

○ DYNAMIC DEVELOPMENT OF COMPLEXITY AND ACCURACY: A CASE STUDY IN SECOND LANGUAGE ACADEMIC WRITING

*Rosmawati Faculty of Education and Social Work, The University of Sydney
rosm4575@uni.sydney.edu.au*

This paper reports on the development of complexity and accuracy in English as a Second Language (ESL) academic writing. Although research into complexity and accuracy development in second language (L2) writing has been well established, few studies have assumed the multidimensionality of these two constructs (Norris & Ortega, 2009) or endeavoured to make long-term observations on the course of their development (Vyatkina, 2012). Given that recent research in the field of Second Language Acquisition has moved towards a more holistic perspective on language learning and development (Larsen-Freeman, 2011, 2012), there is a need to consider the potential dynamics of the longitudinal development of these constructs in L2 writing.

This study addresses this issue by exploring the dynamic unfolding of complexity and accuracy development in the academic writing of an advanced L2 learner during her postgraduate study in Australia. The results suggested that both complexity and accuracy displayed the characteristics of a dynamic system and their development was highly variable, non-linear, and idiosyncratic. Their interaction, too, was dynamic and changed over time. The findings in this study confirm and substantiate the Dynamic Systems Theory (DST) proposition of L2 developmental dynamics, including the development of L2 academic writing.

KEY WORDS: complexity, accuracy, L2 academic writing, dynamic development, interaction

INTRODUCTION

Along with the rapid growth of global education and its corresponding demand on academic writing skills, proficiency in academic English has become a vital factor in determining academic success (Lillis & Curry, 2006). This situation presents itself as a major challenge for international students (Leki, Cumming, & Silva, 2008; Paltridge & Starfield, 2007) who have to write academically in their non-native language, English, and are assessed using the same standard as the native students. Writing, being the most difficult skill to acquire (Richard & Renandya, 2002), has long attracted the attention of many researchers, who have endeavoured to understand its nature (Hyland, 2003; Silva, 1997; Silva & Reichelt, 2003); writing academically in a second language (L2) involves a great deal of additional issues (Ingvarsdóttir & Arnbjörnsdóttir, 2013; Lillis & Curry, 2010) and has, therefore, become a research domain itself (Cumming, 2006; Leki, 2011).

One major issue in second language academic writing concerns the development of L2 writing. Theories and hypotheses have been put forward to explain the phenomenon from the perspectives of sociolinguistics, psycholinguistics, systemic functional linguistics, and many others (Manchón, 2012). A relatively recent approach that attempts a more comprehensive description of L2 writing development is the Dynamic Systems Theory (DST) perspective (de Bot & Larsen-Freeman, 2011). Assuming that language development is dynamic, studies using a DST perspective take into account the multiple factors affecting (and therefore shaping) L2 writing and its development (Verspoor, de Bot, & Lowie, 2004; Verspoor, Schmid, & Xu, 2012).

Development in L2 writing mirrors the development of the learner's language system (Verspoor & Smiskova, 2012), as L2 writing involves the active use of all facets of language, from lexical to grammatical, from semantic to pragmatic, and discourse levels of language features. Therefore, L2 writing provides an avenue for understanding L2 development. However, despite the potential to offer a comprehensive account of L2 development, few studies have been attempted using in this line of inquiry.

This study addresses this gap by exploring the dynamic development of L2 academic writing in terms of the development of complexity and accuracy constructs. The study reported here is part of a larger study that addresses the issue of multidimensionality of the constructs in L2 writing, and explores the dynamic interactions among the three constructs of language performance (complexity, accuracy, and fluency; henceforth CAF) from the perspective of Dynamic Systems Theory. The purpose is to explore both the development and the interaction of those constructs in L2 academic writing. However, only two constructs are discussed in this paper: complexity and accuracy.

LITERATURE REVIEW

In the attempt to gain insight into second language acquisition (SLA), researchers have tried to explore and measure language development by gauging the progress in learners' performance in L2 production. Nonetheless, the constructs of L2 production are multidimensional (Housen & Kuiken, 2009; Norris & Ortega, 2009; Pallotti, 2009) and are therefore not reducible to one single aspect. Consequently, many models have been established to attempt a better description of learner performance. These include the communicative competence framework (Council of Europe, 2001), which has been commonly used in language testing and assessment practice. These models divide the proficiency/performance continuum into several categories (scales and/or subscales) and offer a very detailed description of what learners in each category "can do" with language.

Another way of understanding such a complex phenomenon is through an exploration into its multi-components. Many researchers contend that the pivotal aspects of L2 performance can

effectively be captured by the Complexity, Accuracy, and Fluency (CAF) constructs. Being the principal constructs of language performance, the CAF triad gauges, describes and benchmarks L2 production. An added advantage of the triad is that each construct encompasses various traits which can be assessed by a number of measures, discussed in more detail below.

THE CAF TRIAD: A HISTORICAL OVERVIEW

Along with the rapidly expanding interest in first language development in the 1970s (Brown, 1973; Hunt, 1965, 1970; Smith, 1973), researchers in the SLA field attempted to find reliable measures to evaluate L2 development (Larsen-Freeman, 1976; Larsen-Freeman & Storm, 1977; Wolfe-Quintero, Inagaki, & Kim, 1998). These measures were intended to be the Index of Development—an ‘independent yardstick by which we can expediently and reliably gauge proficiency in a second language’ (Larsen-Freeman, 1978, p. 439). In the early stages, length-based measures were borrowed from the field of first language (L1) acquisition to meet the immediate demand (Cooper, 1976; Flahive & Snow, 1980; Ho-Peng, 1983), and have since been widely adopted in SLA research. The most common ones are the mean length of particular structures (Norris & Ortega, 2009); for example, MLU (Mean Length of Utterances), MLS (Mean length of Sentences), and T-units (minimal terminable units). Although these length-based measures are useful to some extent, they are not free of problems. Beginner L2 learners, for example, rely heavily on rote-learned formulaic chunks (Myles, 2012); therefore, a long string of such structures in their L2 output results in a perceived development when length-based measures are used, hence giving a false impression of progress. A more systematic measurement of development is needed and this measurement should ideally give a more precise identification of the learners’ levels and allow for a more comprehensive description of the multiple aspects of L2 production. In order to arrive at the measurement system, we need to first understand what can be measured in L2 performance.

The notions of Complexity, Accuracy, and Fluency (CAF) were proposed as the principal constructs to capture the multidimensionality of the constructs in L2 performance (Skehan, 1998). In its early introduction in SLA research, the triad functioned mainly as dependent variables in studies on the effects of factors and treatments assumed to affect L2 acquisition (see Housen & Kuiken, 2009 for a list of such studies); however, in its more recent application, the triad increasingly gained significance and has become a very important set of constructs to describe L2 performance (Ellis & Barkhuizen, 2005; see also Wolfe-Quintero et al., 1998 and Ortega, 2003 for a review of studies employing these measures).

Despite their wide application, there is no single unifying definition of the constructs in the CAF triad. This is perhaps due to the fact that each construct in the triad encompasses traits which are multi-faceted in their nature, and are therefore difficult, if not impossible, to fully

define. However, a broad definition of each construct has been put forward; their detailed operationalisation in each study may then derive from these general definitions and include more facets of the traits they encompass. In its broadest sense, complexity is a measure of 'the extent to which the language produced in performing a task is elaborate and varied' (Ellis, 2003, p. 340), while accuracy measures 'the degree of conformity to certain norms' (Pallotti, 2009, p. 592) and reflects 'the conformity of second language knowledge to the target language norms' (Wolfe-Quintero et al., 1998, p. 4). Fluency, on the other hand, gauges 'how comfortable the second language [learner] is with producing [the target] language' (Wolfe-Quintero et al., 1998, p. 9). Therefore, complexity describes the learners' language knowledge while accuracy measures the appropriateness of language use and fluency the automaticity of language use. These three constructs, as a triad, gauge the learners' development as a whole.

