JUA: (. ) I don't know what is the name of that.[+gram]
*JUA: (. ) and the person (. ) &uh who is <in> [/] <in> [/] in charge 0prep the swimming pool [/\].

*JUA: (. ) &um (. ) &um (whispers) and finally I have the [* f:a:i] victim.

*JUA: and the person (. ) who is in charge 0prep swimming pool [/] (. ) just saw him.

*JUA: (. ) and (. ) I know this video.

*JUA: and I know that (. ) he (. ) use [* m:a:0es] <the> [/] (. ) the thing.

*ANN: whistle.

*JUA: yeah.

*JUA: (. ) and then he expulse [* s:r] [* m:a:0es] 0obj (. ) from the swimming pool (. ) <or> [/] or put [* m:a:0es] [* s:r] 0obj in another part of the pool
APPENDIX F

EXCERPTS FROM INTERVIEWS DURING SESSIONS

These excerpts were taken from conversations between the participant and the researcher. They were selected to add context to quotes that are referred to throughout the paper. As before, JUA represents Juan, ANN represents the researcher.

**July 10th, 2013: Excerpt Immediately After Finishing 5Set1 (Month 5)**

JUA: I felt uncomfortable with this exercise.
ANN: Yeah, why?
JUA: Actually because I can't find the word, I don't know it was difficult and it was not difficult it's the that I (XXX)
ANN: Okay. Do you think it's mainly vocabulary?
JUA: it's que?
ANN: vocabulary? the words?
JUA: Yeah, I think so... hmmm.

**December 3, 2013: Conversation Regarding his Musical Compositions being performed at Temple (Month 10)**

JUA: I also crashed with two guys approached me and they were XXX and they were really honest that they like it the piece. You see umm because you know some people I mean like my adviser and other composers is pretty formal so (yeah) and and this music I know I know that contemporary music is not really like umm the fam(?) with people people like music (mhmm) sooo hmmm sometime you say OK it's consultation just a formal a polite way to say ok but they were really...i mean honestly like sometimes some guy talking about the it shocked me I don't know I feel something about well it's very strange

**February 24, 2014: Year Evaluation (Month 13)**

JUA: And I'm different this is the last semester (.) so
ANN: oh yeah, you can see it
JUA: yeah it's almost like (.) so every day I am thinking about that
ANN: mhmm, that makes a big difference too
JUA: it seems too long this month
ANN: mhm but it's not too long you keep telling yourself that
JUA: yeah but it's
ANN: It's not very much longer
JUA: and I didn't tell you but ah my wife told me that she is expecting something
ANN: You didn't tell me that oh that's so exciting!
JUA: because I knew on more thing in my life
ANN: one more reason you want to go home yeah
JUA: so it's difficult because she’s a little sick and she's having a problem and she's alone
  (. ) uh so now I feel like I need to go there
ANN: Yeah, I understand
JUA: Yeah
ANN: Does she have family there? (. ) Her family is she there her family?
JUA: yeah yeah yeah
ANN: okay good so she has someone there (. ) you don't need to worry (laughter)
JUA: yeah yeah I mean
ANN: but I understand you want to be there
JUA: yeah so yeah I'm (. ) yup (. ) so when I think about the Ph.D (. ) I think oh my god (. )
    I don't want it … I mean I really don't want to do it and I think what should I do that
    if I don't want?
ANN: mhm
JUA: I mean I know it's a kind of an important thing for academic focus (. ) but in fact I I
    I don't think so I mean (. ) it doesn't stop like (. ) masters Ph.D whatever (. ) I mean I
    think that that that the academic world create those ideas (. ) I mean they're not really
    important I mean
ANN: mhm yeah
JUA: you know what I mean
ANN: yeah I know what you mean
JUA: uhhh
ANN: It's important to those who think it's important and nobody else
JUA: yeah but and and it's important for just one reason (. ) you're gonna get more money
    when you work
    …
JUA: I want I want to be part of that
ANN: you don't want to miss that
JUA: yea I want to be part of that because my wife says sometimes yeah you can do your PhD it doesn't matter (.).

ANN: Yeah

JUA: I want to be part (.). to to educate the baby

ANN: mhmm

JUA: I want to be part because (.). I have my ideas how to do it

ANN: Okay

JUA: and so (laughter) so I want to try that

ANN: yeah

JUA: and I don't care about I don't care really really about money I mean I'll be happy because okay money it's important but I'm not gonna be a poor person if I don't get a PhD I mean yeah that's not a real problem

ANN: It's something you have to measure the good and the bad and decide what's more important to you but you want yeah

JUA: yeah and so

ANN: Well you're learning to know yourself you're gonna know what you want (laughter)

JUA: Probably yeah because I never I never did what I want to do I usually follow my maestrros

ANN: mhmm

JUA: they say you're good at this but and they okay you say I'm good in that okay and they are right absolutely yeah right but because but it's not because I was good in those stuff it's because I'm very disciplined

