A LONGITUDINAL STUDY OF ACCURACY, COMPLEXITY AND VARIABILITY

Wouter M. Penris
s2067331

MA Thesis
Department of Applied Linguistics
Faculty of Liberal Arts
University of Groningen

Main Supervisor – Dr. Marjolijn H. Verspoor
Secondary Reader – Dr. Wander Lowie

Handed in June 24, 2013

words: 17325
(only body counted: p 5 till 47)
Abstract
The purpose of the present study was to apply Dynamic Systems Theory (DST) methodology, comparing and evaluating the use of various measures of complexity and accuracy in light of DST, looking for different stages of stability or variability in the extended longitudinal development of advanced written English (L2) in a Dutch (L1) subject. Forty-nine texts produced in two non-concurrent phases over a period of 13 years were analysed using 17 different measures of complexity and accuracy. Clear evidence was found for the existence of: stages of high variability followed by attractor states, changing relations between variables, and non-linear development, all in line with DST. Out of the different groups of measures, the ones outlining dynamic development most strongly in our advanced learner were: Customised Lexical Frequency Profile, Average Sentence Length and Average Noun Phase length, Sentence Complexity (counting compound and compound-complex sentences) and Lexical and Syntactic Errors. Lexical Density and Type-Token Ratio were deemed weak. Finally, a new measure, Lexico-Syntactic Finite Verb Ratio (LSFVR), showed moderate dynamic growth at advanced proficiency levels but correlated strongly and significantly with all other measures, implying potential usefulness for future studies.

Keywords: L2 writing; Developmental Variables; Dynamic Systems Theory; Complexity; Accuracy; CAF Measures
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%ACWL</td>
<td>Proportion of Tokens in a Text Belonging to the COCA Academic Word List</td>
</tr>
<tr>
<td>%FLI</td>
<td>Proportion of Frequent Lexical Items</td>
</tr>
<tr>
<td>%ULI</td>
<td>Proportion of Unique Lexical Items</td>
</tr>
<tr>
<td>AL</td>
<td>Applied Linguistics</td>
</tr>
<tr>
<td>ANPL</td>
<td>Average Noun Phrase Length</td>
</tr>
<tr>
<td>ASL</td>
<td>Average Sentence Length</td>
</tr>
<tr>
<td>AWL</td>
<td>Average Word Length of Lexical Words</td>
</tr>
<tr>
<td>CA</td>
<td>Complexity and Accuracy</td>
</tr>
<tr>
<td>CCX+CX</td>
<td>The Combined Compound Complex Sentences and Complex Sentences</td>
</tr>
<tr>
<td>CLFP</td>
<td>Customized Lexical Frequency Profile</td>
</tr>
<tr>
<td>D</td>
<td>Lexical Density</td>
</tr>
<tr>
<td>DST</td>
<td>Dynamic Systems Theory</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a Foreign Language</td>
</tr>
<tr>
<td>FVTR</td>
<td>Finite Verb Token Ratio</td>
</tr>
<tr>
<td>L1</td>
<td>First Language</td>
</tr>
<tr>
<td>L2</td>
<td>Second Language</td>
</tr>
<tr>
<td>LA</td>
<td>Lexical Accuracy</td>
</tr>
<tr>
<td>LC</td>
<td>Lexical Complexity</td>
</tr>
<tr>
<td>LFP</td>
<td>Lexical Frequency Profile</td>
</tr>
<tr>
<td>LSFVR</td>
<td>Lexico-Syntactic Finite Verb Ratio</td>
</tr>
<tr>
<td>NFC</td>
<td>Non-Finite Clause</td>
</tr>
<tr>
<td>NFC/FV</td>
<td>Non-Finite Clauses per Finite Verb</td>
</tr>
<tr>
<td>PM</td>
<td>Punctuation Mechanics</td>
</tr>
<tr>
<td>RUG</td>
<td>Rijksuniversiteit Groningen - University of Groningen</td>
</tr>
<tr>
<td>S+C+F</td>
<td>The Combined Simple Sentences, Compound Sentences and Fragments</td>
</tr>
<tr>
<td>SA</td>
<td>Syntactic Accuracy</td>
</tr>
<tr>
<td>SC</td>
<td>Syntactic Complexity</td>
</tr>
<tr>
<td>SLD</td>
<td>Second Language Development</td>
</tr>
<tr>
<td>TTR</td>
<td>Type-Token Ratio</td>
</tr>
</tbody>
</table>
Table of Contents

Abstract ........................................................................................................................................... 2
List of Abbreviations ....................................................................................................................... 3
1.0 Introduction .............................................................................................................................. 5
2.0 Background ............................................................................................................................. 6
  2.1 Dynamic Systems Theory in Second Language Development ........................................ 6
  2.2 Measuring Development .................................................................................................... 8
  2.3 Signs of Dynamic Development in Previous DST Studies .............................................. 12
3.0 Methodology ......................................................................................................................... 14
  3.1 Subject & Data Collection ................................................................................................. 14
  3.2 Research Design ............................................................................................................... 16
  3.3 Coding & Modifications ................................................................................................. 20
4.0 Results .................................................................................................................................... 23
  4.1 Lexical Complexity ........................................................................................................... 23
  4.2 Syntactic Complexity ....................................................................................................... 28
  4.3 Lexical & Syntactic Accuracy .......................................................................................... 37
  4.4 Interactions between Lexical, Syntactic Variables ......................................................... 40
5.0 Discussion ............................................................................................................................... 42
  5.1 Signs of Dynamic Development ...................................................................................... 42
  5.2 CA Variables that Best Outline the Dynamic Development of an Advanced Learner .. 44
  5.3 Comparing Results with an Extended Longitudinal Focus ............................................ 46
  5.4 Limitations ....................................................................................................................... 47
6.0 Conclusion ............................................................................................................................... 48
Appendix A ..................................................................................................................................... 50
Appendix B ..................................................................................................................................... 51
Appendix C ..................................................................................................................................... 52
Appendix D ..................................................................................................................................... 53
Appendix E ..................................................................................................................................... 54
References ...................................................................................................................................... 71
"Variability - The fact or quality of being variable in some respect; tendency towards, capacity for, variation or change," (Variability, 2000).

In Applied Linguistics (AL) studies of the past, variability was often considered noise (Verspoor, Lowie, & van Dijk, 2008), and statistical tests used in AL were designed to work with coherent data 'untainted' by variability to detect linear causality. However, as the abovementioned definition indicates, variability might also provide information relevant to change. Ellis (1994: 137) was one of the first to identify "free variation [as occurring] during an early stage of development and then [disappearing] as learners develop better organized second language (L2) systems," implying that variability can be used as a source of information instead of irrelevant noise.

Since then, research on Dynamic Systems Theory (DST), originally a mathematical theory, has been steadily applied to second language development (SLD) (e.g. de Bot, Lowie, & Verspoor, 2007; Larsen-Freeman, 1997), which has coalesced to form a new branch of AL that considers SLD to be non-linear, and individual variability to be a source of information instead of noise.

So far, most DST studies were based on variables measuring Complexity and Accuracy (CA) in multiple written texts produced over a period of three years, but as far as we know, there are no DST studies that have looked at development for longer periods of time, even though SLD is a life-long process (cf. de Bot & Larsen-Freeman, 2011). In this paper we present an analysis of extended longitudinal data collected over a period of 13 years. The purpose of the present study is to apply DST methodology, comparing and evaluating the use of various measures of complexity and accuracy in light of DST, looking for different stages of stability or variability in the extended longitudinal development of advanced written English (L2) in a Dutch (L1) subject. More specifically, the primary research questions of the present study are:

1. Which CA variables show clear signs of dynamic development in terms of: variability, attractor states, competitive growth, supportive growth, and precursors?
2. Which CA measures, out of groups of similar ones, provide the best insight into the dynamic development of an (advanced) learner?
3. How do the results presented in this paper relate to the duration of this study and subsequently to the current literature, since the duration of this study is longer than previous DST studies.
2.0 Background

2.1 Dynamic Systems Theory in Second Language Development

In the past fifteen years, research on Dynamic Systems Theory has led to a novel division of AL that regards SLD as non-linear. Originally, mathematical DST looked at systems in which multiple variables dynamically interact with each other. De Bot et al. (2007) offer the example of the double pendulum, a simple system in which only two variables interact, yet the result is that the movement within the system is complex and difficult to determine. Real complex systems, for example human society, are systems in which very large numbers of variables influence each other over time.

Quintessentially, DST looks at change in these systems over time, better expressed as $x_{(t+1)} = f(x(t))$, meaning that the state $(x(t))$ that a system is in at a specific time is the function $(f)$ for the state at the next time interval $(x_{(t+1)})$. Though the function presented here is linear, it illustrates that a system is dependent on its previous state. An example to demonstrate this dynamicity of complex systems is the state of society on any day, which is logically dependent on its state the previous day. In dynamic systems, all variables are dependent on each other; if one variable changes, all the other variables are affected as well. Thus, states at later times are difficult, if at all possible to predict, as it is unfeasible to track all variables unless a very simple system is studied.

Human cognition is a perfect example of a complex system, where initial conditions determine the starting state of variables which interact over time, resulting in new states, signifying the typical iterative nature of development in DST (van Geert, 1994). For instance, a person's experiences along with motivation, intelligence, etc., represent the initial condition of the system at a certain time $(t_1)$, and all these factors influence each other, leading to a later time $(t_2)$, in which the values of these factors have changed. These variables interact, resulting in a continuous dynamic interaction between in- and external forces. Van Geert (2008) consequently makes a distinction between internal and external resources, where the former has a bearing on the resources within the mind and body of a person, such as intelligence and motivation, and where the latter has a bearing on the resources outside the body, such as money and access to education.

Whereas the process of acquiring the first language is heavily dependent on external forces, SLD is dependent on even more forces. O’Grady (2008a; 2008b) mentions a number of internal influences: critical period, knowledge of a first language, intelligence, general language ability, motivation, and so on, leading to the obvious fact that all these internal factors engage with the environment during SLD. If indeed all these factors are different for all individuals, then the assumption that language acquisition progresses similarly and linearly in every person, such as is the case in older language acquisition models like the Information Processing model, must be inadequate (de Bot et al., 2007). Likewise, Sparks et al. (2009) report that differences between individuals, such as language proficiency, can lead to large differences in the trajectory of SLD.

De Bot and Larsen-Freeman (2011) outline the characteristics of the application of DST to SLD:

First of all, when target behaviour is practised, the result is not always positive. Sometimes there is success and sometimes failure before something is mastered, which results
in non-linear development, progressing in steps, representative of the iterative nature of complex systems in SLD.

Secondly, there is a continuous movement in the system due to the self-reorganisation of variables or subsystems interacting with each other and with outside factors, but all variables are dependent on their initial conditions.

Thirdly, these systems develop as a result of available internal and external resources over time. Logically, there is a limit to the available internal and external resources, meaning that these have to be divided among certain tasks (van Dijk, Verspoor, & Lowie, 2011). For example, a person who is learning Spanish invests money (external resource) and energy (internal resource) in a course, but (s)he will not be able to spend these on an unlimited number of other courses, for energy (and often money as well) is limited, meaning that growth is ipso facto limited too. Additionally, these resources are interdependent on each other again. For instance, the ability to sing well is a resource for musicality, but musicality will also have a direct effect on the ability to sing. This is also applicable to other cognitive systems, such as language.

Fourthly, de Bot and Larsen Freeman (2011), but also Van Dijk and Van Geert (2007), and Spoelman and Verspoor (2010), indicate that increased variability is a sign of an unstable phase of learning with trial and error, often leading to a period of overuse or overgeneralisation, which is again frequently followed by a more stable phase with less variability when an observed ability shows more stability. In DST these phases are called attractor states, when target behaviour is mastered, and repellor states, when target behaviour shows high variability. However, when non-target behaviour or an incorrect form becomes the stable phase of behaviour, this is called fossilisation.

Many processes studied in SLD can likewise be viewed in light of DST, as multiple studies have already shown (e.g. Larsen-Freeman, 1997; Verspoor et al., 2008). De Bot et al. (2005) offer the example of the multi-lingual mind as a dynamic and complex system, where the super system contains many more subsystems (e.g. L1 system, L2 system). As L2 learners rarely show mastery of one skill before they move on to the next, language learning does not develop linearly (Larsen-Freeman & Long, 1991). Instead, students revert, then grow, stall, then learn, etc.. As a result, DST can shed light on the trajectory of SLD of an individual by identifying periods of variability; stages with increased variability preceding higher-level attractor states are a sign of learning, and the subsequent attractor state is a sign of having reached a higher level (e.g. van Dijk, 2004; Verspoor et al., 2008). It is even said that the variability itself is the underlying agent of development in complex systems (e.g. de Bot et al., 2007; Larsen-Freeman, 2006). Additionally, common paths of development can still be identified in such individual learning trajectories (Siegler, 2006; Verspoor, Schmid, & Xiaoyan, 2012).