The CAF constructs are measured in three different ways: frequency counts, ratios, and indices (Norris & Ortega, 2009; Wolfe-Quintero et al., 1998). In their survey of more than 50 studies in second language writing, Wolfe-Quintero et al. (1998) found that the T-unit measure was the best general descriptor for each of the three constructs,¹ as it corresponds well with proficiency level. Its application is highly recommended, especially in combination with other specific measures, in order to arrive at a more comprehensive description of learners' performance and development (Byrnes, Maxim, & Norris, 2010; Lu, 2011; Norris & Ortega, 2009; Vyatkina, 2012, 2013).

The CAF triad is both descriptive and normative (Ortega, 2003) and has proven to be a good yardstick for gauging L2 production (Ellis & Barkhuizen, 2005; Housen, Kuiken, & Vedder, 2012; Skehan, 2009). Being predominantly operationalised as a set of quantitative measures, the triad not only offers better perceptibility of development (evidenced by the changes in the numerical value of the indices) but also allows for better comparability across studies.

With the recent integration of Dynamic Systems Theory into the field of second language development research, researchers have become more and more aware of the multidimensionality of language learning and the nonlinearity of language development (de Bot, 2008; de Bot & Larsen-Freeman, 2011; de Bot, Lowie, & Verspoor, 2007; Larsen-Freeman, 1997, 2002; Mercer, Ryan, & Williams, 2012; van Geert, 2008). As the constructs of L2 performance, the development of CAF also reflects nonlinearity (Lowie & Verspoor, 2014). Such dynamic development has not been well accounted for within the contemporary practices in CAF studies, which still show 'a lack of attention to CAF as a dynamic and interrelated set of constantly changing subsystems' (Norris & Ortega, 2009, p. 555). More longitudinal observations are called for in order to explore the dynamic nature of CAF development (Larsen-Freeman, 2009). Dynamic Systems Theory, a meta-theory to explain developmental dynamics (de Bot, Lowie, Thorne, & Verspoor, 2013; Overton & Lerner, 2014; van Geert, 2011; Witherington, 2007), is then proposed as a suitable, and possibly the

most potentially fruitful, framework within which CAF studies can be attempted (Norris & Ortega, 2009). In this framework, the CAF constructs are treated as dynamic (sub)systems, whose growth is expected to be nonlinear and displays a high degree of variability as the expression of development (Larsen-Freeman, 2012; van Dijk, Verspoor, & Lowie, 2011; van Geert, 2014).

CAF STUDIES ADOPTING A DST PERSPECTIVE

In its essence, Dynamic Systems Theory (DST) describes how a dynamic system changes and develops over time and how complexity emerges out of such behaviour (de Bot, 2008; de Bot et al., 2007; de Bot, Verspoor, & Lowie, 2005; Larsen-Freeman, 2002; van Geert, 2008). A dynamic system is defined as a system with many interconnected subsystems which comprise a large number of highly intertwined components (Caspi & Lowie, 2010; de Bot & Larsen-Freeman, 2011). Both the development of these constructs and the interactions among them are dynamic and change over time; therefore, the outcome of such behaviours is nonlinear, seemingly chaotic, and displays a high degree of variability (van Dijk et al., 2011; Verspoor, Lowie, & van Dijk, 2008).

DST emphasises that variability is an integral part of development. Variability allows for room to accommodate changes in the system while the system is going through its self-restructuring process. Therefore, variability is not only an expression of the system's adaptability to changes but also a 'prerequisite to development' (van Dijk et al., 2011, p. 58). A dynamic system displays a high degree of variability during its development (Verspoor & Behrens, 2011). Examining the degree of changes in variability along developmental trajectories will, from the DST perspective, unveil the mechanisms underlying the changes and offer insights into how development unfolds (de Bot & Larsen-Freeman, 2011).

Given the dynamic nature of language development and all its highly interconnected components, the outcome of these interactions is also bound to be dynamic and non-linear. In order to capture the dynamics of language development, it is essential to attempt more observations of longitudinal nature (Larsen-Freeman & Cameron, 2008; Norris & Ortega, 2009; van Dijk et al., 2011) from a dynamic perspective that allows for exploration of the multi facets of development over different time scales.

Following this line of suggestions, some studies have endeavoured to explore the dynamic development of L2 production from a DST perspective. Given the scope of this paper, only the most recent ones are discussed here. Verspoor et al.'s (2008) longitudinal study (over a period of 3 years) on the academic writing of an advanced learner of English found a competitive relation between the sentence length measure and the type token ratio. This finding suggested that the learner might not be able to allocate her resources equally and this resulted in the non-concurrent development of the two levels of language use. Furthermore, the study found a great amount of variability in the learner's development suggesting that the

relationship between the two measures under scrutiny was non-linear and changed over time. These findings substantiate the DST prediction of a dynamic system's behaviour and confirm that language development is indeed dynamic.

Another three-year longitudinal study is that of Spoelman and Verspoor (2010), which looked into the interaction between accuracy measures and complexity measures in a Dutch student learning Finnish in an academic setting. They found that accuracy fluctuated considerably in the early stages, but soon settled down as the learner developed further. However, the authors also noted that the nature of the interactions between accuracy and complexity measures changed over time (i.e., a short-term negative correlation suggesting competition and a positive association suggesting simultaneous growth for most of the time), confirming the DST proposition that the system's behaviour is dynamic and non-linear.

Larsen-Freeman's (2006) study also supported the DST claim of high intra- and inter-individual variability accompanying developmental processes, as each participant in the study showed different developmental trajectories although the overall group average of CAF measures showed a general increasing trend over time.ⁱⁱ

Similar results were also found in a study by Vyatkina (2012) who examined the longitudinal and cross-sectional development of lexicogrammatical complexity in the writing of learners of German at college level. Besides confirming that length-based complexity measures correlated well with proficiency levels, Vyatkina also found a general upward trend in the development of lexicogrammatical complexity measures as established in previous CAF studies. Moreover, she also detected significant variability in individual and cross-sectional data: each participant's developmental pattern was highly dynamic and idiosyncratic. Each participant used different strategies to make his/her writing more complex. Such individual differences are expected within the DST framework, and the findings in this study provide further and more substantial evidence for the DST account of L2 development.

The most recent study of CAF development within the DST framework is that of Polat and Kim (2013). The authors explored the dynamics of complexity and accuracy in the L2 development of a Turkish immigrant in the USA in an untutored situation. It was found that their participant's syntactic complexity and lexical diversity developed well while accuracy seemed very constrained. They concluded that the participant's interlanguage was highly variable, but was perhaps nested within a relatively stable state.ⁱⁱⁱ Polat and Kim's study was the first study to attempt a longitudinal observation of the development of CAF constructs in a naturalistic learning context. More such studies were called for in order to unveil the nature of L2 development, and more specifically, CAF development and interactions.