ANN: uhhuh

JUA: and I can do it I mean I can do whatever I want cuz I'm really strict in this way

ANN: okay

JUA: So but they say ah you're good at that ah you're a good composer (.). okay (.). let's do it

ANN: okay, yeah yeah yeah

JUA: it's like and the other thing that I feel really really good to have been here is because I uh (.). I really like that I have I am a composer

ANN: mhmm

JUA: I have

ANN: It is really amazing

...
ANN: In Columbia you were doing very academic speech and here it’s not necessarily academic speech so almost a whole new language for you
JUA: yea
ANN: Even in the pronunciation
JUA: yea
ANN: cuz I’m sure you have noticed
JUA: yea yea yea absolutely absolutely absolutely yea but it seems like they have the time to learn better
ANN: OK
JUA: yea I mean that I know I know the difference when I find some people that I don’t know I have a mental block
ANN: umhum well you have a whole life to work on it so
JUA: yea I know I mean no worries I mean I’m worried because I’m leaving here
ANN: yea
JUA: I would and I think why I should I should be better I think for two years I
ANN: I spent a lot of time saying that to myself about my Spanish because I lived in Mexico for five years my Spanish should be much much better than it is but it isn’t
JUA: Yea Yea

March 27, 2014 Excerpt from Writing – English Learning History, (Month 14)

When I started the master program at Temple University in Philadelphia I realized that my level was still very low to speak to English native speakers. Before that I had lacked to the opportunity to speak with real people from United States and I felt disappointed and frustrated because my English was and is not enough to communicate satisfactorily with others. Now, I feel that is really difficult to improve if we do not speak a lot with others, but I am not that person who want to speak to others or to have friends. So I keep improving my writing, reading and listening skills, but my speaking is everyday worst and worst.

April 25, 2014 Conversation (from Month 15)

JUA: Oh... I mean I miss everything in my country. I miss my wife I miss you know that I'm expecting this baby, so (yeah) um... yeah I don't feel comfortable because I feel like a I like a stranger everywhere. It's difficult every day to wake up and you are you are a stranger (mhm) I talked to my friends in Columbia and they say the same. It’s like Yeah you you can be fine I mean every day you know that
you aren't part of this culture. but for some people it's okay, so you're different but at least for me and others like I don't like to feel that kind of feeling everyday like you are different every day you feel the same (yeah) and you know that others are different but also as I told you I have this kind of ... cultural difference and in some ways I feel that they are irreconcilable (uhhuh, good word) yeah, so I don't feel comfortable. I feel okay, but not really comfortable.

ANN: Ok, I understand. (you know) Yeah, I've felt similarly but yeah I do understand. umm Here's one I wanted to ask you for a while, when you go back to Columbia do you think you'll come back to the United States ever again?

JUA: uhhh, it depends on ... umm where I'm gonna do my PhD, it depends (OK) I have the chance here, (mhm) so I have the... doors open here so there's my first option to come back to Temple University but I don’t know maybe one year I have such a instruct Switzerland (Switzerland) Switzerland (yeah) and my wife and I we can go there maybe so I don't know but my first option is Temple University.

... ANN: Um, what does your family think of you being here?

JUA: What does my family think about why I'm getting thinking what I'm doing here? (yeah) What do you mean by

ANN: Are they happy you're here are they.. do they think it's a good idea, obviously they miss you

JUA: I found that they are proud (good ok) yeah, absolutely I mean ... my mother and my father, they are proud that I'm here ... I never say I feel bad they don't know about my feelings (ok) just my wife (mhm) and maybe my cousin because we are really good friends they told him I don't like to here I want to go back I need to be there and but not like maybe you know more about that than they are (yeah) they I mean I try to be more

ANN: Well I ask you (laughs)

JUA: No, I'm really I'm really happy he oh I'm fine or whatever yeah mmy mother doesn't know about this kind of feeling that my father doesn't know they know my thought about American culture because we talk about that but not this kind of
feeling that happens so (. ) they they are proud they feel oh it's nice that's he's there I'm the first the first person in my family (mhmm) that I study outside of Columbia so yeah they they have these kind of feeling of proudness? (Oh good) and my wife is like (. ) you have to do this ... she she was one of the first ( . ) yeah the main person to say you have to do it even if you don't want to (mhmm) you have to do it you have time there are many people that want to do that and they don't have this opportunity and they say I don't care I don't care I know but you have to do this [pounding on something for emphasis] you have to you have to and I say ok yeah I know this is a good opportunity that was a good experience whatever .. you know (yeah) but if you don't I'm doing well I mean they feel that they I mean they think that I'm happy that I'm ok yeah fine ok I don't tell my problem (ok) I don't want to

ANN: I understand I did the same thing with my parents (laughs) yeah it makes sense (just) well good I'm glad they're proud of you that's that's nice

JUA: yeah yeah they feel for everything when they come my parents (yeah) that they ... they want to be here they feel it's something important and yeah it's important ...