Since the development of a second language in a dynamic, complex system is attributed to the presence or absence of resources, growth can only occur when resources are available, and when they are diverted to the proper subsystem (van Geert, 1994). Since resources are limited, so is growth, so when a learner spends all his available time on the acquisition of vocabulary in German, (s)he cannot spend this resource on English. Van Geert (2008) adds that each person’s individual mix of talents and resources has a direct effect on
the maximum growth at a certain time, or carrying capacity. Finally, Van Geert (1991) notes that it is also possible that a variable nears its ceiling levels, when an increase in carrying capacity does not lead to an increase in a grower any more.

When multiple linguistic subsystems are measured, it is possible to identify three special types of relations between them, outlined by e.g. Spoelman & Verspoor (2010):

1. When intelligence grows due to language instruction, linguistic talent can also increase. This is called connected growth in DST.
2. An example of the second relation occurs when all resources are diverted to the acquisition of German, leading to the attrition of other languages; DST calls this a competitive relation, where the diversion of resources to the one, will lead to a stall in, or negative growth in the other.
3. DST describes the existence of growth precursors in SLD, where a minimum level of the one is required before the other can start to grow. For example, syntactic complexity can only start to develop when a minimum number of words is present in the language system.

Thelen & Smith (1994) and Van Geert (1994) argue that DST and Emergentism are undeniably compatible, because language is a natural emergent property that is the result of basic cognitive processes interacting under influence external stimuli to become more than the sum of its parts. Therefore only the natural human ability to acquire language is necessary (de Bot et al., 2007).

Verspoor and Behrens (2011) point out that DST is compatible with many current language acquisition theories. First of all, like in DST, Cognitive Linguistics Theory looks at language as an emergent complex system in which multiple subsystems interact over time, influencing each other, never being static (Langacker, 2008). Secondly, Emergentist theories are in line with DST, as these also assume that complex structures emerge and continually evolve through iterative behaviour patterns (e.g. Hopper, 1998). Thirdly, usage-based accounts of language learning are compatible with DST (e.g. MacWhinney, 2008); stating that frequency of input is essential, the acquisition of complicated grammar in second language can only occur through many iterations, where many linguistic subsystems play a part, detecting patterns in input. This is in line with the iterative nature and the interconnectivity of subsystems described in DST.

The effect of frequency on SLD is essential, as has been reported by numerous studies (e.g. Ellis, 2002). Hart & Risley (1995) found that children perform better in school in later life when the number of interactions, as opposed to the quality of interactions, was raised. Repeated contact with words or syntactic structures causes more advanced and more extensive linguistic neural networks to form (Bybee, 2008), meaning that words and grammar will become entrenched through repetition.

2.2 Measuring Development.

As an L2 learner progresses and his/her control of the language improves, (s)he is expected to create ever more complex utterances, meaning that simple utterances will also become less frequent (Tomasello, 2003; Van Geert, 2009). More recently, Verspoor et al. reconfirmed that "learners move from the simplest and most frequent constructions to more
complex and less frequent ones" (2012: 256). Moreover, the number of errors is expected to decrease (Verspoor & Behrens, 2011). As sub-systems develop, this should become visualised when various operationalisations measuring complexity and accuracy of written language are compared (Schmid, Verspoor, & MacWhinney, 2011). Instances of increased variability followed by low variability are indicative of periods of learning, leading to attractor states, as has been discussed above. Moreover, when comparing variables, it is also possible to determine which ones show supportive or competitive growth, or changing relations (Verspoor et al., 2012). Finally, as the subject in this study is writing at quite an advanced level, ceiling levels may become visualised when a variable remains in a high-level attractor state for an extended time. The primary aim of visualising development across these different variables is to gain insight into the process of advanced SLD. Since the present thesis employs 17 different variables measuring syntactic and lexical complexity as well as accuracy at various levels, the literary grounding of these variables will now be presented, and each time their developmental expectations will be presented alongside.

**Lexical Complexity 1 - Diversification - TTR & D**

Leki et al. (2008) have recently provided an overview of studies pointing out that lexical measures alter as L2 learners become more proficient. Therefore we will look at several variables that investigate Lexical Complexity (LC). Two measures of lexical diversification are Type-Token Ratio (TTR) and Lexical Density (D). By counting the number of words (tokens) divided by the number of different words (types) it is possible to obtain a ratio (TTR) that indicates how diverse word use is in a text. For instance, when a text of 200 tokens has 100 types, then the TTR is 0.5, while if more different types, say 180, are used, TTR will go up to 0.9. Hence, a higher TTR is indicative of more advanced writing, as more different words are used in a text. A common problem with TTR is that it is unreliable with varying sample sizes (Richards & Malvern, 1998). Moreover, TTR is known to become unreliable for larger samples due to the high number of function words present in the English language (Schmid et al., 2011). MacWhinney (2000) notes that D, an alternative version of TTR, proposed by McKee, Malvern & Richards (2000), is a better measure because it circumvents the abovementioned flaws in TTR by applying mathematical curve fitting; as mentioned above, TTR shows a decline or down curve in growth due to the high occurrence of function words, and this curve is mathematically compensated for, so, as the values of D increase, lexical diversity grows, implying more varied vocabulary use (Johansson, 2008). Miralpeix (2006) mentions that D is accurate across various languages and contexts, and that D indeed circumvents the known problems with TTR. However, van Hout and Vermeer (2007) mention that D is an insufficient tool for supplying information about an individual text. As such, TTR and D are two complementary measures for lexical diversification that provide general insight into how LC develops, but the meaning of these variables becomes more interesting when we compare them to related measures of lexical complexity, as dynamic relations among variables can then be traced (Verspoor et al., 2012).

**Lexical Complexity 2 - Sophistication - AWL, Academic Words & CLFP**

Next to lexical diversification, lexical sophistication can measure which percentage of a text is either simple, unusual or advanced (Read, 2000). Average Word Length (AWL) is a
general lexical sophistication measure that indicates how sophisticated word use is by looking at their average length. As a writer of English becomes more skilled, (s)he tends to use longer words (Grant & Ginther, 2000), which will lead to a higher AWL. Furthermore, Jarvis et al. (2003) point out that AWL is an accurate indicator of the complexity of essays. Wolfe-Quintero et al. (1998) write that in English more advanced words are longer and also lower in frequency. Schmid et al. (2011) add that the above-described high occurrence of function words in English also negatively influences AWL, as these function words are short and very frequent, leading to possible deflation of the measure. To resolve this problem, Schmid et al. (2011) propose using the AWL of lexical words. More recently, Verspoor et al. (2012) found that AWL could not be used to distinguish between 5 levels ranging from beginner to intermediate, as the differences between their groups were not significant.

Another more specific variable measuring lexical sophistication is the proportion of tokens in the text belonging to the Academic Word List. Originally an idea by Laufer and Nation (1995), Coxhead (2000) refined their idea and created an official list of academic words; when in a text the proportion of words from the Academic Word List is high, it is arguably more lexically sophisticated. More recently, the Corpus of Contemporary American English (COCA, Davies, 2008-) have created a new, more comprehensive academic word list than the one devised by Coxhead. While the latter looks at word families, COCA researchers found that some words occurred more frequently in their corpus than expected. The COCA website specifies that all texts in it have been evenly taken from 5 genres, namely: Spoken, Fiction, Magazine Fiction, Academic and Newspaper. If an average lemma occurs 50 times, then these occurrences should be evenly divided among the five genres, which is 10 occurrences per genre. However, if it occurred more than 20 times in the academic part of the corpus, it was given an academic label. Since the academic part of the COCA alone consists of texts with a grand total of 110 million words, the COCA academic word list is much larger, more precise and arguably more comprehensive than the list supplied by Coxhead (Davies, 2008-). Though Brezina (2012) has recently argued that the academic corpus provided by Google Scholar is even more extensive, and that the COCA academic list is not large enough, Davies (2013) defends his word list, outlining several flaws in Brezina's study through the incorrect querying of the COCA, which resulted in faulty data. Theoretically, as the proportion of COCA academic words in texts increases over time, this is a sign of increased lexical sophistication.

Furthermore, lexical sophistication can be established using Lexical Frequency Profile (LFP), developed by Laufer and Nation (1995; 1999). This concept is built around the idea that learners learn the most frequent words in a language first, in line with the studies on frequency discussed above, precisely because of their frequent nature, meaning that more complex words are acquired at a later time as these are infrequent. Thus, a beginning writer of L2 English is likely to use many of the 1000 most-frequent words in English, but as (s)he progresses and is ever more exposed to language, it is likely that (s)he will also use more infrequent words. Likewise, the proportion of words in various frequency bands can be calculated. Verspoor et al. (2012: 244) indicate that the LFP bands are “sensitive to genre,” which is a potential problem for texts produced by L2 learners, for these often lack academic register and are genre specific. Instead, they employed a Customized LFP (CLFP), construed by Schmid et al.(2011), where all of the texts produced by the studied subjects, who had been
writing on similar subjects, were used to create a custom corpus which was equally divided into five frequency bands. Consequently, lexical sophistication was deemed "the originality of the vocabulary in relation to the corpus at large" (Verspoor et al., 2012: 245). They found that CLFP did not accurately discriminate between their 5 proficiency groups, but overall they reported that the proportion of most frequent words decreased as a learner progressed, while the proportion of least frequent words increased.

**Syntactic Complexity 1 - Length Measures - Sentence, Clause and Noun Phrase Length**

Though many researchers in AL considered length measures indicators of syntactic complexity, Wolfe-Quintero et al. (1998) point out that these measures do not actually describe how the objects become more complex, and Ishikawa (1995) reports that length measures are insufficient indicators of development when the objects measured contain too many mistakes. Moreover, Wolfe-Quintero et al. (1998) indicate that ratios are generally better markers for proficiency growth. However, Norris and Ortega (2009: 562) offer that when length measures are combined with other more specific ones, e.g. clause type, the view resulting from the combined variables is a "complementary" one that is more "distinct". Leki et al. (2008) add that parts of sentences, or objects within them may also provide useful insight into SLD.

Average Sentence Length (ASL) can be calculated by taking the number of words in a text, divided by the number of sentences in it. Verspoor et al. (2008) indicate that ASL increases as a texts become more advanced, though this measure fails to reveal how, and it is difficult to use this measure on a holistic level due to the length measure insufficiency problem caused by errors discussed above (Ishikawa, 1995). Therefore, Finite Verb Token Ratio (FVTR) may be a better indicator of complexity, for it includes the number of finite verbs, which are known to become less frequent in more advanced texts (Verspoor et al., 2008). FVTR is calculated by dividing the number of words in a text by the number of finite verbs in it. Verspoor et al. (2008) also describe that noun phrases (NPs), which are part of clauses and sentences, tend to become longer and that non-finite constructions (these are defined below) tend to increase as a learner becomes more advanced, consequently, so will ASL and FVTR. Verspoor & Sauter (2000) describe the complex nature of NPs, which are sentence elements that have a noun at its head and can take on pre- and post-modifiers, which, when expanded, can become very complex, because they can harbour yet more NPs in them subordinate levels, e.g. [The man [the little boy] met yesterday at the party] is kind. In sentences, NPs can also occur as the complement in Preposition Phrases. e.g. of the man. Since NPs can become extremely long in Dutch, while in English this is less common, if average NP length (ANPL) is already very large from the onset, then it is possible that this is an example of L1 interference (Verspoor et al., 2008).

**Syntactic Complexity 2 - Sentence Complexity & Clause Type**

Very recently, Verspoor et al. (2012) have measured clause type and sentence complexity to see if varying constructions occurred more or less frequently at certain developmental levels. They manually coded each sentence in the corpus as belonging to one of the following complexity types: fragment, simple, compound, complex or compound complex. Moreover, they manually coded each text for Non-Finite Clauses (NFCs), and
finite/dependent clause type (adverbial, relative, nominal). For both complexities, they followed Verspoor and Sauter's (2000) sentence and clause type classifications, as described in Appendix A.

Subsequently, Verspoor et al. (2012) report that more and more complex structures are used as proficiency increases, while the number of simple structures decreases. Both the number of NFCs and the number of dependent clauses increases as learners become more advanced, the latter of which is in line with Wolfe-Quintero et al. (1998). Finally, they report that the number of dependent clauses accurately predicts proficiency level.

**Lexical & Syntactic Accuracy**

It is a well-known fact that L2 learners make errors, and that these slowly diminish as proficiency increases (e.g. Leki et al., 2008; Wolfe-Quintero et al., 1998). Verspoor et al. (2012) have recently shown that beginners up to intermediate learners show fewer grammar than lexical and spelling errors but that these particular errors did not allow for discrimination into proficiency levels. Moreover, they hypothesised that their findings could be the result of either the fact they only used beginner and intermediate students, or that there simply is little difference between these groups. Additionally, they report high levels of variation across the different types of errors and across the various proficiency levels, but they speculate that this variation might decrease at even higher levels. Finally, in another study, Verspoor et al. (2008) mention, in a similar case to the present thesis, that their advanced L2 subject, who was a student at the same university, made too few errors for error variables to become informative.

**2.3 Signs of Dynamic Development in Previous DST Studies**

Thus far, there are several studies which report on possible signs of dynamic development, as have been discussed above, but there are only a few reporting on the interaction between variables. First of all, Bassano and van Geert (2007) show evidence for the existence of precursors, where children first had to develop simple L1 word constructions before they could use more complex ones. Secondly, Robinson and Mervis (1998) discovered that L1 lexical growth is a precursor for more syntactically complex L1 constructions. These two studies were performed in time windows smaller than 3 years.