Very much in line with Polat and Kim's suggestion, the current study explores the dynamic unfolding of complexity and accuracy development in L2 academic writing. This paper reports on the first part of a larger project which investigates the development of multidimensional

constructs in second language writing to look for developmental transitions and to identify interactional pattern(s) between the CAF measures over time by applying variability analyses within the DST framework (van Dijk & van Geert, 2007). Although the larger project involves four participants over an academic year, only one semester of data from one participant is reported in this paper. The data reported here were collected from an advanced English learner's academic writing during her postgraduate study in Australia. This paper traces the development of complexity and accuracy constructs in her academic writing over a period of one academic semester and explores the interaction between these two constructs along the course of development. The current study contributes to unveiling the dynamic relationship between these two measures and advancing our understanding of L2 academic writing development.

RESEARCH QUESTIONS

This study was designed to answer the following two research questions:

- 1) How do complexity and accuracy measures develop in L2 academic writing?
- 2) How do these two measures interact over time during the one academic semester observation period?

RESEARCH DESIGN

This study is a quantitative case-study investigation, based on a longitudinal observation of a single participant's written production over one academic semester. It examines the development of two constructs (complexity and accuracy) in the participant's academic writing. The data were collected and coded using a quantitative approach and submitted to statistical analyses to answer the research questions.

PARTICIPANT AND SETTING

The participant in this study was Machiko (pseudonym), a 32-year-old female Japanese student who came to Australia for the first time to study at a postgraduate level in an Australian university. Prior to her arrival in Australia, she had been studying English in her country (in a foreign language context) for about 15 years. Her exposure to English was mainly in a classroom context. Prior to coming to Australia, she had worked for a few years as a high-school English teacher in Japan. Her university enrolment in Australia was her first experience of studying in an English speaking country. As required for the program admission, she took a standardised English test (an IELTS test) prior to commencing her study and achieved a score of 6.5, which was equivalent to the B2 level on the CEFR scale; hence she was an advanced learner.

DATA

This study adopted a time-series approach to trace the development of complexity and accuracy constructs in the written output (academic essays) of the participant over one academic semester. The essays were the assignments for the courses she was enrolled in. These assignments were mainly argumentative essays (of approximately 2000-2500 words per essay) on topics in the TESOL area in which she majored. The topics for the assignments were announced at the beginning of the semester and the submission dates^{iv} were also specified beforehand. The assignments were completed at home and she had access to both the internet and word processing software. The researcher was not present during the writing process, which was also untimed. These assignments abided by the guidelines set up by the faculty and the course coordinator. Following their submission, copies of the assignments were sent to the researcher and served as the data for this study.

The rationale for choosing to collect the data from the very first semester was underpinned by the assumption within the DST framework that sudden proliferation of both input and use of an L2 may set the whole interlanguage system into chaos (de Bot & Lowie, 2010; Lowie, Verspoor, & de Bot, 2009; Verspoor, de Bot, & Xu, 2011; Verspoor, Lowie, & de Bot, 2009), resulting in high degree of variability which eventually leads to development (de Bot, Chan, Lowie, Plat, & Verspoor, 2012; de Bot et al., 2013; Larsen-Freeman, 2011; van Dijk et al., 2011; van Geert, 2009a). DST maintains that any dynamic system has the potential to fall into chaos and then restructure; hence, contrary to the common belief, 'even for an advanced learner, the system can be far from stable' (Verspoor et al., 2008, p. 229). Therefore, a high degree of variability was expected in this study and will offer key information unveiling the nature of development (de Bot & Larsen-Freeman, 2011; van Dijk et al., 2011).

SAMPLING AND CODING

The data were coded for complexity and accuracy constructs. However, as some measures for complexity, including those adopted in this study, are very sensitive to text length and correlate negatively with word count (Wolfe-Quintero et al., 1998), the text length in this study was controlled. A purposive sampling of approximately 200 words ($\pm 10\%$ of the original text) was conducted to filter out paragraphs with dense paraphrases and quotations as these may give a false impression of the learner's performance. As a result of this purposive sampling, a total of 10 pieces of sample text were obtained. These sample texts were then submitted to two coders and two independent raters for coding procedures.

Firstly, the data were coded for two complexity traits: syntactical diversity and overall grammatical complexity. Two types of measures were employed to measure these two traits, respectively: a frequency count of sentence types and a ratio measure of grammatical complexity. The sample texts were first coded for sentence types: simple (Si), compound (Co), complex (Cx) and compound-complex (CoCx) sentences (after Verspoor and Sauter's

[2000] definitions of each sentence type). The results were then tallied. The aim of this procedure was to see the distribution of sentence types in the participant's writing, hence a description of syntactical diversity.

Then, the sample texts were further coded for the second trait: overall grammatical complexity. Instead of employing some commonly used indices like MLT (Mean Length of T-units) or MLS (Mean Length of Sentences) and DepC (Dependent Clause for subordination amount) to measure general complexity, this study adopted another type of measure. As Verspoor et al. (2008) pointed out, such separate indices do not capture the entire degree of complexity because 'they do not bring to light additional complex construction such as longer NPs [Noun Phrases] or non-finite constructions' (p. 220). Especially in the case of advanced learners, coordination and subordination are not the only two strategies used in their writing. Advanced learners also use complex nominal structures and/or non-finite structures in their writing, especially in their academic writing (Biber, Gray, & Poonpon, 2011). A good measure of grammatical complexity should therefore account not only for subordination and coordination but also for complex nominal structures and non-finite structures. Separate indices, however, can only measure one complexification strategy at a time. Therefore, this study employed a W/FV (word per finite-verb) ratio^v to calculate the overall degree of sentence complexity. Besides measuring subordination and coordination, this ratio also takes into account complex noun structures, which are prevalent in academic writing but are not measurable by the other complexity measures, and combines them into a single measure to reflect the overall degree of complexity. This ratio measures the complexity of the sentences as a whole (Verspoor et al., 2008). Higher indices mean that the sentences are more complex.

Two accuracy traits were also explored: error types and overall accuracy ratio. The errors detected in the sample texts were coded, following Thewissen's (2013) study,^{vi} as Global Errors (GE), Local Errors (LE), and Mechanical Errors (ME). Global errors span entire clauses and affect the sentences as a whole; they include grammatical errors, and errors such as unclear/incomplete sentences. Local errors affect accuracy at the phrasal level and include errors at the word, lexical, and lexico-grammatical levels. Mechanical errors are errors at the stylistic level including form/spelling and punctuation. All occurrences of these errors in the participant's writing were then tallied and combined accordingly into the above-mentioned three categories of errors. The resulting distribution chart shows the error types and their frequency of occurrence.

Then, clauses that contained no error at all were coded as error free clauses (EFC) and each one of such clauses was counted as one EFC. The total number of EFCs in each sample was counted and then compared to the total number of clauses (C) to obtain the EFC/C ratio which described the portion of error-free structures in comparison to the entire text in the samples. The use of such a ratio of accurate production is highly recommended in the

literature as it tells ‘the true story in measuring the accuracy of the learners’ written production’ (Jiang, 2013, p. 21).

The following table summarises the coding scheme employed in this study.

