A Case Study in Complexity and Accuracy Development in ESL Academic Writing: A Dynamic Perspective

Rosmawati

The University of Sydney, Australia

0039

The Asian Conference on Education 2013

Official Conference Proceedings 2013

Abstract

This paper reports on a case study on the development of complexity and accuracy in an advanced English learner's academic writing over one academic semester. Studies on complexity and accuracy measures in second language (L2) development have shown diversified results. While some studies confirmed the Trade-off hypothesis prediction on the trade-off relationship between complexity and accuracy (Skehan and Foster, 1999, Skehan, 2009, Skehan and Foster, 2007), other studies demonstrated a joint-rise in both measures (Robinson, 2001, Robinson, 2003) due to cognitive demands of the task, hence advocating the Cognition hypothesis. Some other studies found no significant interactions between the two measures (Gunnarson, 2012, Levkina and Gilabert, 2012) and doubted the previous two hypotheses. Given the divergent results, there is a need to look into the nature of interactions between these two measures and unveil their developmental process to offer more insights into L2 writing development.

This study addresses this gap by exploring the dynamic unfolding of complexity and accuracy development in an advanced L2 learner's academic writing during her postgraduate study in Australia. The results suggested that complexity and accuracy measures displayed the characteristics of a dynamic system and their development was highly variable and non-linear. A moderate negative association was detected in the interaction between complexity and accuracy though not to a statistically significant level. The findings suggested that the developmental patterns of both the measures and the learner are highly dynamic and idiosyncratic. However, more longitudinal data are needed to explore the nature of interactions between complexity and accuracy in L2 development.

Keywords: complexity, accuracy, L2 academic writing, dynamic development, interaction

Introduction

In the attempt to gain insights into language acquisition and its subsequent development, researchers have been devising tools and measures to tap into learners' progress and gauge their development. The need for such means was first voiced out in the early 70s. Along with the expanding interest in the vast growing developmental studies, the search for an index to work as a standard yardstick to evaluate second language (L2) learners' progress was attempted (Wolfe-Quintero et al., 1998). In the early stage, length-based measures were borrowed from the field of first language (L1) acquisition to meet this demand and have since been widely adopted in the Second Language Acquisition (SLA) research enterprise, the most common ones being the mean length of particular structures (Norris and Ortega, 2009). Although these length-based measures are useful to some extent, they are not free of problems. Beginner L2 learners, for example, rely much on rote-learned formulaic chunks (Myles, 2012); therefore, perceived longer production of such structures results in an increase in length-based measures and gives false impressions of progress. Therefore, as a remedy, Larsen-Freeman (1976, 1978, Larsen-Freeman, 1977) proposed an Index of Development and this index was further operationalized as measures of Complexity, Accuracy, and Fluency (CAF).

Complexity measures "the extent to which the language produced in performing at task is elaborate and varied" (Ellis, 2003), while *accuracy* measures "the degree of conformity to certain norms" (Pallotti, 2009) and reflects "the conformity of second language knowledge to the target language norms" (Wolfe-Quintero et al., 1998). *Fluency*, on the other hand, gauges "how comfortable the second language [learner] is with producing [the target] language" (Wolfe-Quintero et al., 1998). Therefore, complexity taps into the learners' language knowledge while accuracy measures the appropriateness of language use and fluency the automaticity of language use. These three measures, as a triad, gauge learners' development.

Literature Review

One way of understanding how the L2 development unfolds is by tracking the development of CAF traits in the L2 learners' production (Skehan, 2009, Pallotti, 2009, Vyatkina, 2012). It is suggested that the development of these measures, along with their interactions, comprehensively captures the multidimensional facets of L2 development and is hence a measure of progress in language learning (Housen and Kuiken, 2009, Housen et al., 2012).

Researchers within the SLA enterprise have long taken interest in unveiling the nature of relationships among these three measures. Hypotheses have been put forward to explain and, to a certain degree, predict the interactions in the CAF triad. Among them, there are two relatively more influential hypotheses; they are: *Trade-Off Hypothesis* and *Cognition Hypothesis*.

As the name suggests, *Trade-off hypothesis* proposes trade-off relations among the CAF components due to the limited attentional capacity and working memory (Skehan, 2009, Skehan and Foster, 2012, Skehan and Foster, 1999, Skehan and Foster, 2007). That is to say that an increase of performance in one area, say complexity for example, is at the expense of the others, that is accuracy and/or fluency. In other words, directing attention to one aspect of performance may lead to an increase in that particular area but, at the same time, results in a lower performance in other aspects; hence the trade-off. As the result of such relationships, a rise of all the three components at the same time is not feasible although a joint rise of two is possible at the expense of another (Skehan and Foster, 2012).

On the other hand, *Cognition hypothesis* argues that given the possible simultaneous access to several attentional resources, higher task complexity may lead to concurrent rise in both complexity and accuracy. The argument was conceived by Robinson (2001) and one of the basic tenets in his hypothesis is that manipulating task difficulty will increase the cognitive demands on the learners and result in elevated performance in both complexity and accuracy (Robinson, 2003).