Thirdly, Verspoor et al. (2008) reports on the interactive nature of variables in the developing system of an L2 English learner who was studied longitudinally over a period of 3 years; she discovered evidence for competitive relationship between TTR and ASL, signifying that a focus on word diversity took a toll on syntactic complexity and vice versa. Moreover, they suspected that a supportive relation between ANPL and FVTR existed, possibly signifying that increased NP complexity also led to increased sentence complexity.

Fourthly, Spoelman and Verspoor (2010: 1), who studied a Finnish L2 learner longitudinally for 3 years, contradict Verspoor et al. (2008), reporting that ANPL and FVTR are competitive growers, as "NP complexity and sentence complexity alternate in development," though they also found that there was supportive growth between word and sentence, and word and NP complexity.

Fifthly, Caspi (2010) performed a 36-week DST study using differing L1, L2 English subjects, in which she studied the same operationalisations as in the present study, Lexical
Accuracy (LA), Lexical Complexity (LC), Syntactic Accuracy (SA) and Syntactic Complexity (SC), and she reports that LC is a precursor to LA, which is a precursor to SC before finally SA is achieved. Finally, in a five-point proficiency cross-sectional study, Verspoor et al. (2012) report that relationships among multiple variables change between proficiency levels, and that this change should be studied in all its possible forms and directions, from multiple viewpoints.
3.0 Methodology

3.1 Subject & Data Collection

The subject studied was an academic English as a Foreign Language (EFL) student who went through two separate phases of university education over a period of thirteen years. From that period, 49 different writing samples were analysed for dynamic development using 17 complementary CA measures. It is hypothesised that by longitudinally analysing variability patterns and the relations between variables, the non-linear development of the student's proficiency will become clear; something that would not be as visible in cross-sectional studies. Moreover, the varying measures of complexity and accuracy were compared in order to ascertain their worth and use in a longitudinal study of an advanced learner. The study hopes to contribute to the field as it looks at development over a period of 13 years, rather than three.

The texts were produced by an L1 Dutch speaker of L2 English, hereafter referred to as Walter, who had been studying English for quite an extensive time already. From the age of 10 Walter had received English as a Foreign Language (EFL) instruction during primary and secondary school, which led him to receive a final secondary school degree at HAVO level (Hoger Algemeen Voortgezet Onderwijs - Higher General Secondary Education), which is comparable to A Level and is the second-highest final level attainable in secondary education in the Netherlands. English is omnipresent in Dutch society; children are exposed to English through TV, radio, and other media on a daily basis, receiving very large amounts of English input. Therefore, arguably English as a Foreign Language does not exist in the Netherlands, but English as a Second does, meaning that the Walter already had an advanced level at the time the study started. From this moment onwards, he produced various texts over a period of 13 years, which can be divided into three major phases. During the first phase (1), which lasted from September 2000 till July 2004, Walter studied at the University of Windesheim in Zwolle, majoring in Teaching English, which led him to receive a degree in English education in 2004. In the second phase (2), which lasted from August 2004 till September 2009, he worked as an English teacher in two Dutch secondary schools, and he also studied music in Amsterdam, which was mainly taught in English. No texts were produced during this phase. During the third (3) and final phase, which lasted from September 2010 till January 2013, Walter continued his work as a teacher of English, and he also studied at the University of Groningen (RUG), majoring in English Language and Culture and Applied Linguistics, which led to an academic degree.

There is a considerable difference in level between the two universities; the English study at the University of Windesheim in Zwolle was done at HBO level, a type of Dutch higher education in applied science, directly preparing for teaching, while the final English studies at the RUG were done at the highest possible university level in Holland, leading to an academic degree. Hence it is likely that especially academic English proficiency will have developed more in the latter phase.

As Walter had saved all of his written work from both the Windesheim and RUG years, it was possible to analyse and compare these texts; however, proper selection criteria had to be devised first in order have a set of texts that were of similar genre:
• Only academic texts were included; all informal texts were excluded (no journals, no week reports, only official assignments). As a result, of the roughly 300 original texts created during phases 1 and 3, only 49 were deemed of a similar nature in that they were all holistically academic, as was assessed by two researchers.
• Moreover, the production date of all of these texts still had to be verifiable by means of study guides, emails, diaries, etc.; when this was not possible, a text was excluded.
• Finally, when Walter had produced two acceptable texts in the same week, only one was randomly included; the True Random Number Service was used to generate a number between 1 and 10 and if the number was even, the first text found was included, while if it was odd, the second text found was included (Haahr, 2012).
• All texts had to have been produced at home, based on homework assignments, and though there had been set deadlines, there had been no immediate time pressure, such as is the case during a written test. Furthermore, Walter had free access to reference materials on the internet, or materials related to various proficiency courses.

After the text selection was finalised, from each text, sections were taken which contained about 200 words (with a 10 percent margin), beginning and ending with complete sentences. While previous DST studies used other sample sizes, for instance 100-word samples (e.g. Spoelman & Verspoor, 2010) or samples ranging between 25 to 200 words (e.g. Verspoor et al., 2012), in this study, sample size was set at 200 words because these longer selections arguably lead to a better resolution during the visualisation process. Additionally, the 49 texts allowed a maximum sample size of 200 words, as the shortest text was 200 words in length. The start of a section was randomly chosen by asking Haahr's True Random Number Service (2012) to provide a number between 1 and e.g. 100, if the text was 300 words long, or 1 and 1800 if the text was 2000 words long (\(X - 200\)). Likewise, the resulting number represented a word in a sentence, and this sentence was the start of the random selection.

It is relevant to mention that the difference in university level described above ipso facto means that the writing assignments done during phase 1 are less academic and less formal in nature than the ones that were written during the RUG years; the latter were aimed at preparing a student for conducting science and writing scientific publications, while the former were aimed principally at the Walter's personal language development. As such, the text topics of the writing assignments during the RUG years text are more formal and scientific as well. The following two writing samples serve to illustrate that academic register has indeed developed between the 3rd and 48th sample.

Sample from the third text (010103W)

*The interview by X consisted of questions about what people knew and thought about them. Everyone had his or her share of knowledge about them. One of the questions of the interview was about the covers that have been made after the Beatles split up. Almost everyone named Joe Cocker in specific with `What Would You Do`. Sixty percent thought that the Beatles` version was better. People said it was more original. When we asked them if they had contributed to the current music, several answers*
were given. Everyone said yes but people had different reasons. One said they were revolutionary because they combined pop music with an orchestra...

Sample from the 48th text (120105R)

The results of experiment 1 (table 1) seem to strongly support the existence of homoiophobia; even though all expressions that were tested for acceptability were in reality acceptable, only the first two ranked out of the 22 seem to be completely accepted. Where Kellerman reasoned that these differences in acceptability were the result of homoiophobia, it is perhaps also possible to account the present results to the rather low proficiency level of the participants. If proficiency is low, then some sentences might be automatically rejected because a participant does not understand an expression...

As is visible above, all texts were given a temporal code (e.g. 030306W or 110102W) to keep track of the order they had been written in. While compiling results into graphs, we discovered that the codes caused the graphs to become cluttered, hence we opted for a simple rank order in the present thesis, where samples 1 till 31 were produced during phase 1, and samples 32 till 49 were produced in phase 3.

3.2 Research Design

The primary aims of the present study are (1) to outline dynamic development in the accuracy and complexity of the longitudinal L2 English development of Walter, (2) to compare and contrast various CA measures to ascertain which ones provide the best insight into the dynamic development of an (advanced) learner, and (3) to compare the results found in the two separate phases, focusing on the extended longitudinal character of the present study, as the duration of the present study is longer than previous DST studies.

While the first and third aim will be discussed in more detail later, the second aim can be achieved by identifying which variable(s) out of the groups discussed below show(s) the most signs of periods of increased variability followed by periods of attractor states, in conjunction with growth or decline. Measures that show little variability and/or growth do not visualise dynamic development and are therefore weak dynamic indicators. Seventeen variables measuring accuracy and complexity were operationalised in line with the DST perspective designed by Verspoor et al. (2011), discussed in the background section above. These variables Moreover, hypotheses on which variable outlines dynamic development most strongly are provided when it was possible to make realistic predictions based on current literature.

Lexical Complexity - Diversification - TTR & D

Lexical Complexity (LC) was broken up into lexical diversification and lexical sophistication. The former was measured using TTR and D, two general complementary variables. TTR was calculated using the freq command in CLAN (MacWhinney, 2000), taking the number of different words or types, divided by the number of words or tokens in a text. The problem of differing sample sizes was circumvented by taking samples of 200 words. D was calculated with the built-in vocd command in CLAN (MacWhinney, 2000). D is
calculated using the formula $TTR = \frac{[D/N][1+2N/D]^{1/2}-1]}{N}$, where N stands for the number of types. Since we are comparing multiple texts by a single author instead of groups, results from D are not subject to the problems mentioned by van Hout and Vermeer (2007) above. In all, both TTR and D provide information on the variety of words in each text, and they are expected to go up over the course of the 13 years in this study. Based on the literature it is likely that D will more accurately reflect proficiency development.

**Lexical Complexity - Sophistication - AWL, Academic Words & CLFP.**

To measure lexical sophistication four complementary variables were used. Even though Verspoor et al. (2012) found that average word length (AWL) did not differentiate between proficiency levels at the lower levels, AWL could be a valuable dynamic indicator of LC, as has been indicated by the other studies mentioned above. Since the present study only looks at a single subject, as opposed to the cross-sectional design applied in Verspoor et al. (2012), AWL will be employed here despite the ambiguous findings in the literature. Thus, Schmid et al.’s (2011) suggestion of calculating word length of lexical words will be followed, in order to combat deflation due to the high occurrence of function words in English. As such, average word length of lexical words (from now on AWL) was calculated by taking the number of letters in the content words in each text, divided by the number of content words. AWL gives general information on how long, and therefore how complex these words are.

The second variable to measure lexical sophistication was based on Laufer and Nation’s (1995) original idea to count the proportion of academic words in a text, but instead we used the COCA website (Davies, 2008-), which has the built-in option to analyse which tokens in a text are academic, as was explained above. The above-mentioned problems, outlined by Brezina (2012) and countered by Davies (2013), are likely to be of no consequence, since the COCA was used consistently for all texts. The resulting tokens were counted and divided by the total number of tokens in a text, resulting in the Proportion from the Academic Word List (%ACWL). This variable is especially suitable for providing insight into the development of academic register during phase 3: the higher the %ACWL, the more advanced the writing has become. Taken together, AWL provides general information on word length growth, while %ACWL provides specific information on where that growth is concentrated, and both are expected to go up over the course of the 13 years in this study.

Thirdly, Schmid et al.’s (2011) suggestion was followed, and a custom corpus of lexical items was created for Walter, which was equally divided into five bands. Subsequently, the proportion of tokens in a text that belonged to the 50 most prevalent words in the entire personal corpus, excluding function words, was measured and dubbed the Proportion of Frequent Lexical Items (%FLI): as described above, the lower the %FLI, the more original a text is in relation to the corpus (Verspoor et al., 2012). However, the other end of the frequency spectrum was also measured by calculating the proportion of tokens per text that occur only once in the entire corpus, called the Proportion of Unique Lexical Items (%ULI) (Schmid et al., 2011). When the %ULI increases, a text is arguably more original in relation to the corpus. Taken together, these measures can indicate how lexical sophistication dynamically develops over time. For exact specifications on how these measures were calculated, please refer to Schmid et al.(2011).
**Syntactic Complexity 1 - Length Measures - Sentence, Clause and Noun Phrase Length**

Syntactic Complexity (SC) was measured using eight variables, three of which measure the length of text objects and five of which describe clause complexity. To measure the complexity of sentences, clauses and NPs, the length of these units was calculated as follows: average sentence length (ASL) was calculated by taking the total number of tokens in a text, divided by the number of utterances in it. For clause length, Finite Verb Token Ratio (FVTR) was used, which was calculated by taking the number of tokens in a text, divided by the number of finite verbs in it. Finally, each NP was manually extracted from each text, though only the NPs at the highest level possible were taken, and intra-NP NPs were not. Here, the NP definitions of Verspoor & Sauter (2000), described earlier, were used. Subsequently, average NP length (ANPL) was calculated by taking the number of words in the NPs of a text, divided by the number of NPs in it, using Microsoft Excel. The resulting variables are expressed as follows: ASL (words/sentence), FVTR (words/finite verb), ANPL (words/NP).

Since Walter already had an advanced level from the start, the problem of errors in length measures, described by Ishikawa (1995) and Wolfe-Quintero et al. (1998) above, do not apply. In line with the literature discussed above, we expect all three variables to increase as the texts become more advanced. Taken together, these three variables can provide complementary insight into the development of sentence complexity on three levels.

Finally, there is a certain degree of overlap present in these length measures. For instance, the measure FVTR is very general, supplying information about finite clause length; however, NPs and NFCs are also part of these clauses, so FVTR also outlines the development of NP and NFC length to a degree; it is an averaging out of various sub-measures. As a result, these general measures (ASL/FVTR/LSFVR(see below)) are strong indicators for general development, but since the various components are different at different times, they do not show what changes when; they are weak(er) indicators of dynamic change.