Table 1. Complexity and Accuracy Measures

Measures	Frequency Counts	Ratios
COMPLEXITY	Sentence types:	
	• Simple (Si)	Word/Finite Verb (W/FV)
	• Compound (Co)	
	• Complex (Cx)	
	• Compound-complex (CoCx)	
ACCURACY	Error types:	
	• Global Errors (GE → grammar & style)	Error-Free Clause/ Clause (EFC/C)
	• Local Errors (LE → word, lexical, & lexico-grammar)	
	• Mechanical Errors (ME → form & punctuation)	

INTER-CODER RELIABILITY

As mentioned in the previous section, the sample texts were submitted to two coders and two independent raters. The author was coder 1. Coder 2 was an English teacher with more than 8 years of teaching experience. Coder 2 was given training sessions prior to doing the coding. Two editors who had more than 20 years of experience editing and proofreading L2 writing were also invited as the independent raters for this study. Their role was mainly to assist accuracy coding. After the data had been coded by the two coders, the coded data were passed to these two native raters. They then checked the coding for accuracy and gave suggestions to resolve any discrepancy between coders.

To ensure inter-coder reliability, a positive overlap ratio (POR) was calculated instead of the more commonly used Pearson’s correlation R value. The decision to employ POR was statistically motivated. Pearson’s R value measures whether the total number of confirmed cases in each sample picked up by coder 1 increases (or decreases) correspondingly with the total number picked up by coder 2. It results in a high value if the increases (or decreases) correspond although the two coders may actually pick up totally different cases. Hence, it is not a genuine reflection of agreement or reliability between coders. To measure reliability,

this study therefore adopted an overlap index, which reflects the percentage of overlapping positive cases confirmed by both coders. It is a relatively better measure of reliability as it shows the portion of the cases that are actually picked up and confirmed by both coders. For the formula to calculate the POR value, please refer to van Geert and van Dijk (2003).

In this study, the POR value reached 95% for the complexity measure and 78% for the accuracy measure. As accuracy measures the appropriateness of language use and is hence a relatively more ambiguous concept in its nature, a lower percentage had been expected. In fact, for such ambiguous phenomena, 'high agreement would be an indicator of low quality rating, for instance based on common errors and shared biases' (van Geert & van Dijk, 2003, p. 273). Justification of the quality of coding can be attempted through explicitly stated procedures; the results are then considered valid and trustworthy (van Geert & van Dijk, 2003). In this study, disagreement between coders was resolved by a discussion between the coders and the independent raters.

STATISTICAL PROCEDURES

The coded data were then submitted to two stages of analysis: the descriptive and the correlation analyses. Each stage corresponded to and answered the research questions respectively: the nature of development of the complexity and accuracy measures and the nature of interaction between them. At the descriptive stage, the data were plotted in developmental graphs to visualise the dynamics along the trajectories. Distributional bar charts were also presented to show the portion of each sentence type and each error type in the sample texts. This stage of analysis explored the nature of the development and hence answered the first research question.

A correlation analysis was then conducted to explore the association between the two constructs. The data were first smoothed to capture the general trends; a moving averages smoothing method^{vii} was chosen to accommodate the non-linear nature of data (van Geert & van Dijk, 2002; Verspoor & van Dijk, 2011). Both data sets were then normalised for comparison purposes. Finally, a moving correlation analysis^{viii} was conducted to explore the association between the two constructs. The results of this analysis answered the second research question regarding the relationship between the two constructs.

RESULTS

The development of complexity and accuracy measures in Machiko's writing in the observation period showed a great deal of variability. The data collected from her academic writing were analysed and the results are presented below.

RESULT 1: THE NATURE OF THE DEVELOPMENT OF COMPLEXITY AND ACCURACY

Figure 1 presents the developmental graph of the overall grammatical complexity level in Machiko’s academic writing throughout the semester. As shown by the fluctuation in the graph, the development of complexity was non-linear and very dynamic, characterised by a high degree of variability along the trajectory. The changes over time showed that Machiko’s writing became relatively more complex at measurement point no. 3, but then ebbed and flowed until a point (no. 9) at which the measure dropped to a level that almost equalled the start of the semester. On the surface, it seemed that Machiko had regressed; however, from a DST perspective, such a seemingly regressing trend is also an expression of development (de Bot, 2008; de Bot, Lowie, & Verspoor, 2005; van Dijk et al., 2011).

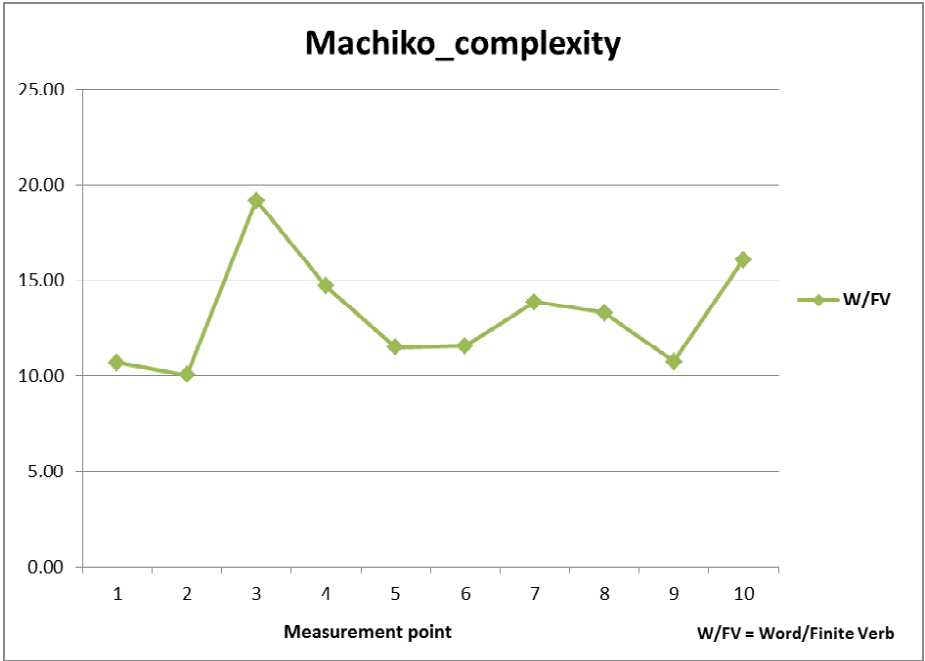


Figure 1. Complexity development over one semester

Figure 2 gives a better illustration of how complex the sentences in the texts were, as it shows the distribution of sentence types in each text. As can be seen in Figure 2, Machiko produced simple, complex, and compound-complex sentences^{ix} in these 10 sample texts. Nonetheless, their distribution was not, at all, balanced. She produced complex sentences most of the time.

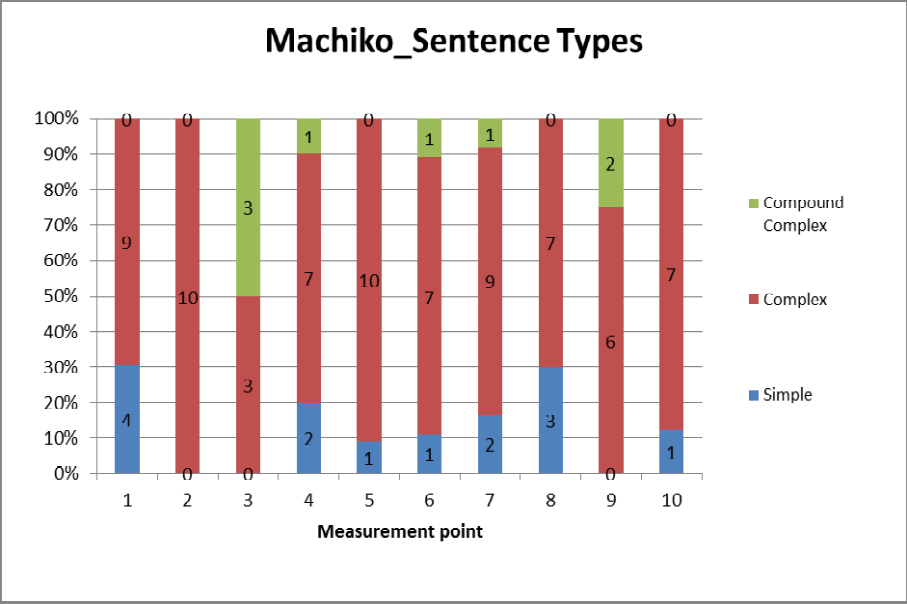


Figure 2. Sentence type distribution

In terms of accuracy development, Machiko’s writing shows a great degree of variability with more visible fluctuations. Figure 3 shows that Machiko’s accuracy level shot up to a perceived peak at measurement point 5 but dropped afterward and stayed low for a while. It then rose, again, to a quite similar peak at measurement point 8 and maintained a high accuracy level afterwards. Machiko’s accuracy level seemed to have improved toward the end of the semester.