ANN: Ok, when you are in Columbia will you need English

JUA: mmmmm probably I mean in percentage probably 10 percentage (ok so not a lot) no, it depends on ... could be important could be not I mean in my work in classes, I have to read English oh I mean we have to read everything in English for students like books magazines and whatever but in class now maybe and probably if I am in a different position like I was several years ago they had to do presentations in English because they're some people from Delaware University were there (mhmm) so and I had to speak in English at that time but it was a very special case, so ... probably I would have to do something with English I think that 10 percentages

ANN: Alright, so not but not your everyday life is not going to change much

JUA: no absolutely not

ANN: You teach at a university right? the books are in English?

JUA: the most of them, yeah
ANN: Umm How do you feel about English now as opposed to when you got here... You improved a great deal, but what else do you think? Where do you think you show the most improvement? Where do you think has changed for you in your English?

JUA: um you know I ... did a good job with writing (mhmm) I think I have a XXX If I focus on writing I I can do a very good job I know that ohhhhh ... absolutely my speaking has improved but not what I expected but I think

ANN: what did you expect?

JUA: to speak more fluently (OK) I wanted to speak really... (yeah yeah) really really fluent (.) but I have to think a lot .(Ok ok) I have to think about the type to find the words and when I don't think sometimes I I say something that it is understandable (uhhuh) something like that you .... you sometimes ask me what did you say? (yeah) because when I don't think I don't know it's like the words goes go out from my mouth but I don't control them ... and what's that ohhh I said something wrong

June 7, 2016: Personal Correspondence

I really don't know how to answer your questions...probably I started to use "gonna" and "wanna" because people speak in that way. About the second thing I'm not sure what to say... when I'm saying something and then I have to stop because... "I don't know" why to say. That's it!

I don't remember if I made a decision about my speech in that way... I just tried to use what I heard by that time; sometimes I did it consciously and others I didn't. What I think is that sometimes I just tried to speak freely, trying to think not much (probably I' writing in that way right now!).

I think probably your guess is right but I can't confirm it 100%. However, when I tried to speak informally, after a time I thought that it was ridiculous and then I changed it again.
ANN (Sent July 29, 2106)

First, a few detail questions:

1. Did you apply to more than one graduate school? I have Colombia in part of my notes and Princeton in another. Did you apply to both?

2. Why did you decide to participate in this study? I don't think I ever asked you that.

…

In our sessions you start telling me that you're struggling (more than before) to talk to English speakers. I realize this was always difficult for you, but you talk about it more after this point. By this time, you'd made a decision to go back to Colombia and be with your wife and new baby. My guess (and what I say in the article) is that your focus changed to writing and language that would benefit you in Colombia (formal) - in other words, away from informal speaking and maybe speaking at all. Again, does this sound right to you?

JUA (response on August 1, 2016)

Your questions:

1. I applied to Brandeis too.

2. I don't really know why I did it, but I saw the notice and I thought that would be nice to speak in English to someone else... I think it was like that.

I think you're right about what I did at the end of my studies in Temple, I focused on writing.
APPENDIX G: EMAIL CORRESPONDENCE GRANTING PERMISSION TO 
REPRINT FIGURE 1. LEVELT’S 1999 MODEL OF SPEAKING

To whom it may concern:

I am a doctoral student at Temple University who is in the process of writing my dissertation. I am writing to ask your permission to reprint figure 4.1 (on page 87) from the following chapter in a book that your company published:


The article will be used in my dissertation's literature review in the section on Levelt's speech model (Levelt, 1989; 1992; 1999). Thank you for your attention to this matter. Please let me know if you require any additional information.

Respectfully,

Elizabeth Hepford
Ph.D. Candidate in Applied Linguistics
Department of Education
Temple University
Philadelphia

Sent: August 15, 2016

Dear Ms. Hepford,
I herewith grant you permission to reprint figure 4.1 of my 1999 paper in Brown & Hagoort (Eds).
With kind regards,
Willem Levelt

Willem J.M. Levelt, Director Emeritus
Max Planck Institute for Psycholinguistics
P.O. Box 310
6500 AH Nijmegen, the Netherlands
(+31) 0622949316
www.mpi.nl/Members/PimLevelt
Dear Elizabeth,

Thank you for your enquiry. You have our permission to use the OUP Material you list in your email below in your dissertation for submission to Temple University.

If at some future date your dissertation is published, it will be necessary to re-clear this permission.

Please also note that if the material to be used is acknowledged to any other source, you will need to clear permission with the rights holder.

Best wishes,

Ben Kennedy
Permissions Manager, Academic Rights & Journals Permissions
Oxford University Press
Great Clarendon Street
Oxford OX2 6DP
Direct tel. +44 (0)1865 354728
Direct fax +44 (0)1865 353429
e mail: ben.kennedy@oup.com

Received: September 2, 2016