**Syntactic Complexity 2 - Sentence Complexity & Clause Type**

In accordance with Leki et al. (2008), the length measures were complemented with other complexity measures in order to construct a more complementary picture of the dynamic development of SC. This was done by manually coding for Sentence Complexity and Clause type, using the definitions in Appendix A (Verspoor & Sauter, 2000; Verspoor et al., 2012). As simple and compound sentence types are both very common in non-advanced writing (Verspoor et al., 2012), the numbers of simple and compound sentences, as well as fragments were added (S+C+F) and juxtaposed to the added numbers of compound complex and complex sentences (CCX+CX), indicative of advanced sentence types.

Moreover, while studying the dependent clause type variables, we realised that the separate finite/dependent clause variables did not reveal anything. Therefore the Relative, Nominal and Adverbial clauses were added to form one Dependent Clause variable. Moreover, the number of Main Clauses was calculated by taking the number of finite verbs and subtracting the number of Dependent Clauses in a text. Together, the relation between the number of dependent and number of main clauses can be studied.

Finally, the number of Non Finite Clauses was converted into a ratio per Finite Verb
(NFC/FV), by taking the number of Non Finite Clauses divided by the number of finite verbs in a text. This measure was chosen because it indicates how many NFCs are present for each finite verb, similar to how many dependent clauses are present for each Main Clause. Additionally, this is in line with Wolfe-Quintero et al.’s (1998) findings that ratios more accurately describe development.

In accordance with the sources mentioned above, we expect S+C+F to decrease and CCX+CX and Dependent Clauses to increase as the subject becomes more advanced. As sentences become longer, fewer finite verbs will be used in a 200-word selection because of the increased use of NFCs and longer NPs. Therefore we expect that NFC/FV will increase.

**Lexico-Syntactic Complexity**

Finally, we also propose that a new Lexico-Syntactic Complexity variable might provide useful insight into the development of a writer. Where FVTR is an SC variable based on the number of tokens in a text, the number of letters in a text will give a more accurate indication of how lexically advanced a text is, as word length increases when texts become more advanced (Grant & Ginther, 2000; Jarvis et al., 2003). Since 200-word text selections were used throughout, we propose that a new variable that combines these two variables might provide useful insight into the dynamic development of these samples. As such, Lexico-Syntactic Finite Verb Ratio (LSFVR), calculated by taking the number of letters in a text, divided by the number of finite verbs, is an advanced measure of complexity which might provide useful general insight into how complex a text has become both on the lexical and syntactic plain. An overview of all complexity variables can be found directly below in Table 1.

**Table 1 - Variables Measuring Lexical (LC) and Syntactic Complexity (SC)**

<table>
<thead>
<tr>
<th>LC Variable Name</th>
<th>Calculated by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type Token Ratio (TTR)</td>
<td>Types divided by tokens</td>
</tr>
<tr>
<td>Lexical Density (D)</td>
<td>[ TTR = D/N \left(\frac{1 + 2N}{D}\right)^{1/2} - 1 ].</td>
</tr>
<tr>
<td>Average Word Length (AWL)</td>
<td>Lexical Items divided by the number of letters in these items (disregards function words)</td>
</tr>
<tr>
<td>Frequent Lexical Items (%FLI)</td>
<td>Proportion of tokens in a text belonging to the 50 most frequent words in subject's own corpus, which is formed on the basis of subject's own 49 texts</td>
</tr>
<tr>
<td>Unique Lexical Items (%ULI)</td>
<td>Proportion of tokens in a text that only occur once in subject's own corpus</td>
</tr>
<tr>
<td>Academic Word List (%ACWL)</td>
<td>Proportion of tokens in a text belonging to the academic word list, as supplied by the COCA website (Davies, 2008-).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SC Variable Name</th>
<th>Calculated by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Sentence Length (ASL)</td>
<td>Total number of tokens divided by number of utterances.</td>
</tr>
<tr>
<td>Finite Verb Token Ratio (FVTR)</td>
<td>Number of tokens divided by the number of finite verbs</td>
</tr>
<tr>
<td>Average NP Length (ANPL)</td>
<td>Number of words in the longest possible NPs divided by the number of NPs.</td>
</tr>
<tr>
<td>Sentence Complexity (S+C+F, CCX+CX)</td>
<td>Number of sentences per text that are: fragment / simple / compound / complex / compound complex, added into S+C+F and CCX+CX</td>
</tr>
<tr>
<td>Dependent Clauses</td>
<td>Number of combined relative, nominal, and adverbial clauses.</td>
</tr>
<tr>
<td>Main Clauses</td>
<td>Number of finite verbs minus the number of dependent clauses.</td>
</tr>
<tr>
<td>Non-Finite Clauses per Finite Verb (NFC/FV)</td>
<td>Number of Non-Finite Clauses divided by the number of finite verbs.</td>
</tr>
<tr>
<td>Lexico-Syntactic Finite Verb Ratio (LSFVR)</td>
<td>Number of letters per text, divided by the number of finite verbs per text.</td>
</tr>
</tbody>
</table>
Lexical & Syntactic Accuracy

Lexical Accuracy (LA) and Syntactic Accuracy (SA) were operationalised into three and six variables respectively, visible in Table 2. Each variable represents an error category, and Verspoor and Sauter (2000) were used as a basis; Appendix B provides a clear overview of these variables and their conditions, as well as a number of examples of errors encountered during coding.

At the outset, while studying the results for the nine different accuracy variables, we discovered that Walter had already attained such a high level of proficiency that he produced few errors from early on, leading to a problematic visualisation of accuracy, in line with Verspoor et al. (2008). To resolve this problem, the three variables for LA and the six variables for SA were grouped together into two variables respectively, called Lexical Errors and Syntactic Errors.

Both Lexical and Syntactic errors should diminish as proficiency increases (e.g. Leki et al., 2008; Wolfe-Quintero et al., 1998), and there should be relatively more Lexical than Syntactic Errors (Verspoor et al., 2012). It will also be interesting to see if the high variability reported by Verspoor et al. (2012) in beginners and intermediate/level students will also be present in our advanced subject.

Table 2 - Overview of the Errors Belonging to the Lexical and Syntactic Error Variables.

<table>
<thead>
<tr>
<th>Lexical Error Variable</th>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical Error</td>
<td>Incorrect word use due to literal L1 translations of words or expressions, preposition errors, pronoun errors, errors caused by the incorrect use of a word semantically related to target form, blends of English and Dutch.</td>
</tr>
<tr>
<td>Spelling Error</td>
<td>Incorrect spelling due to L1 interference, phonetic spelling, homophone spelling of target form, typos</td>
</tr>
<tr>
<td>Authentic Error</td>
<td>Incorrect use of chunks/expressions beyond the word-level</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Syntactic Error Variable</th>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb Error</td>
<td>Incorrect predicate form or predicate use</td>
</tr>
<tr>
<td>Grammatical Error</td>
<td>Incorrect use of apostrophe, congruence, word class, number, articles</td>
</tr>
<tr>
<td>Mechanical Error</td>
<td>Incorrect use of capitals or spaces</td>
</tr>
<tr>
<td>Punctuation Error</td>
<td>Incorrect use of comma, full-stop, colon, semi-colon leading to problems such as Comma Splice, Fused Sentence, Fragments, or a restrictive or non-restrictive modifier punctuated incorrectly</td>
</tr>
<tr>
<td>Punctuation Mechanics Error</td>
<td>No comma before a conjunction that separates two clauses</td>
</tr>
<tr>
<td>Word Order Error</td>
<td>Word order (L1 interference or not)</td>
</tr>
</tbody>
</table>

3.3 Coding & Modifications

The 49 selected texts were coded and analysed using Codes for the Human Analysis of Transcripts (CHAT) and Computerized Language Analysis (CLAN), which were both created by MacWhinney and Snow (1990) for the Child Language Data Exchange Systems (CHILDES) project, allowing for the coding of errors, grammatical features and any other type of variable. As Polio (1997) indicates that error analysis is subject to individual interpretation, precluding inter-rater reliability, the coding was executed by a single researcher. Moreover, the researcher and the subject studied were one and the same person. However, after the initial coding had been done by the researcher, it was checked by another
one. Subsequently, the coding was finalised using the feedback supplied by the latter researcher. Appendix C provides an example of a number of finalised coded utterances.

To warrant accurate word length and unique word counts, we replaced all proper names with name, all numbers with numb, all geographic names with place, which are all four or five-letter words. Since CLAN does not recognise numbers, these had to be spelt in full, causing problems with the word length counts, as was also the case with long geographical and proper names. Furthermore, especially during phase three, many utterances with quotes interlaced the original writings of Walter, as is expected in academic writing. Therefore, these utterances were either deleted, or the quote was replaced with the word quote if the sentence structure allowed it. Finally, all enumerations longer than three words were cut. In Appendix D a more comprehensive overview of all modifications is presented, and also the reasoning behind them is explained.

3.4 Analysis Techniques

The analytical tools that were used to analyse the variables were developed by Van Dijk & Van Geert (2002). First the raw data was plotted and visually inspected for possible developmental patterns. Where needed, polynomials of the second degree were added to visualise general growth or declining trends. Whenever raw variability needed to be visualised, the data of a variable was detrended to take away growth. This was traditionally done by taking away the slope and intercept of the data (Spoelman & Verspoor, 2010; Verspoor et al., 2008), but the method had to be adapted for the presented study. While previous studies dealt with concurrent texts in one phase, the present case consisted of 2 separate phases (1 & 3), segregated by an extended time in which no texts were produced, leading to skewed detrending. Hence the phases were separately detrended to maximise variability, to take away growth trends, and to prevent skewedness due to growth in phase 2. To identify periods with different degrees of variability, moving min-max diagrams were created, where the minimum and maximum values of 5 instances was taken, as the following example shows:

\[ \text{min (t1...t5), min (t2...t6), min (t3...t7), etc.} \]
\[ \text{max (t1...t5), max (t2...t6), max (t3...t7), etc.} \]

Consequently, it was impossible to calculate min-max values over the first and last two data points of phase 1 and 3, as is apparent, for example, in figure 2 on page 24.

Moreover, Pearson-correlations between variables were calculated with significance in order to determine if there were general connections between variables, to visualise to what extent these variables measure the same or different constructs. It is noteworthy that this traditional statistic method will only show general connections, while we are primarily interested in the change in relation between two variables over time; therefore, these relations (precursor, competitive growth, supportive growth) between variables were determined by using moving correlation windows, as the following example outlines:

\[ \text{correlation (t1...t5), correlation (t2...t6), correlation (t3...t7), etc.} \]
Again, no values could be calculated over the first and last two data points in phase 1 and 3. Each time, the moving correlations were calculated both on the raw data, and on the detrended data. The resulting moving correlations are the best tools for visualising variable interaction. To outline local changing trends, polynomials to the fifth degree were added, so stages of competitive, supportive or changing relations became visible.

Finally, resampling techniques were used on the data of all variables, along with Monte Carlo analyses, as described by Van Dijk et al. (2011), using Poptools, a statistical analysis tool developed as a plug-in for Microsoft Excel by Hood (2004). A Monte Carlo analysis calculates whether an observed peak in variability is significant or not by randomly reshuffling the data 5000 times and by checking whether a similar peak occurs during those reshufflings. The $\alpha$ was set at 0.05, so if a peak occurred less than 250 times in the simulation, the $p$-value was deemed significant. Likewise, peaks in variability were analysed for significance, and developmental jumps were identified.
4.0 Results

In the results section, each time, two variables are presented side by side, starting with the more general measures first, leading to the more specific ones. In order to visually outline that the texts originate from 2 separate phases (1 & 3; the Windesheim & RUG years), an open space has been left between the samples from either phase, more specifically between samples 31 and 32, making it easier to look at the two phases separately, but still making it possible to see coherence between the two. Each time when significant Pearson Correlations were found, the accompanying scatter plots can be found in Appendix E.

4.1 Lexical Complexity

The most general variables of LC were Type-Token Ratio (TTR) and Lexical Density (D), which are visualised in figure 1, along with their polynomials to the second degree.

![Figure 1 - The Development of Lexical Density (D) & Type-Token Ratio (TTR) over Time in the Writing Samples of Walter, Including Polynomial Trendlines (2nd).](image)

The polynomials indicate that little to no growth is visible in TTR, while a slight growth, followed by slight decline is visible in D. A Pearson correlation revealed that there was a very strong positive relationship between D and TTR, \( r = .857; \ p<0.01 \) (two-tailed). Figure 1 also shows that the levels in variability remain rather constant in both variables and that D shows a greater range of variability than TTR. The only visible peak occurs in D at sample 29, which has been marked orange, but a resampling and Monte Carlo analysis revealed that it was not significant \( (p = 0.21) \), so this peak could also be attributed to chance. Furthermore, the bandwidth of variability generally stays at the same level, hence a Min-Max diagram will not be shown here.
As the moving window of correlations in figure 2 above shows, the main difference in variability between the two variables predominantly occurred at the initial stages of phase 1, while later on the moving correlation showed very little fluctuation, and it remained positive.