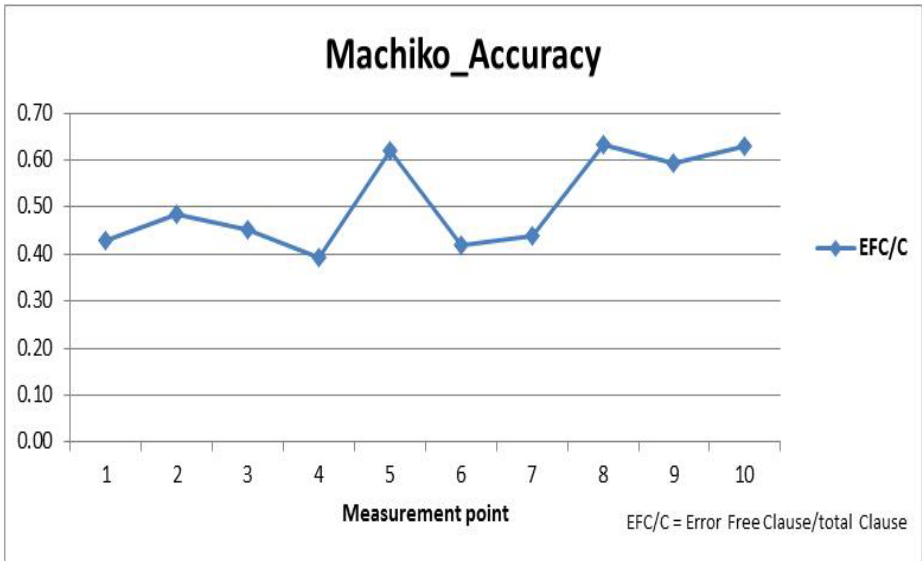


Figure 3. Accuracy development over one semester

To gain further insight into the types of errors Machiko made in her writing, each error occurrence was coded and calculated, and the results are shown in a distributional bar chart. Figure 4 shows the distribution of error types in the sample texts in percentage terms. Evidently, Machiko made many grammatical errors with the highest portion reaching 65% of the total errors at measurement point 7. As for the local errors, their occurrence was also evidenced in every sample text, with the highest portion being 60% of the total errors at measurement point 6. Interestingly, mechanical errors were also detected in the sample texts. The fact that Machiko was an advanced learner and that the assignments were written with a word processor would have given the impression that spelling and punctuation errors would not occur much. However, the occurrence of such errors was still evidenced in 90% of the sample texts (in 9 out of the 10 samples). A further investigation into the sample texts revealed that most punctuation errors resulted from the absence of a comma in its obligatory contexts, such as the one preceding a non-defining relative clause.

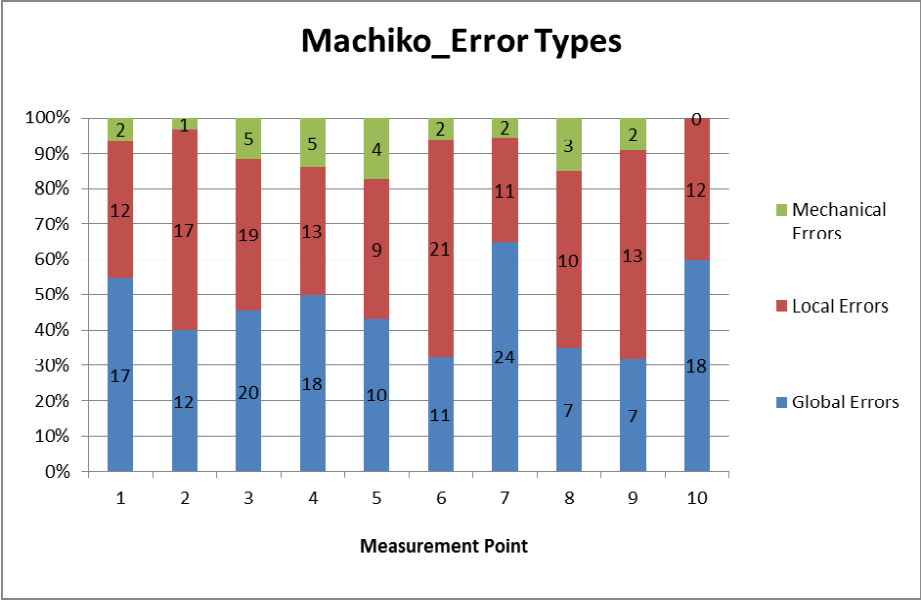


Figure 4. Error type distribution

RESULT 2: THE INTERACTION BETWEEN COMPLEXITY AND ACCURACY

To explore the interactions between complexity and accuracy in the sample texts, the data were first smoothed to accentuate the trends in the data. However, given the dynamic nature of the data, a linear regression was highly unlikely to provide the best fit for the data. Therefore, a two window moving averages method was chosen and the data were smoothed for both the complexity and accuracy indices. The results were then normalised for comparison purposes and presented in Figure 5.

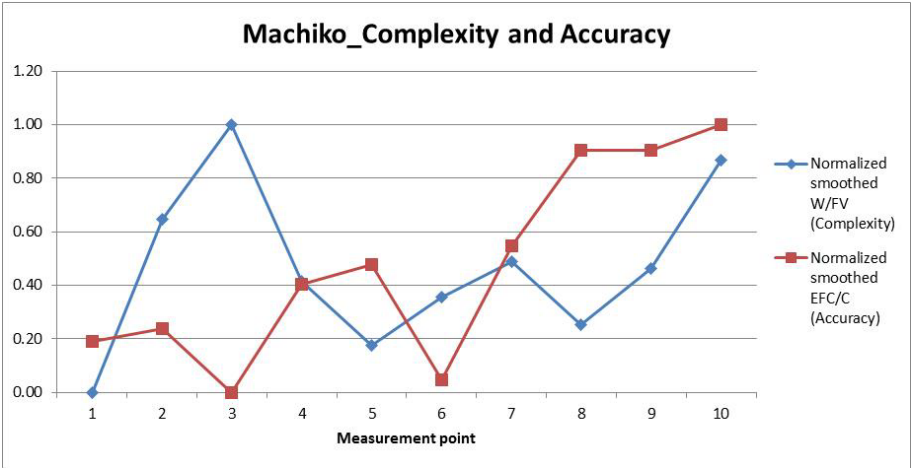


Figure 5. The interaction between complexity and accuracy

Figure 5 shows that the two indices were moving in the opposite direction for a while at the beginning of the semester, i.e., when complexity increased, accuracy decreased, although not to the same extent. However, towards the end of the semester, the two indices seemed to be moving together in the same direction. Visual inspection of the graph reveals that the nature of the relationship between the two variables changed from a negative association to a positive one. In order to illustrate such dynamics, a 3-window moving correlation analysis was performed to reveal the temporal interactions between the two measures. The result is shown in Figure 6.