Two more specific variables of LC were Average Word Length of Lexical Words (AWL) and the proportion of tokens in a text belonging to the Academic Word List (%ACWL), which are visualised in figure 3 along with their polynomials.

The polynomials show little to no growth in phase 1, but during phase 3 there is a medium increase in AWL, which is clearly visible in the polynomial. Moreover, the %ACWL, the most specific LC variable in this study, shows a great increase, generally doubling in phase 3 alone. Compared, AWL shows more steady growth, while %ACWL shows more peaks. A Pearson correlation revealed that there was a very strong positive relationship between %ACWL and AWL, \( r = .849; p<0.01 \) (two-tailed). During phase 3,
increased variability can be observed, especially in %ACWL, which is better visualised in the Min-Max diagram in figure 4 below.

Figure 4 - Min-Max Graph Representing the Development of the Proportion of Tokens in a Text Belonging to the COCA Academic Word List (%ACWL) in the Writing Samples of Walter.

During phase 1, academic word use stays within a confined bandwidth, but this bandwidth greatly increases between samples 36 and 44, after which we see a small period of decreased variability between 44 and 46, followed by yet again another period of increased variability. A resampling and Monte Carlo analysis was not significant ($p = 1.0$), so the actual peaks in the observed variability can also be attributed to chance.

Figure 5 - Moving Window of Correlation Between the Detrended Proportion of Words in a Text Belonging to the COCA Academic Word List (%ACWL) & Detrended Average Word Length of Lexical Words (AWL) Including Polynomial Trendline ($5^{th}$) in the Writing Samples of Walter.

As the moving window of correlation between detrended %ACWL and detrended AWL in figure 5 above shows, the variability found in the two variables is predominantly positively correlated except at the initial stages of phase 1. Moreover, there is considerable
fluctuation in the correlation, though this does not result in a negative correlation except at sample 3.

The final two variables of LC that were studied were the proportions of Frequent Lexical Items (%FLI) and Unique Lexical Items (%ULI) in Walter’s own corpus, which are visualised in figure 6 below, along with their polynomials to the second degree.

The polynomials indicate that initially the %ULI is high in phase 1 but steadily declines over the course of this study, while the inverse is visible with the %FLI, which steadily increases during phase 1. Moreover, during phase 3, the %ULI increases again, while the %FLI decreases. A Pearson correlation revealed that there was a strong negative relationship between %FLI and %ULI, \( r = -0.673; p<0.01 \) (two-tailed). The higher the %ULI, the lower the %FLI, and vice versa. During phase 3, increased variability can be observed in %ULI, which is better visualised in the Min-Max diagram in figure 7 below.
During phase 1, there are two stages of increased variability at writing samples 3-8 and 13-22, alternated by two relatively stable periods at writing samples 9-12 and 23-26. During phase 3 the bandwidth of variability is even higher (34-42), and eventually returns to a smaller bandwidth of variability (43+). A resampling and Monte Carlo analysis was significant ($p < 0.05$), so the largest peak in the observed variability, between 39 and 40, cannot be attributed to chance, indicating that the decrease in bandwidth of variability at 42 is indeed significant.

In figure 8 above, the moving window of correlation between the detrended %FLI and the detrended %ULI generally supports the negative Pearson correlation found earlier. The variability in both variables correlates negatively, but there was an instance when the correlation turned into a positive one at writing samples 21-24. Moreover, there is considerable fluctuation in the correlation between the two measures, but generally it remains negative. Finally, a Pearson correlation was run to see how all the LC variables interact with each other, resulting in Table 3.

### Table 3 - Pearson Correlations of all Possible LC Variable Interactions, along with Significance.

<table>
<thead>
<tr>
<th>LC Correlations</th>
<th>TTR</th>
<th>D</th>
<th>AWL</th>
<th>%ACWL</th>
<th>%ULI</th>
<th>%FLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTR Pearson Correlation Sig. (2-tailed)</td>
<td>1</td>
<td>.857</td>
<td>.251</td>
<td>.199</td>
<td>.182</td>
<td>-.232</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.000</td>
<td>.082</td>
<td>.170</td>
<td>.212</td>
<td>.108</td>
</tr>
<tr>
<td>D Pearson Correlation Sig. (2-tailed)</td>
<td>.857</td>
<td>1</td>
<td>.142</td>
<td>.141</td>
<td>-.141</td>
<td>.081</td>
</tr>
<tr>
<td></td>
<td>.000</td>
<td>.142</td>
<td>.331</td>
<td>.335</td>
<td>.332</td>
<td>.597</td>
</tr>
<tr>
<td>AWL Pearson Correlation Sig. (2-tailed)</td>
<td>.251</td>
<td>.142</td>
<td>1</td>
<td>.849</td>
<td>.149</td>
<td>-.410</td>
</tr>
<tr>
<td></td>
<td>.082</td>
<td>.331</td>
<td>.335</td>
<td>.000</td>
<td>.307</td>
<td>.003</td>
</tr>
<tr>
<td>%ACWL Pearson Correlation Sig. (2-tailed)</td>
<td>.199</td>
<td>.141</td>
<td>.849</td>
<td>1</td>
<td>-.040</td>
<td>-.326</td>
</tr>
<tr>
<td></td>
<td>.170</td>
<td>.335</td>
<td>.000</td>
<td>.784</td>
<td>.022</td>
<td>.000</td>
</tr>
<tr>
<td>%ULI Pearson Correlation Sig. (2-tailed)</td>
<td>.182</td>
<td>-.141</td>
<td>.149</td>
<td>-.040</td>
<td>1</td>
<td>.673</td>
</tr>
<tr>
<td></td>
<td>.212</td>
<td>.332</td>
<td>.307</td>
<td>.784</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>%FLI Pearson Correlation Sig. (2-tailed)</td>
<td>-.232</td>
<td>.081</td>
<td>-.410</td>
<td>-.326</td>
<td>.673</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>.108</td>
<td>.579</td>
<td>.003</td>
<td>.022</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

Correlations in cells marked blue are significant at the 0.01 level (2-tailed).

Correlations in cells marked grey are significant at the 0.05 level (2-tailed).
It is apparent that there are two more, hereto before unknown significant variable interactions between %FLI and AWL, and %FLI and %ACWL respectively, which can also be visualised in a moving window of correlation, as has been done in figure 9 below.

![Figure 9 - Moving Window of Correlation Between the Detrended Proportions of Frequent Lexical Items in Walter's own Corpus (%FLI) & Detrended Average Word Length of Lexical Words (AWL), as well as Detrended %FLI and the Detrended Proportion of Words in a Text Belonging to the COCA Academic Word List (%ACWL) Including Polynomial Trendlines (5\(^{th}\)) in the Writing Samples of Walter.]

Even when detrended, the correlations are erratic in nature, which corresponds with their weak Pearson correlation strength. The polynomial of detrended correlation of %FLI and detrended AWL shows a gradual change from positive, to negative, to positive again, while the polynomial of the correlation of detrended %FLI and detrended %ACWL shows quite acute changes from negative, to positive, negative to positive, to negative again.

### 4.2 Syntactic Complexity

The first two SC variables that were studied were Average Sentence Length (ASL) and Average NP Length (ANPL), which have both been plotted with their respective polynomial trendlines in Figure 10 below.
Both variables show relative stability during phase 1, but increased growth during phase 3. A Pearson correlation revealed that there was a very strong positive relationship between ASL and ANPL, $r = .769; p<0.01$ (two-tailed). During phase 3, alternating stages of increased and decreased variability can be observed in both variables which is better visualised in the Min-Max diagram in figures 11 and 12 below.
During phase 1, the bandwidth of ANPL starts out very small, then expands to double its size after writing sample 5, and this medium variability remains until sample 27, after which it decreases again (28-29). During phase 3 we see a small bandwidth at 34-35, expanding to higher (36-41) and even higher variability (42-45), which eventually decreases again at 46. A resampling and Monte Carlo analysis was not significant ($p = 0.26$), so the peaks in the observed variability can also be attributed to chance.

Figure 12- Min-Max Graph Representing the Development of Average Sentence Length (ASL) in the Writing Samples of Walter.

During phase 1, the bandwidth of ASL starts out very small, which leads to a phase of medium variability between 16-24, after which it decreases again. During phase 3 we see an initially small bandwidth at 34-37, expanding to higher (38-41) and even higher variability (42-47). A resampling and Monte Carlo analysis was not significant ($p = 0.36$), so the peaks in the observed variability can also be attributed to chance.

Figure 13 - Moving Window of Correlation Between Average Noun Phrase Length (ANPL) & Average Sentence Length (ASL) in the Writing Samples of Walter, Including Polynomial Trendline (5th).

In figure 13 above, a moving window of correlation between detrended ANPL and detrended ASL generally supports the positive Pearson correlation found above: the
polynomial indicates that the variability in both measures generally correlates positively as well, except at the end of phase 1, where the correlation turned negative at writing samples 24-28, which is also reflected in the polynomial. Furthermore, there is considerable fluctuation in the correlation between the two measures over time, but during phase 3 there is more stability.

The next two SC variables that were studied were Finite Verb Token Ratio (FVTR) and Non-Finite Clauses per Finite Verb (NFC/FV), which have both been plotted with their respective polynomial trendlines in Figure 14 below.

Figure 14 - The Development of Finite Verb Token Ratio (FVTR) & Non-Finite Clauses per Finite Verb (NFC/FV) over Time in the Writing Samples of Walter, Including Polynomial Trendlines (2nd).

FVTR shows little growth during either the phase 1 or 3, but in comparison with the former phase, the average level of FVTR is higher during the latter phase, so the polynomial shows growth. Moreover, the levels in variability remain rather constant in FVTR. A resampling and Monte Carlo analysis were not significant ($p = 0.37$), so the small peaks could also be attributed to chance. The polynomial for NFC/FV indicates that considerable growth takes place during phase 1, while that growth seems to decline by the end of phase 3. A Pearson correlation revealed that there was a strong positive relationship between FVTR and NFC/FV, $r = .502; p<0.01$ (two-tailed). NFC/FV shows alternating stages of increased and decreased variability, which is better visualised in the Min-Max diagram in figures 15 below.
During phase 1, the bandwidth of NFC/FV is medium (3-6), after which there is a stage of lower variability (7-13). Subsequently, the bandwidth of variability increase greatly at samples 14-29, before it decreases again at a higher average level at samples 28-29. During phase 3 an initially extremely wide bandwidth slowly decreases over time. A resampling and Monte Carlo analysis was not significant ($p = 0.96$), so the peaks in the observed variability can also be attributed to chance.

In figure 16 above, a moving window of correlation between detrended NFC/FV and detrended FVTR generally supports the positive Pearson correlation found above: there is mostly a positive correlation between the variability in both measures. However, the figure also shows that there were three instances where the correlation turned into a negative one, at writing samples 9-12, at 36-39 and at 46, which is also reflected in the polynomial. Moreover, there is considerable fluctuation in the correlation between the two measures over time.

The next two SC variables that were studied were the Combined Simple, Compound Sentences and Fragments (S+C+F) and the Combined Compound Complex and Complex Sentences (CCX+CX), which have both been plotted with their respective polynomial trendlines in Figure 17 below.
The polynomials indicate that the number of S+C+F slowly decreases, while the number of CCX+CX stays the same over time. However, since the total number of main clauses per text slowly decreases over time (ref. figure 21), this decline is also reflected in S+C+F and CCX+CX. As a result, there is a relative increase in the proportion of CCX+CX as opposed to the proportion of S+C+F, which has been plotted in Figure 18 below.

It is immediately apparent that the polynomial for %CCX+CX indeed shows an increase over time. A Pearson correlation on the original counts (not the proportions) of S+C+F and CCX+CX revealed that there was a moderate negative relationship between the variables, \( r = -0.326; p<0.05 \) (two-tailed). Variability in figure 18 remains rather constant with the exception of the final samples of phase 1, where the low peaks indicate that variability is lower, and the final samples of phase 3, where variability is higher, as is indicated by the high
peaks. A resampling and Monte Carlo analysis was not significant for S+C+F ($p = 0.99$), so the peaks in the observed variability can also be attributed to chance. However, it was significant for CCX+CX ($p<0.05$), so the highest peak, visible writing sample 3, which has been highlighted orange, is significant. The variability is better visualised in the Min-Max diagram in figure 19 below.

Initially, the bandwidth of variability in CCX+CX is very high in phase 1 (3-5), which is followed by a period of low variability (6-11). Afterwards, subsequent periods of high (12-17), low (18-25) and even lower variability (26-29) are visible. Similarly, during phase 3 variability is medium at first (34-38), and lower at the end (39+). Combined with the positive Monte Carlo, it means that the first movement from high (3-5) to low variability (6-11) is significant.

In figure 20 above, a moving window of correlation between detrended S+C+F and detrended CCX+CX generally supports the negative Pearson correlation found earlier: the variability in both measures correlates negatively, as the polynomial indicates. However, the
The next two SC variables that were studied were the number of dependent clauses and the number of main clauses per text, which have both been plotted with their respective polynomial trendlines in Figure 21.