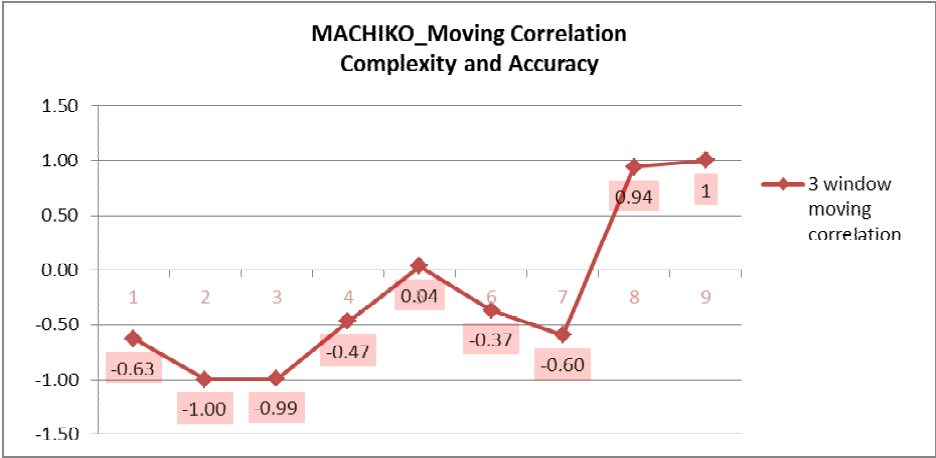


Figure 6. Moving correlation between complexity and accuracy

Within the ten sample texts, the association between complexity and accuracy measures fluctuated between a moderate and a strong negative association and then it changed to a strong positive relationship toward the end of the semester. The results suggest a potential trade-off relationship (Ferrari, 2012; Skehan, 2009; Skehan & Foster, 1999, 2007, 2012) between the two constructs at the beginning of the semester but such a competition vanished toward the end of the semester where the two constructs supported each other's development and became connected growers, in dynamic systems parlance. However, a further examination of the specific levels of both the complexity and accuracy measures, and an analysis of the rest of the texts, is needed before such a claim can be made. The possibility of other factors being at play, such as fluency, task characteristics, and task difficulty, etc., which affect and shape the interactions between the two constructs under observation (Housen et al., 2012; Levkina & Gilabert, 2012) should also be taken into consideration in the further attempt to explain the nature of their interaction.

DISCUSSION

The findings in this study show that both complexity and accuracy development reflect the behaviour of a dynamic system, i.e., non-linear, dynamic, and displaying a high degree of variability along its trajectories. Such findings are in line with the findings of Verspoor et al. (2004), Spoelman and Verspoor (2010), Vyatkina (2012, 2013), Polat and Kim (2013) and Thewissen (2013), in which a substantial amount of variability was evidenced in L2 development. These findings also lend support to the DST proposition of the existence of intra-individual variability in any (sub)system of language development (de Bot et al., 2012; de Bot et al., 2013; de Bot et al., 2007; van Dijk et al., 2011; van Geert, 2009a, 2009b, 2012; van Geert, Steenbeek, & van Dijk, 2011). Given the nature of the system, such dynamics, according to DST, are to be expected (de Bot, 2008; de Bot & Larsen-Freeman, 2011; Larsen-Freeman, 2009, 2010, 2011, 2012; Larsen-Freeman & Cameron, 2008; van Geert, 2008; Verspoor et al., 2008).

There is no single direction of development (Verspoor et al., 2008). Learners progress and regress; development ebbs and flows. Therefore simple terms of general trends are not sufficient to describe such a dynamic process. Each measure has its own pattern of development and is idiosyncratic (Jiang, 2013; Polat & Kim, 2013; Thewissen, 2013; Vyatkina, 2012, 2013). In fact, this micro level idiosyncrasy reflects what is going on at the macro level, i.e., each learner is idiosyncratic in terms of his/her development (de Bot, 2008; Kim & Sankey, 2009, 2010; Verspoor et al., 2008). The findings in this study therefore substantiate the DST proposition of L2 developmental dynamics.

Moreover, the interaction between (sub)systems in language development is also dynamic and non-linear. The interaction between complexity and accuracy in this study, too, was dynamic and non-linear. The two constructs were competing for a period of time before

changing their interaction towards a positive supporting relationship. Such a change in the nature of relationship is in accordance with the DST account of dynamic interaction in complex systems.

CONCLUSIONS

The results of this study support the DST proposition regarding the non-linearity of development and the high degree of variability during development. The two variables measured in this study displayed very different patterns. This augments the findings of Vyatkina (2012, 2013) and supports the claim that every development is idiosyncratic in its nature. The existence of variability throughout the developmental process also confirms the DST hypotheses about the behaviour of a dynamic system. Such high intra-individual variability was also detected in Spoelman and Verspoor (2010), and lends further support to the DST proposition about the central role of variability in shaping development.

Nevertheless, this study has demonstrated that complexity and accuracy constructs are highly dynamic and their development shows characteristics of a dynamic system, being non-linear and highly variable. The two constructs develop through different patterns and there is no single comprehensive term to name such pattern other than dynamic. Not only are the constructs dynamic and idiosyncratic, but also the learner, being the macro level of development, is idiosyncratic. No two learners experience the same development, and there is, therefore, no such thing as a typical/average learner.

LIMITATIONS AND FUTURE DIRECTIONS

This study is limited by the small number of text samples and the length of observation period (only one academic semester). The data analysed in this study were also sampled through a process of purposive sampling which was both statistically^x and methodologically^{xi} motivated, resulting in a corpus of about 2,100 words. Therefore, this study is limited in the amount of data analysed. Any findings, including the association between the two measures (complexity and accuracy) can therefore only be interpreted within the context of the sample texts; no claim beyond the scope of this study and beyond this set of data can be made. Whether or not the association will continue to exist, or change into another manifestation, will require more data points for further examination. Such an attempt is currently being undertaken and the results will be reported upon completion of the project.

Potential remedies^{xii} to the limitations in this study include attempting a more in-depth study with more participants and for a longer observation period. Ideally, the whole texts would be analysed instead of sample texts. Expanding the measures to include both global and specific measures of each construct would have the potential to unveil the dynamics underpinning the behaviour of the constructs. Lastly, qualitative analysis of the writing quality to complement

the quantitative findings would offer further insights into the development of second language academic writing.

ACKNOWLEDGEMENTS

The author of this paper would like to offer special thanks to Dr Aek Phakiti (Faculty of Education and Social Work, The University of Sydney) and the two anonymous reviewers for their valuable comments on the draft of this paper. All remaining weaknesses are my responsibility.

The author of this paper would also like to acknowledge the generous support of the Prime Minister's Australia Asia Endeavour Award (award no. PMPGI_DCD_3027_2012) without which this project would not have been possible.

REFERENCES

- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5–35.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment [Monograph]. *The Modern Language Journal*, 94(Supplement S1), 1–202. doi: 10.1111/j.1540-4781.2010.01136.x
- Caspi, T., & Lowie, W. (2010). A dynamic perspective on L2 lexical development in academic English. In R. Chacón-Beltrán, C. Abello-Contesse, & M. d. M. Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 41–60). Bristol: Multilingual Matters.
- Cooper, T. (1976). Measuring written syntactic patterns of second language learners of German. *Journal of Educational Research*, 69(5), 176–183.
- Council of Europe. (2001). *Common European Framework of Reference for language learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
- Cumming, A. (Ed.). (2006). *Goals for academic writing: ESL students and their instructors*. Philadelphia/Amsterdam: John Benjamins Publishing Company.
- de Bot, K. (2008). Introduction: Second language development as a dynamic process. *The Modern Language Journal*, 92(2), 166–178. doi: 10.1111/j.1540-4781.2008.00712.x
- de Bot, K., Chan, H., Lowie, W., Plat, R., & Verspoor, M. H. (2012). A dynamic perspective on language processing and development. *Dutch Journal of Applied Linguistics*, 1(2), 188–218. doi: 10.1075/dujal.1.2.03deb
- de Bot, K., & Larsen-Freeman, D. (2011). Researching second language development from a dynamic systems theory perspective. In M. H. Verspoor, K. de Bot & W. Lowie (Eds.), *A dynamic systems approach to second language development* (pp. 5–23). Amsterdam/Philadelphia: John Benjamins Publishing Company.

- de Bot, K., & Lowie, W. (2010). On the stability of representations in the multilingual lexicon. In M. Pütz & L. Scola (Eds.), *Cognitive processing in second language acquisition* (pp. 117–133). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- de Bot, K., Lowie, W., Thorne, S. L., & Verspoor, M. (2013). Dynamic Systems Theory as a comprehensive theory of second language development. In M. G. Mayo, M. G. Mangado, & M. M. Adrián (Eds.), *Contemporary approaches to second language acquisition* (pp. 199–220). Amsterdam: John Benjamins Publishing Company.
- de Bot, K., Lowie, W., & Verspoor, M. H. (2005). *Second language acquisition: An advanced resource book*. Oxon: Routledge.
- de Bot, K., Lowie, W., & Verspoor, M. H. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7–21. doi: 10.1017/S1366728906002732
- de Bot, K., Verspoor, M. H., & Lowie, W. (2005). Dynamic systems theory and applied linguistics: The ultimate “so what”? *International Journal of Applied Linguistics*, 15(1), 116–118. doi: 0.1111/j.1473–4192.2005.0083b.x
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Ferrari, S. (2012). A longitudinal study of complexity, accuracy, and fluency variation in second language development. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 277–297). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Flahive, D., & Snow, B. (1980). Measures of syntactic complexity in evaluating ESL compositions. In J. Oller & K. Perkins (Eds.), *Research in language testing* (pp. 171–176). Rowley, MA: Newbury House.
- Ho-Peng, L. (1983). Using T–unit measures to assess writing proficiency of ESL students. *REL C Journal*, 14(2), 35–43.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461–473. doi: 10.1093/applin/amp048
- Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy, and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1–20). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hunt, K. (1965). *Grammatical structures written at three grade levels* (NCTE Research Report No. 3). Urbana, IL: The National Council of Teachers of English.
- Hunt, K. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development*, 35(1), 1–67.
- Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
- Ingvarsdóttir, H., & Arnbjörnsdóttir, B. (2013). ELF and academic writing: A perspective from the Expanding Circle. *Journal of English as a Lingua Franca*, 2(1), 123–145.
- Jiang, W. (2013). Measurements of development in L2 written production: The case of L2 Chinese. *Applied Linguistics*, 34(1), 1–24. doi: 10.1093/applin/ams019

- Kim, M., & Sankey, D. (2009). Towards a Dynamic Systems Approach to moral development and moral education: A response to the JME special issue, September 2008. *Journal of Moral Education*, 38(3), 283–298.
- Kim, M., & Sankey, D. (2010). The dynamics of emergent self-organization: Reconceptualizing child development in teacher education. *Australian Journal of Teacher Education*, 35(4), 79–98.
- Larsen-Freeman, D. (1976). Evidence of the need for a second language acquisition index of development. In W. Ritchie (Ed.), *Second language acquisition research: Issues and implications* (pp. 127–136). New York: Academic Press, Inc.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12(4), 439–448.
- Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 141–165. doi: 10.1093/applin/18.2.141
- Larsen-Freeman, D. (2002). Language acquisition and language use from a Chaos/Complexity theory perspective. In C. Kramsch (Ed.), *Language acquisition and language socialization* (pp. 33–46). London: Continuum.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590–619. doi: 10.1093/applin/aml029
- Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, 30(4), 579–589. doi: 10.1093/applin/amp043
- Larsen-Freeman, D. (2010). The dynamic co-adaptation of cognitive and social views: A complexity theory perspective. In R. Batstone (Ed.), *Sociocognitive perspectives on language use and language learning* (pp. 40–53). Oxford: Oxford University Press.
- Larsen-Freeman, D. (2011). A complexity theory approach to second language development/acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 48–72). London & New York: Routledge.
- Larsen-Freeman, D. (2012). Complex, dynamic systems: A new transdisciplinary theme for applied linguistics. *Language Teaching*, 45(2), 202–214.
- Larsen-Freeman, D., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal*, 92(2), 200–213. doi: 10.1111/j.1540-4781.2008.00714.x
- Larsen-Freeman, D., & Storm, V. (1977). The construction of a second language acquisition index of development. *Language Learning*, 27(1), 123–134. doi: 10.1111/j.1467-1770.1977.tb00296.x
- Leki, I. (2011). Learning to write in a second language: Multilingual graduates and undergraduates expanding genre repertoires. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 85–110). Philadelphia/Amsterdam: John Benjamins Publishing Company.
- Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English* (Vol. Routledge): New York.
- Levkina, M., & Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 171–197). Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Lillis, T., & Curry, M. J. (2006). Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts. *Written Communication*, 26(1), 3–35. doi: 10.1177/0741088305283754
- Lillis, T., & Curry, M. J. (2010). *Academic writing in a global context: The politics and practices of publishing in English*. New York: Routledge.
- Lowie, W., & Verspoor, M. (2014, March). Variability and learning mechanisms. In K. de Bot & D. Larsen-Freeman (Chair), *From universality to variability in second language development*. Colloquium conducted at the meeting of the American Association for Applied Linguistics 2014, Portland, Oregon.
- Lowie, W., Verspoor, M., & de Bot, K. (2009). A dynamic view of second language development across the lifespan. In K. de Bot & R. W. Schrauf (Eds.), *Language development over the lifespan* (pp. 125–145). New York: Routledge.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 44(1), 36–62. doi: 10.5054/tq.2011.240859
- Manchón, R. M. (Ed.). (2012). *L2 writing development: Multiple perspectives*. Boston/Berlin: Walter de Gruyter.
- Mercer, S., Ryan, S., & Williams, M. (2012). Conclusion: Final Remarks. In S. Mercer, S. Ryan & M. Williams (Eds.), *Psychology for language learning: Insights from research, theory, and practice* (pp. 239–247). Hampshire: Palgrave Macmillan.
- Myles, F. (2012). Complexity, accuracy, and fluency: The role played by formulaic sequences in early interlanguage development. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 71–93). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578. doi: 10.1093/applin/amp044
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518. doi: 10.1093/applin/24.4.492
- Overton, W. F., & Lerner, R. M. (2014). Fundamental concepts and methods in Developmental Science: A relational perspective. *Research in Human Development*, 11(1), 63–73. doi: 10.1080/15427609.2014.881086
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590–601. doi: 10.1093/applin/amp045
- Paltridge, B., & Starfield, S. (2007). *Thesis and dissertation writing in a second language: A handbook for supervisors*. Oxon: Routledge.
- Polat, B., & Kim, Y. (2013). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics*, 1–25. doi: 10.1093/applin/amt013
- Richard, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge: Cambridge University Press.