The polynomials indicate that the number of main clauses decreases rapidly during phase 1 and more slowly during phase 3, while the number of dependent clauses slowly increases during phase 1 and shows a slight decline at the end of phase 3. However, since the total number of main clauses per text slowly decreases over time, there is a relative increase in the number of dependent clauses, as these are tied to their main clauses. Therefore, Figure 22 shows their relative proportions (%Main Clauses, %Dependent Clauses).

The figure also shows that there were two instances where the correlation turned into a positive one, at writing samples 8-9 and 28-29, though the polynomial indicates that the negative trend remained.
The proportion of dependent clauses shows an increase over time. A Pearson correlation on the original numbers of dependent and main clauses (not the proportions) revealed that there was no significant relationship between the variables, $r = -0.261; \ p = 0.070$ (two-tailed). Moreover, a resampling and Monte Carlo analysis was not significant for either variable (Dependent Clauses: $p = 0.52$; Main Clauses: $p = 0.95$), so the peaks in the observed variability can also be attributed to chance. Below, the development of the number of dependent clauses is shown in a Min-Max graph.

In phase 1, the number of dependent clauses shows variability alternating in two cycles from high (3-5) to low (6-7), to high again (8-24), to low again (25-29). In phase 3 there are two more cycles from high (34-35) to medium (36-40), to high (41-45) to medium (47-46). The last SC variable that was studied was Lexico-Syntactic Finite Verb Ratio (LSFVR), which has been plotted alongside FVTR, including their respective polynomial trendlines, in Figure 24 below.

Both variables seem to behave in an identical manner, even though word length is included in LSFVR. As described earlier, FVTR shows little growth during either phase 1 or
3, but in comparison with the former phase, the average level of FVTR is higher during the latter phase, so the polynomial shows growth, which is also visible in LSFVR. Since average word length was higher during phase 3, LSFVR is also higher, therefore the polynomial indicates more growth. Like with FVTR, LSFVR also shows little variability, and a resampling and Monte Carlo analysis revealed that the observed small peaks were not significant ($p = 0.48$), indicating that they could be the result of chance. A Pearson correlation revealed that there was a very strong positive relationship between FVTR and LSFVR, $r = .957; p<0.01$ (two-tailed). Finally, Pearson correlations were run to see how all the LC variables interact with each other, resulting in Table 4.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ASL</td>
<td>1</td>
<td>.769</td>
<td>.611</td>
<td>.213</td>
<td>-.796</td>
<td>-.185</td>
<td>1.60</td>
<td>-.7.90</td>
<td>695</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.141</td>
<td>.000</td>
<td>.203</td>
<td>.271</td>
<td>.000</td>
<td>.000</td>
<td>000</td>
</tr>
<tr>
<td>ANPL</td>
<td>.769</td>
<td>1</td>
<td>.515</td>
<td>.140</td>
<td>-.617</td>
<td>-.035</td>
<td>.194</td>
<td>-.6.81</td>
<td>584</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.336</td>
<td>.000</td>
<td>.810</td>
<td>.182</td>
<td>.000</td>
<td>.000</td>
<td>000</td>
</tr>
<tr>
<td>FVTR</td>
<td>.611</td>
<td>.515</td>
<td>1</td>
<td>.502</td>
<td>-.3.93</td>
<td>-.536</td>
<td>-.414</td>
<td>-.7.33</td>
<td>957</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.005</td>
<td>.000</td>
<td>.003</td>
<td>.000</td>
<td>.000</td>
<td>000</td>
</tr>
<tr>
<td>NFC/FV</td>
<td>213</td>
<td>.140</td>
<td>.502</td>
<td>1</td>
<td>.095</td>
<td>-.401</td>
<td>-.333</td>
<td>-.2.92</td>
<td>467</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.141</td>
<td>.336</td>
<td>.000</td>
<td>.515</td>
<td>.044</td>
<td>.019</td>
<td>.042</td>
<td>.001</td>
<td>011</td>
</tr>
<tr>
<td>S+C+F</td>
<td>-.796</td>
<td>-.617</td>
<td>-.393</td>
<td>-.095</td>
<td>1</td>
<td>-.326</td>
<td>-.524</td>
<td>.832</td>
<td>-.494</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>-.000</td>
<td>-.000</td>
<td>.005</td>
<td>.515</td>
<td>.022</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>000</td>
</tr>
<tr>
<td>CCX+CX</td>
<td>-.185</td>
<td>-.035</td>
<td>-.536</td>
<td>-.401</td>
<td>-.326</td>
<td>1</td>
<td>.751</td>
<td>.055</td>
<td>-.436</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.203</td>
<td>.810</td>
<td>.000</td>
<td>.004</td>
<td>.022</td>
<td>.000</td>
<td>.709</td>
<td>.002</td>
<td>002</td>
</tr>
<tr>
<td>Dependent Clauses</td>
<td>.160</td>
<td>.194</td>
<td>-.414</td>
<td>-.333</td>
<td>-.524</td>
<td>.751</td>
<td>1.000</td>
<td>-.2.61</td>
<td>-.297</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.271</td>
<td>.182</td>
<td>.003</td>
<td>.019</td>
<td>.000</td>
<td>.000</td>
<td>.070</td>
<td>.038</td>
<td>003</td>
</tr>
<tr>
<td>Main Clauses</td>
<td>-.7.90</td>
<td>-.6.81</td>
<td>-.733</td>
<td>-.2.92</td>
<td>.832</td>
<td>.055</td>
<td>-.261</td>
<td>1.000</td>
<td>-.7.63</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.042</td>
<td>.000</td>
<td>.709</td>
<td>.070</td>
<td>.000</td>
<td>000</td>
</tr>
<tr>
<td>LSFVR</td>
<td>.695</td>
<td>.584</td>
<td>.957</td>
<td>.467</td>
<td>-.4.94</td>
<td>-.436</td>
<td>-.297</td>
<td>-.7.63</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.001</td>
<td>.000</td>
<td>.002</td>
<td>.038</td>
<td>.000</td>
<td>.000</td>
<td>000</td>
</tr>
</tbody>
</table>

Correlations in cells marked blue are significant at the 0.01 level (2-tailed).

Correlations in cells marked grey are significant at the 0.05 level (2-tailed).

It is apparent that there are more, hereto before unknown significant variable interactions. Moreover, FVTR is highly significant ($p<0.01$) with all other SC variables, and LSFVR is also highly significant ($p<0.01$) with all SC variables, except with the number of dependent clauses, where it is significant ($p<0.05$).

### 4.3 Lexical & Syntactic Accuracy

The two language accuracy variables that were studied were the number of Lexical Errors and the number of Syntactic Errors, the developments of which can be traced in figure 25 below, to which polynomial trendlines have been added to outline general tendencies.
This graph visualises how, over the course of this study, both variables show a steady decline, eventually leading to practically no errors being present at the end. The polynomials show that there are twice as many Syntactic Errors present at the start of the study in comparison to Lexical Errors, but the decline in the number of errors is also twice as large, as can be seen in the polynomial.

A Pearson correlation revealed that there was a very strong positive relationship between Lexical Errors and Syntactic Errors, $r = .722; p<0.01$ (two-tailed). It is noteworthy to mention that an average of 44% of the Syntactic Errors made consisted of punctuation and punctuation mechanics errors. Without these two error categories, development of both variables would run nearly concurrently, as is visible in figure 26 below.
The Min-Max diagram in figure 27 above shows that periods of high and low variability succeed each other during phase 1, and only at the beginning of phase 3 do the Syntactic Errors show high variability, after which a final period of low variability is achieved. However, a resampling and Monte Carlo analysis was not significant ($p = 0.92$), so the actual peaks in the observed variability can also be attributed to chance.

The Min-Max diagram in figure 25 above again shows that periods of high and low variability succeed each other during phase 1, leading to ever fewer errors. At the beginning of phase 3, the Lexical Errors already show low variability, which becomes even lower at samples 46 and 47. However, a resampling and Monte Carlo analysis was not significant ($p = 1.0$), so the actual peaks in the observed variability can also be attributed to chance. Finally, a moving window of correlation between detrended Lexical Errors and detrended Syntactic Errors was made, which is visible in figure 29 below.
It is apparent that there is considerable fluctuation in the moving correlation of the variability between the two measures, though the polynomial indicates that an initial positive correlation generally turns into a negative one at the end of the 13-year period.

4.4 Interactions between Lexical, Syntactic Variables

Table 4 presented evidence that FVTR is highly significant \((p<0.01)\) with all other SC variables, and that LSFVR is also highly significant \((p<0.05)\) with all SC variables, except with the number of dependent clauses, where it is significant \((p<0.05)\).

However, when we ran correlations for each possible variable combination in this study, we found that LSFVR and FVTR were the only variables that significantly correlate with all other variables, regardless of their syntactic or lexical nature, as is visible below in Table 5. In it Pearson correlations and significances of LSVR and FVTR with the all other variables are supplied. LSFVR generally has stronger correlations with high significances \((p<0.01)\) with all but two lexical variables, for which it has significance \((p<0.05)\). FVTR has less strong correlations with high significances \((p<0.01)\) with all but three lexical variables, for which it has significance \((p<0.05)\). Moreover, FVTR has correlations with high significances \((p<0.01)\) with all syntactic variables, while LSFVR has correlations with high significances \((p<0.01)\) with all but one syntactic variable, for which it has significance \((p<0.05)\). Finally, group descriptives (lexical/syntactic) indicated that LSFVR has stronger average correlation strengths with increased standard deviations relative to FVTR, as is also visible in Table 5 below.
Table 5 - Pearson Correlations of LSFVR & FVTR with all other Variables, with Significance, & Lexical and Syntactic Group Means and Standard Deviations.

<table>
<thead>
<tr>
<th></th>
<th>LSFVR Pearson Corr.</th>
<th>LSFVR Sig. (2-tailed)</th>
<th>FVTR Pearson Corr.</th>
<th>FVTR Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic Complexity Variables</td>
<td>CCX+CX -0.436</td>
<td>0.022</td>
<td>-0.536</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>S+C+F -0.494</td>
<td>0.000</td>
<td>-0.393</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>LSFVR 1</td>
<td></td>
<td>0.957</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>ASL 0.695</td>
<td>0.000</td>
<td>0.611</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>FVTR 0.957</td>
<td>0.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ANPL 0.584</td>
<td>0.000</td>
<td>0.515</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>NFC/FV 0.467</td>
<td>0.001</td>
<td>0.502</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Main Clauses -0.763</td>
<td>0.000</td>
<td>-0.414</td>
<td>0.003</td>
</tr>
<tr>
<td>Syntactic Complexity Variables</td>
<td>Dependent Clause -0.297</td>
<td>0.038</td>
<td>-0.414</td>
<td>0.003</td>
</tr>
<tr>
<td>Lexical Complexity Variables</td>
<td>TTR 0.420</td>
<td>0.000</td>
<td>0.381</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>D 0.308</td>
<td>0.031</td>
<td>0.294</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>AWL 0.698</td>
<td>0.000</td>
<td>0.537</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>%FLI -0.563</td>
<td>0.000</td>
<td>-0.531</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>%ULI 0.323</td>
<td>0.023</td>
<td>0.293</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>%ACWL 0.659</td>
<td>0.000</td>
<td>0.497</td>
<td>0.000</td>
</tr>
<tr>
<td>Lexical Accuracy</td>
<td>Lexical Errors -0.423</td>
<td>0.002</td>
<td>-0.341</td>
<td>0.017</td>
</tr>
<tr>
<td>Syntactic Variables Mean and Stand. Dev</td>
<td>M: .579 SD: .199</td>
<td></td>
<td>M: .548 SD: .172</td>
<td></td>
</tr>
</tbody>
</table>

Correlations in cells marked blue are significant at the 0.01 level (2-tailed).

Correlations in cells marked grey are significant at the 0.05 level (2-tailed).
5.0 Discussion

5.1 Signs of Dynamic Development

Our first research question was whether the signs of dynamic development would be visible or not. First of all, we did not see any linear development, which is in line with Larsen-Freeman & Long (1991). Additionally, in this study we have indeed come across various examples of stages of high variability leading to attractors, which were best visualised in the moving Min-Max diagrams. This is in line with previous DST studies (e.g. Spoelman & Verspoor, 2010; Verspoor et al., 2008). The best example of variability and attractors was found in figure 7, where %ULI showed varying stages of low and high variability, as the bandwidth of variability increased and decreased several times. In phase 1 we saw that the %ULI showed high variability followed by lower attractors, indicative of a process where Walter tried to use fewer ULIs, while in phase 3 we saw the inverse, where high variability led to a higher-level attractor. Since the Monte Carlo simulation indicated that the highest peak, present in phase 3, was significant, this is a perfect example of a developmental jump in a DST learning process (Verspoor et al., 2011).

The second best example of variability and attractors was found in figure 19, where CCX+CX showed similar behaviour, but there now there was a continuous growing trend as the relative proportion of CCX+CX sentences had increased, in line with Verspoor et al. (2012). Effectively, this means that as time progressed Walter used more grammatically complex sentences in his writing. Since the Monte Carlo simulation indicated that the first and highest peak in phase 1 was significant, this is another perfect example of a developmental jump in a DST learning process.

We do want to note that out of all of the 17 variables studied, we found only two instances where the Monte Carlo simulation indicated that a peak was significant, even though nearly all Min-Max graphs in this study showed stages of high variability and attractors. Additionally, the visual interpretation of Min-Max diagrams led to problems in exactly defining the stages of high and low variability. Figure 11 is a perfect example, where ANPL shows signs of an attractor at writing samples 3 to 5, leading to higher variability in samples 6 to 27, leading to another attractor at samples 28 and 29. However, when we look at phase 3, we see that the concept of "high variability" is suddenly completely different, for the variability observed in these samples is three times as high as during phase 1. Variability in DST can take on any kind of shape and general trends have been defined (e.g. Verspoor et al., 2011), one of which is the developmental jump which can be discovered with Monte Carlo analyses. However, we suggest that it essential to keep looking for more refined criteria and statistical tools to pinpoint other types of variability stages; for example, it should be possible to more precisely identify and calculate (the significance of) the relative surface changes of Min-Max diagrams over time, something that will have to be defined in future studies. Alternatively, applying models of growth, such as described by Lowie et al. (2011), could also lead to better identification of growth patterns, but this was beyond the scope of the present thesis.

In this study we have also come across various examples of interactive variability, where we could define the relations of competitive and supportive growth, or changing relations, using general Pearson correlations along with moving correlations diagrams on
detrended data. This is in line with previous DST studies (e.g. van Geert, 1994; Verspoor et al., 2012). First of all, the best example of supportive growth was found between %ACWL and AWL; when the number of academic words, which are generally longer, increased, average word length of lexical words grew as well, in line with expectations. This was determined by the very strong Pearson correlation and the positive moving correlation in figure 2, meaning that there was a clear supportive growth relationship between %ACWL and AWL.

Secondly, the best example of competitive growth was found between S+F+C and CCX+CX. There was a decline in S+F+C and a relative growth in CCX+CX, and a Pearson analysis indicated that there was a moderate negative correlation between them. Figure 20 subsequently showed that the variability in both measures correlated even more negatively, though there was some fluctuation. This means that Walter first used many simple sentence structures, which slowly decreased and gave rise to more complex structures, which is in line with e.g. Tomasello (2003); as the variability also showed a negative correlation, Walter invested energy in either simple, or more complex structures, so there was a clear competitive relationship between S+F+C and CCX+CX.

Thirdly, the best example of changing relations was found between %FLI and AWL, as outlined by figure 9. Here the polynomial indicated that there was a positive correlation between the variables at the initial stages of phase 1, after which it turned into a negative correlation. Moreover, in phase 3, the relationship was initially negative, but turned into a positive relation again at the end. This means that text originality and word complexity sometimes took a toll on and sometimes supported each other, indicating a change in the allocation of resources, which is in line with van Geert (2008).

Additionally, when we looked at the moving correlation between %FLI and %ACWL in the same figure (9), we saw even more changing roles. Here, text originality and academic word use varyingly supported or competed with each other. When we subsequently created moving correlations between random variables, we found the same image over and over again. These changes seemed erratic, and though it was impossible to define why these changes occurred, they indicate that available resources were spread over different variables at different times, sometimes causing competitive, and sometimes supportive growth. Since these changing roles are identified visually, it yet has to be proven statistically that the observed changes are not the result of chance. More refined analyses through modelling need to be done, but this is beyond the scope of this thesis.

Fourthly, we were unable to observe a precursor relationship in this study. We propose that this might be due to the fact that Walter's level was already too high to allow for these to be visualised. We did find strong evidence for the existence of ceiling levels; both D and TTR showed very little growth and variability, but they stayed at nearly the same (high) levels over the 13 years. Since Walter's proficiency evolved from advanced to very advanced, arguably, these findings indicate that Walter’s word diversification was already at a ceiling level.
5.2 CA Variables that Best Outline the Dynamic Development of an Advanced Learner

Our second research question was to ascertain which CA measures, out of groups of similar ones, provided the best insight into the dynamic development of an (advanced) learner. The groups will now be discussed, one or more variables will be chosen, and the reasoning behind the choice will be explained.

Lexical Complexity - Sophistication - AWL, %ACWL, %FLI & %ULI.

Of the four variables measuring lexical sophistication, the most general measure, AWL, showed little growth and little variability (figure 3), meaning that it did not outline dynamic development strongly. However, these findings are in line with the growth described by Grant & Ginther (2000), Jarvis et al. (2003), and Wolfe-Quintero et al. (1998). The more specific measure %ACWL showed greatly increased variability in phase 3, followed by an attractor, along with increased growth (figure 4). Effectively, when Walter started his academic training in phase 3, his use of academic words grew, in line with Coxhead (2000). Subsequently, this was successfully dynamically outlined in figure 4. As a result, %ACWL is only a good indicator of dynamic development for the academic phase 3.

The other more specific measures showed the most interesting dynamic patterns of all variables measuring lexical sophistication through CLFP. Representing the originality or unoriginality of word use in a text as compared to the corpus, %ULI and %FLI both showed stages of increased variability followed by attractors, but the growth of both measures (figure 6) showed difficult to explain patterns. For example, in phase 1, Walter started using more FLI, while the inverse was visible in phase three. We had expected the use of FLI to go steadily (NB, not linearly!) down, for fewer FLI would be indicative of more original word use (Verspoor et al., 2012). Walter was asked about this and said that initially he had been told by his teachers that he used too many complex words, and that he lacked many normal, everyday words in his vocabulary, resulting in his attempts to include more everyday vocabulary. However, since the attractor in %ULI in phase 3 moved to roundabout the same levels as the attractors in phase 1, this could also be indicative of an optimal level: a corpus with more or longer text samples will result in relatively fewer ULI in the corpus because words are then likely to be used more than once. Since our corpus of 49 texts is quite large, this might be the case, or perhaps the subject was indeed developing his colloquial, more frequent register during phase 1. During phase 3, Walter received a lot of new input on various academic topics, but later he specialised in linguistics. Therefore it is logical that we see more initial variability during phase 3, leading to an attractor at optimum level. Since %ULI also showed a significant peak, we conclude that %ULI and %FLI are two strong complementary variables which outline the general direction of the development of originality in the lexicon, relative to optimum levels. However, a complete picture of the development Walter's lexical sophistication was only possible by using all four variables together, in line with Leki et al. (Leki et al., 2008).

Lexical Complexity - Diversification - TTR & D.

Out of the two variables measuring lexical diversification, D was the better variable to outline dynamic development, as it showed more growth and more peaks in development (fig. 1), though the difference was very small. Above we speculated that the lack of growth could
be the result of D and TTR being at a ceiling level, which would prevent them from showing growth patterns. Alternatively, the problems known in literature (Schmid et al., 2011) could have been the cause for insufficient visualisation of development. Based on present findings, we conclude that both TTR & D are weak measures for outlining the development of lexical diversification in an advanced L2 learner.

**Syntactic Complexity 1 - Length Measures - ASL, FVTR & ANPL**

Out of the three variables measuring SC through length, FVTR (fig. 14) showed growth over the thirteen years, but most of it must have occurred in phase two, and there were no obvious patterns of variability and attractors; in phase 1 and 3, Walter did not increase the number of words tied to each finite verb, but he did in phase 2, and overall, developmental peaks were not visible. Therefore, FVTR weakly outlined the dynamic development of this advanced learner, contrasting with e.g. Verspoor et al. (2008). As the plot for FVTR resembled the plot for TTR, potentially, FVTR might also be nearing a ceiling level, or similar mathematical curve problems as have been described with TTR may be at the root of this lack in dynamicty.

ASL and ANPL (fig. 10-14) showed much more dynamic development, especially in phase 3, where growth and variability increased in both variables. The data suggests that Walter did not really develop either his sentence or NP length during phase 1, while he really started experimenting with these during phase 3. This is to be expected in the more academic setting of phase 3, since more precise NPs are needed to describe complicated subject matter, automatically leading to increased ASL too. Hence, the findings of Verspoor et al (2008) seem partially replicated. Since ANPL showed the highest peaks and ASL showed the most growth, there is no better outliner of dynamic development; both are equally strong length complexity measures, which are specifically well suited for tracking dynamic changes in very advanced learners.

**Syntactic Complexity 2 - S+F+C & CCX+CX, NFC/FV, and Main & Dependent Clauses**

Out of the five variables measuring SC through clause typology, the most general variables measuring main and dependent clauses provided good dynamic developmental information about Walter: over time, in 200-word samples, he started using more dependent clauses relative to the number of main clauses, which slowly decreased, which is in line with e.g. Verspoor et al.(2012) and Wolfe-Quintero et al.(1998). This signifies that more dependent clauses were tied to ever fewer main clauses, which means sentence complexity increased over time. Moreover, the dependent clauses showed several periods of high and low variability, meaning that this pair of variables outlined the dynamic development of our advanced writer strongly.

The number of Non-Finite Clauses per Finite Verb also provided strong dynamic developmental information: as expected, Walter began to use more NFC/FV over time, and there were two clear movements from attractor to high variability in phase 1, and from high variability to attractor in phase 2, as was visible in figure 15. Here it was also striking that the peaks at samples 16 and 32 were not significant, while visually they seemed important. We conclude that NFCs are developed rather early in advanced writing, but that variability
remains high once it has started developing, though this will have to be confirmed in future studies.

The best variable pair measuring SC through clause typology was S+F+C & CCX+CX. As described above, Walter started using relatively more and more complex, while using fewer and fewer simple sentence types (fig. 18), in line with Verspoor et al. (2012). Also, there were various visible stages of variability in CCX+CX, with a significant peak. Therefore, this is the strongest indicator of SC through clause typology.

**Lexical & Syntactic Accuracy - Lexical Errors and Syntactic Errors**

Of the variables measuring accuracy, both outlined dynamic development strongly, for they showed negative growth and various stages of variability and attractors. As was expected, over time Walter made fewer and fewer errors, both in the lexical and syntactic area (in line with e.g. Leki et al., 2008; Wolfe-Quintero et al., 1998), and variability also decreased in the end, which is in line with predictions made by Verspoor et al. (2012). More notable is the fact that more syntactic errors were present initially, though punctuation errors were included. However, these also declined at a faster rate, contrasting with the same study. Eventually both measures decreased to practically zero in parallel fashion, contradicting Caspi (2010) who reported that LA preceded SA. As the error rates neared zero in phase 3, the obtained image in the moving correlation (fig 29) is potentially unreliable. Taken together, these measures illustrated well how Walter dynamically moved from advanced to very advanced proficiency.

**Lexico-Syntactic Complexity - LSFVR**

When we compared our new variable LSFVR with FVTR, we found that the former was a slightly stronger indicator of dynamic development because it showed more growth (in fig. 24), though both variables showed rather little variability, meaning that they were good indicators of general growth, but weak indicators of dynamic growth at advanced levels, as they fail to show what changes when. More interesting was the fact that LSFVR significantly correlated with all other CA variables in the present study and did so more potently than FVTR, implying that LSFVR is a very good general growth indicator. However, future studies will have to determine whether this claim holds up at different proficiency levels as well.

**5.3 Comparing Results with an Extended Longitudinal Focus**

Our third research question was to find out how the results presented in this paper relate to the duration of this study and to the current literature, since the duration of this study is longer than previous DST studies. In general, Walter's development in phase 1 was very different from his development in phase 3. In phase 1 the variables generally showed less variability and less growth, in comparison to phase 3. If the present study had only looked at either the former or latter phase, which was the setup of previous DST studies, then many developmental patterns would not have been discovered.

For instance, D and TTR remained rather steady during phase 1 and 3, but the general levels in both phases differed, leading to a growth in the polynomials in figure 1 that otherwise would not have been detected. Likewise, the unique growth patterns between %ULI
and %FLI described above would not have become visualised if only one phase had been described. Moreover, changes in relations such as were described between %FLI and AWL, where the moving correlation in figure 9 outlined a move from supportive to competitive to supportive growth, would not have become apparent in such detail, showing multiple changes.

The extended nature also allowed visualisation across developmental levels, from advanced, to very advanced academic proficiency. Thus, the extended longitudinal focus of the present study has outlined various intricate developmental patterns, and we propose that more of these studies should be performed in the future, in order to better and more completely determine typical developmental patterns of measures and learners across multiple proficiency levels.

### 5.4 Limitations

Though the present study was performed as carefully and meticulously as possible, there are also a number of limitations. First of all, the main researcher and the subject studied were one and the same person, which could result in a degree of bias. However, there was a second researcher who checked all the initial data, precluding large-scale bias and inconsistencies.

Finally, while trying to interpret the intricate, yet difficult-to-explain growth patterns in %ULI and %FLI described above, we wondered whether we were still looking at normal L2 development or at a very specific type of L2 specialisation; in phase 3 Walter acquired very specific academic writing skills intended for use in scientific publications, which was reflected in the %ULI. Even though all combined variables in the present study show that there has been development, the direction of that development, which was deemed to be very advanced L2 English, could perhaps also be pointed towards a sub-type of L2 English: is this a scholarly jargonised English that has in fact little to do with regular English? Therefore, for future studies, we propose that advanced written work from various sources, both academic and non-academic, is compared across multiple languages to ascertain whether this can indeed be called a sub-type of L2 English, or if it is indeed very advanced L2 English.
6.0 Conclusion

Concluding, the present thesis has provided valuable insight into the development of an L1 Dutch learner of L2 English from advanced to very advanced proficiency, using 17 different CA variables. Development seemed to adhere to the principles described by DST (Larsen-Freeman, 1997).