- Silva, T. (1997). Differences in ESL and Native-English-Speaker writing: The research and its implications. In C. Severino, J. Guena & J. Butler (Eds.), *Writing in multicultural settings* (pp. 209–219). New York: Modern Language Association of America.
- Silva, T., & Reichelt, M. (2003). Second language writing up close and personal: some success stories. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 93–114). Cambridge: Cambridge University Press.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2009). Modeling second language development: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532. doi: 10.1093/applin/amp047
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93–120. doi: 10.1111/1467-9922.00071
- Skehan, P., & Foster, P. (2007). Complexity, accuracy, fluency and lexis in task-based performances: A meta-analysis of the Ealing research. In S. V. Daele, A. Housen, F. Kuiken, M. Pierrard & I. Vedder (Eds.), *Complexity, accuracy and fluency in second language use, learning and teaching* (pp. 207–226). Brussels: Contactforum.
- Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA* (pp. 199–220). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge, UK: Cambridge University Press.
- Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31(1), 1–22. doi: 10.1093/applin/amq001
- Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *The Modern Language Journal*, 97(S1), 77–101. doi: 10.1111/j.1540-4781.2012.01422.x
- van Dijk, M., & van Geert, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development*, 16(1), 7–33. doi: 10.1002/icd.506
- van Dijk, M., Verspoor, M. H., & Lowie, W. (2011). Variability and DST. In M. H. Verspoor, K. de Bot & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 55–84). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *The Modern Language Journal*, 92(2), 179–199. doi: 10.1111/j.1540-4781.2008.00713.x
- van Geert, P. (2009a). A comprehensive dynamic systems theory of language development. In K. de Bot & R. W. Schrauf (Eds.), *Language development over the lifespan* (pp. 60–104). New York: Routledge.
- van Geert, P. (2009b). Nonlinear complex dynamical systems in developmental psychology. In S. J. Guastello, M. Koopmans & D. Pincus (Eds.), *Chaos and complexity in psychology: The theory of nonlinear dynamical systems* (pp. 242–281). Cambridge: Cambridge University Press.

- van Geert, P. (2011). The contribution of Complex Dynamic Systems to Development. *Child Development Perspectives*, 5(4), 273–278.
- van Geert, P. (2012). Dynamic Systems. In B. Lauren, T. Little, & N. Card (Eds.), *Handbook of developmental research methods* (pp. 725–741). New York: Guilford Press.
- van Geert, P. (2014, March). L2 acquisition and the study of change in complex systems. In K. de Bot & D. Larsen-Freeman (Chair), *From universality to variability in second language development*. Colloquium conducted at the meeting of the American Association for Applied Linguistics 2014, Portland, Oregon.
- van Geert, P., Steenbeek, H., & van Dijk, M. (2011). A dynamic model of expert-novice co-adaptation during language learning and acquisition. In M. S. Schmid & W. Lowie (Eds.), *Modelling bilingualism: From structure to chaos: In honor of Kees de Bot*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behaviour and Development*, 25, 340–374. doi: 0163–6383/02
- van Geert, P., & van Dijk, M. (2003). Ambiguity in child language: The problem of inter-observer reliability in ambiguous observation data. *First Language*, 23(3), 259–284. doi: 10.1177/01427237030233001
- Verspoor, M. H., & Behrens, H. (2011). Dynamic systems theory and a usage-based approach to second language development. In M. H. Verspoor, K. de Bot & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 25–38). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Verspoor, M. H., de Bot, K., & Lowie, W. (2004). Dynamic Systems Theory and variation: A case study in L2 writing. In H. Aertsen, M. Hannay & R. Lyall (Eds.), *Words in their places: A festschrift for J. Lachlan Mackenzie* (pp. 407–421). Amsterdam: Department of English Language and Culture, Vrije Universiteit.
- Verspoor, M. H., de Bot, K., & Xu, X. (2011). The role of input and scholastic aptitude in second language development. *Toegepaste Taalwetenschap in Artikelen*, 86, 47–60.
- Verspoor, M. H., Lowie, W., & de Bot, K. (2009). Input and second language development from a dynamic perspective. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA*. Bristol: Multilingual Matters.
- Verspoor, M. H., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214–231. doi: 10.1111/j.1540–4781.2008.00715.x
- Verspoor, M. H., & Sauter, K. (2000). *English sentence analysis: An introductory course*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Verspoor, M. H., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21(3), 239–263. doi: 10.1016/j.jslw.2012.03.007
- Verspoor, M. H., & Smiskova, H. (2012). Foreign language writing development from a dynamic usage based perspective. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 17–46). Boston/Berlin: Walter de Gruyter.

- Verspoor, M. H., & van Dijk, M. (2011). Visualizing interactions between variables. In M. H. Verspoor, K. de Bot & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 85–98). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal*, 96(4), 576–598. doi: 10.1111/j.1540-4781.2012.01401.x
- Vyatkina, N. (2013). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, 97(S1), 11–30. doi: 10.1111/j.1540-4781.2012.01421.
- Witherington, D. C. (2007). The Dynamic Systems Approach as metatheory for Developmental Psychology. *Human Development*, 50, 127–153.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Manoa: Second Language Teaching and Curriculum Center, University of Hawai'i.

ENDNOTES

- i Wolfe-Quintero et al. (1998) recommended that the best measure for complexity was C/T, fluency W/T, and accuracy EFT/T; all of them were ratio measures using T-unit count as the denominator.
- ii The study, however, was designed as a repeated-task experiment and used the same task over the six-month observation period. Therefore, one may doubt whether the progress at the end of the study resulted from familiarity with the task through repetition or genuine development.
- iii A relatively stable state, in the DST parlance, refers to a state in which not much perceptible development takes place.
- iv The submission dates were: 03 Sept, 19 Sept, 21 Sept, 02 Oct, 05 Oct, 23 Oct, 05 Nov, 07 Nov, 21 Nov, and 23 Nov 2012 (in the order from the first task to the 10th task).
- v This ratio is calculated by first dividing the number of finite verbs in each sentence by the word length of each sentence and then the resulting values are averaged for the entire text in each sample.
- vi This study contains a very detailed list of errors in its EFL corpus.
- vii In this study, a two-window moving average smoothing technique was applied: T1 and T2 were first averaged, then T2 and T3 were averaged. Next, T3 and T4 were averaged, and this procedure was repeated until T9 and T10 were averaged.
- viii In this study, a three-window moving correlation analysis was conducted to explore the interaction between the complexity and accuracy measures. T1-T3, T2-T4, T3-T5, and so on were the time window for the moving correlation in this study. The correlation between the two measures in each time window was calculated and the resulting values were then plotted in a moving correlation graph.

-
- ix That compound sentences were not evidenced in these sample texts does not necessarily mean that Machiko did not produce compound sentences in her writing at all. A further examination of the rest of the texts is needed to complement the findings reported here.
 - x to normalise the size of each text for comparison purposes and to achieve better measures of complexity.
 - xi to filter out paragraphs with a substantial number of paraphrases and/or quotations.
 - xii These recommendations are taken into consideration and applied in the bigger project.