There were clear signs of stages of high variability followed by attractors at a higher/lower developmental level in most of the 17 CA variables described, signifying the typical trial and error phase before new skills can be used more consistently. Also, in two variables we found significant evidence for developmental jumps. It is proposed that the mathematical tools for defining stages of variability should be expanded. Even though developmental trajectories can take on any shape in DST (Verspoor et al., 2011), the Monte-Carlo analyses were only able to identify a very specific type of jump. It is also proposed that the surface change of Min-Max diagrams over time may also be used to define changing stages of variability, but these will have to be developed in future studies. Alternatively, modelling techniques (e.g. Lowie et al. (2011)) may be applied to visualise patterns of growth, but that was beyond the scope of the present thesis.

Based on the principles of limited resources (van Geert, 1994), multiple changing relations among variables were observed, indicating that different subsystems received resources at different times and that variables sometimes grew competitively or supportively. Some variables showed only one type of relation for an extended time (e.g. AWL & %ACWL: supportive growth). Others showed continuously changing relations (e.g. %FLI and %ACWL). These could be defined because of the longitudinal nature of the present study. However, since the present study used data collected over a period of 13 years, instead of the regular three years in previous DST studies, more and more intricate patterns in changing relations over time became visible, which would otherwise not have been possible. Hence, it is recommended that future longitudinal studies try to extend the duration, crossing developmental levels of learners.

Evidence for the existence of ceiling levels was found in two variables, TTR and D, because they showed little growth and remained at nearly the same (high) level, without showing variability stages, for the full duration of the study. However, this could also be the result of inherent flaws in these variables. More research will have to be performed to confirm either hypothesis.

The present thesis has also compared 17 different variables measuring accuracy and complexity to determine which one(s) outlined the dynamic development of our advanced learner most strongly. The results show that development is tracked best using complementary variables (Norris & Ortega, 2009). Two groups of lexical and two groups of syntactic complexity measures were studied. Out of four variables measuring lexical sophistication, CLFP measures %FLI and %ULI were the strongest complementary pair, showing interesting dynamic patterns which supposedly move around optimum levels, though future research will have to confirm this. Lexical Diversity was best measured using D, though it was still a weak dynamic growth indicator for advanced and very advanced levels, since it presumably was already at a ceiling level. Out of the syntactic complexity measures, ASL and ANPL were the strongest length measures, especially well-suited for very advanced learners. S+F+C and
CCX+CX were the strongest sentence complexity variables, outlining that complex constructions increase, while simpler ones decrease (Verspoor et al., 2012). Finally, the accuracy measures Lexical Errors and Syntactic Error were very strong dynamic growth indicators, though the relations between them were difficult to establish as the number of errors neared zero. Moreover, the conjecture that the variability in the error measures declines as a learner moves to a higher-than-intermediate level was confirmed (Verspoor et al., 2012).

It was also discovered that LSFVR was a weak indicator of dynamic growth at the present advanced level, but it was a strong general measure, as it correlated significantly with all other measures in this study. FVTR did so too, but LSFVR showed stronger correlations. Therefore it is proposed that LSFVR is especially well-suited for use in general L2 development studies, though future studies will have to determine the worth of this measure for other proficiency levels.
Appendix A

Sentence Complexity and Clause Types, as devised by Verspoor and Sauter (2000).

A sentence was called

1) **Fragment** when it is not a complete sentence, for example, when it has no finite verb.
2) **Simple** when it only has one finite clause, where clauses are henceforth defined as consisting of at least a subject and finite verb. e.g. *I love Susan.*
3) **Compound** when it has two or more main clauses, linked by coordinating conjunctions or semi-colons, and the order of the clauses is fixed, meaning that that clauses are untransposable without changing the meaning of the sentence. e.g. *I love Susan for I hate Stephen* (different in meaning when turned around: *I hate Stephen for I love Susan*).
4) **Complex** when it has one main clause and one or more dependent clauses.
5) **Compound Complex** when it has two or more main clauses and at least one or more dependent clauses. e.g. *What you are is cool, and the girl who is standing over there is nice* (dependent clauses underlined, main clauses separated by comma).

A clause was called

1) **Adverbial Clause** when it contained a finite verb and had an adverbial function in the sentence: e.g. *I love him because he is so handsome.*
2) **Relative Clause** when it contained a finite verb and provided extra information about a certain noun (post-modification). e.g. *A man who plays tennis is kind.*
3) **Nominal Clause** when it contained a finite verb and took on the role of one of the essential parts of the sentence (Subject, Object, Complement, Attribute), which can be discovered by replacing the dependent clause with a single word. e.g. *What is done is done* (replaceable by *It*).
4) **Non-Finite Clause** when a non-finite construction contained a non-finite verb, functioning as nominal, adverbial or modifier, and it consists of either to-infinitives, ing-forms or infinitives. e.g. *I wanted to go to school.*

Please refer to Verspoor & Sauter (2000) for a clear and complete overview.
Appendix B

Error Categories

For most of the error categories, we used De Vries (2009) as a basis for error categorisation, unless indicated otherwise. Provided error examples have been taken from the 49 samples.

LA was operationalised into three variables or types of errors, the first of which is the Lexical Error, which is described by De Vries (2009) as a type of error at the lexical level that has arisen through the incorrect use of a word, potentially due to L1 interference. Some examples are: literal L1 translations of words or expressions, preposition errors, pronoun errors, errors caused by the incorrect use of a word which is semantically related to the target form, and blends of English and Dutch. (e.g. makes them able, vocab, death tolls, etc.) Secondly, the Spelling Error is described by De Vries (2009) as an error that occurs because a word is spelt incorrectly, potentially due to: L1 interference, phonetic spelling, incorrect spelling of a homophone of the target form, typos, etc. (e.g. base (bass), you sessions, emissary, etc.) Thirdly, errors were considered Authentic Errors when an expression was grammatically (almost) correct but the English expression was not the idiomatic way to express it (e.g. gave the blame, taught about sex education, take this job on their hands, etc.) or a vital part of the expression was left off as in as you see rather than as you can see. If only one word was incorrect the error was considered a lexical error if it was not part of an idiomatic expression.

SA was operationalised into six variables or types of errors, the first of which is the Verb Error. For each predicate, the coder indicated whether the Form and Use were correct or not. Sometimes errors occurred separately in just Form (e.g. lets start), or Use (e.g. ...the covers that have been made after names split up), and sometimes both Form and Use were incorrect (e.g. I would reached). Secondly, errors were considered Grammatical Errors when there was a problem with grammar at either the word level, phrase level or sentence level, excluding errors in predicates. De Vries (2009) gives a number of examples: errors with apostrophes, congruence, word class, and number (e.g. ...a race implanted their offspring..., such a potential devastating device, etc.). Thirdly, errors were considered Mechanical Errors when there was a problem with capitals, or spaces (too few or many). Fourthly, errors were considered Punctuation Errors when there was a problem with any type of punctuation, e.g. full stops, commas, colons and semi-colons, which led to problems such as comma splice, fused sentences or fragments. Fifthly, errors were considered Punctuation Mechanics Errors when a comma was missing directly before a conjunction that separates two clauses, or when an initial adverbial phrase was too short to allow offsetting. Finally, errors were considered Word Order Errors when there was a problem with the word order in a sentence, either through L1 interference or not.
Appendix C

An Example of CHAT.

@Begin
@Languages: eng
@Font:
@Participants: XYZ 2067331 Student
@ID: eng|corpus|XYZ||male|group||||010101W|
*XYZ: there is also a second problem.
%syn: simple
%xten: [is] PRES+ PERF- PASS- COND- PROG- FORM+ USE+,
*XYZ: because of the road blocks, <fuelsupply> [: fuel supply] trucks have not been able to deliver
their fuel at the <gasstation> [: gas stations].
%syn: simple
%xten: [have not been] PRES+ PERF+ PASS- COND- PROG- Form+ Use+,
%err: [gasstations] err mechanics, [fuelsupply] err mechanics,
%xcla: [to deliver their fuel at the <gasstation> [: gas stations]] non finite,
*XYZ: because of that, people cannot buy any fuel anymore.
%syn: simple
%xten: [cannot buy] PRES+ PERF- PASS- COND+ PROG- Form+ Use+,
*XYZ: they are getting angry about this.
%syn: simple
%xten: [are getting] PRES+ PERF- PASS- COND- PROG+ Form+ Use+,
*XYZ: the truckdriver's motives are the high fuel prices.
%syn: simple
%xten: [are] PRES+ PERF- PASS- COND- PROG- Form+ Use+,
%err: [truckdriver's] err grammar,
*XYZ: people think they are paying too much tax to the state &P and now they want the taxes to be
reduced.
%syn: compound
%xten: [think] PRES+ PERF- PASS- COND- PROG- Form+ Use+, [are paying] PRES+ PERF- PASS- COND-
PROG+ Form+ Use+, [want] PRES+ PERF- PASS- COND- PROG- Form+ Use+,
%err: [&P] err punctuation,
%xcla: [to be reduced ] non finite, [they are paying too much tax to the state] nominal,
*XYZ: the state would not cooperate at first &P but when the entire country ran out of fuel because
of the blocks, the state sent their minister of traffic as an <emisary> [: emissary] , &P to negotiate
about the road blocks and about lowering the taxes.
%syn: compound complex
%xten: [would not cooperate] PRES- PERF- PASS- COND+ PROG- Form+ Use+, [ran] PRES- PERF- PASS-
COND- PROG- Form+ Use+, [sent] PRES- PERF- PASS- COND- PROG- Form+ Use+,
%err: [emisary] err spelling, [&P] err punctuation, [&P] err punctuation,
%xcla: [to negotiate about the road blocks and about lowering the taxes ] non finite, [when the
total country ran out of fuel because of the blocks ] adverbial,
*XYZ: until now, &P < there > [: : their] negotiations are not going very well.
%syn: simple
%xten: [are not going] PRES+ PERF- PASS- COND- PROG+ Form+ Use-,
%err: [until now ] err authentic, [&P] err punctuation, [there] err spelling,
@End
Appendix D

Overview of Modifications and Reasoning:

- All proper nouns were replaced with *name* for privacy reasons and to ensure that average word length counts or unique word counts would not be affected by long proper names.
- All plural proper nouns were replaced with *names* (e.g. The Beatles) for the same reasons.
- All geographic names were replaced by *place*, or when the geographic reference was an adjective pertaining to the people or language of a certain country, it was replaced by *dutch*, for the same reasons.
- All numbers (e.g. five, 5, 188, etc.) were replaced with the word *numb* because CLAN does not allow numbers in sentences, and the full written-out numbers affect word length counts and unique word counts for the same reasons as described above.
- Since CLAN considers all capitalised words to be proper nouns, all capitalised words other than person names had to be decapitalised (e.g. sentence-initial capitals, and examples provided below).
- Some proper nouns included actual references, (e.g. Berlin Conservatory). In this case, the geographical name was replaced with *dutch* and the referent was decapitalised (*conservatory*), as the referring function, instead of the naming function, was intended. Another example is *Universal Translator*, which is a proper noun for a device, but also an accurate compound noun and descriptive reference, so here only the capital letters were removed, instead of replacing the proper name by *name*. Another example is *Greenhouse Effect*.
- When geographical names consisted of two parts but referred to only one place (e.g. *Edinburgh, Schotland*), they were replaced with a single instance of the word *place*.
- Complex quantifiers such as *many hundreds of thousands (of years)* were replaced with a single word *numb*.
- There was one percentage in all samples, which was written as 40%, and replaced with *numb percent*.
- The common numeral *one* was not replaced when it is a personal pronoun.
- Ordinals were not replaced.
- Simplified representations of complex terms in essays (e.g. *L2, EFL*) were replaced with *name*.
- Whenever, two proper nouns were linked using the genitive (e.g. *Van Gogh's Sunflowers*), they were replaced with *name's name*. Here we did not use *name's names*, as this is a single referent (the painting), and not a plural one (e.g. *the Beatles*).
- All contractions were replaced with their full forms (e.g. *let's, we're*).
- All quotes were replaced by *quote* or the utterance was deleted if the structure did not allow replacing.
- During the building of the corpus (CLFP), instances of the word *name, place, number* and *quote* were excluded from it.

For an extensive overview of the coding requirements, please visit childes.psy.cmu.edu/.
Appendix E
Scatter plots of Significant Pearson Correlations
COMPLEXITY, ACCURACY AND VARIABILITY

Diagram 1: Relationship between Main Clauses and ASL with an R² Linear of 0.024.

Diagram 2: Relationship between S+C+F and ANPL with an R² Linear of 0.330.

Diagram 3: Relationship between Dependent Clauses and NFC/FV with an R² Linear of 0.111.
References