Both hypotheses propose different explanations to account for the nature of the interactions among the components of the CAF triad; the main difference being the underpinning reason to explain the perceived joint rise of complexity and accuracy. While Robinson maintains that the simultaneous increase of both measures corresponds to task difficulty level, Skehan (2009) argues against this proposition and asserts that the observed elevated performance in the two measures reflects the selective effects of task characteristics (Housen and Kuiken, 2009, Skehan and Foster, 2012). Both hyphotheses defend their own ground and refute the other party's stance.

This debate on the exact nature of the interactions in the CAF triad carries on into and is reflected in empirical studies. Since the 70's, research has been designed and studies attempted not only to observe but also to tease out these interactions. Given the scope of this paper, only the more recent studies are discussed here. For a synthesis of previous empirical findings, see Norris and Ortega (2009), Wolfe-Quintero et al. (1998), Ortega (2003).

Ferrari's (2012) study found traces of trade-off effects between complexity and accuracy in a certain time period in her participants' longitudinal development. Such interactions were also confirmed in Myles (2012), not only among the CAF dimensions but also between the triad and the learners' communicative adequacy. Both studies provided empirical evidence to support the Trade-off hypothesis to a certain extent. Skehan and Foster (2012) added that the hypothesis, coupled with the selective effect of task characteristics, was sufficient to explain the interactions between complexity and accuracy measures, independent of task difficulty as purported by Robinson's Cognition hypothesis. Kuiken and Vedder's (2012) study further augmented this stance and concluded that no evidence was found in their study

to show the effects of task cognitive difficulty on syntactic and lexical complexity although an increase in accuracy was detected. Further examination on the data showed that the rise of accuracy level reflected the fact that the learners were directing their attention to forms and hence fewer lexical errors.

Gunnarson (2012), on the other hand, found no competition between complexity and accuracy and pointed out that there was also no significant interactions detected between syntactic complexity and fluency in her study. The results of her study challenged the propositions advocated by the Trade-off hypothesis. More doubts were put forward by Levkina and Gilabert (2012) and their study showed the combined effect of task complexity and planning time on fluency (negative effect) and lexical complexity (positive effect) but no significant change was detected in both syntactical complexity and accuracy. This study questioned both the Trade-off hypothesis and the Cognition hypothesis as none of them "can satisfactorily explain" their findings (Levkina and Gilabert, 2012).

Given the multidimensional facets of L2 development, it is not surprising to find such divergent results. In fact, linear relationships among the CAF components and their simple correspondence to a given condition/task are not to be expected from such a multi-componential and dynamical system (Housen et al., 2012). With regard to this, Norris and Ortega (2009) raised an ontological issue concerning the way CAF triad is perceived and operationalized. Pointing at the contemporary practice, they suggested that there is "a lack of attention to CAF as a dynamic and interrelated set of constantly changing subsystems" (Norris and Ortega, 2009) and called for more longitudinal observations if the nature of the CAF development and interactions were to be explored. Similar concern was voiced by Larsen-Freeman (2009) and Dynamic Systems Theory was proposed as a suitable, and possibly the most potentially fruitful, framework within which CAF studies could be attempted and CAF components treated as dynamic (sub)systems.

In its essence, Dynamic Systems Theory (DST) is a theory about how dynamic system changes and develops over time and how complexity emerges out of such behaviour (de Bot, 2008, de Bot et al., 2007, Larsen-Freeman, 2002, van Geert, 2008, de Bot et al., 2005b). Originated in the field of natural sciences, DST was originally developed to find explanation to account for systems that "seem to be chaotic and self-organising" (de Bot et al., 2005a). Though in its early development DST was a purely mathematical approach, its later stages offer practical tools and perspectives into researching the process of *change* (de Bot, 2008). Its current application includes a great diversity of fields ranging from meteorology for weather forecast to ornithology for explaining the bird flock flight patterns, and also across disciplines like economics and laws (Kellert, 2008).

The integration of DST into the SLA research enterprise introduces novel perspectives into the field. Instead of the popular way of looking for causal

relationships between affective factors and learning outcomes and establishing a neat pattern to account for the acquisition process in general, DST suggests a more coherent view on the developmental process itself. Development is viewed a dynamic and non-linear process, coloured by both progress and regress, and its trajectory displays a great amount of variability (Larsen-Freeman, 1997, Verspoor and Behrens, 2011). Language acquisition is one such developmental process in which the dynamic interactions among its variables result in the complexity of the system's behaviour.

Given the dynamic nature of the system and all its highly interconnected components, the outcome of their interactions is bound to be non-linear too, i.e., they change over time. The divergent results evidenced so far are in fact snapshots of separate moments along the development and hence the diversity. Such discrete pictures of different time junctures, however, are not sufficient to capture the dynamism. In order to gain more insights into the nature of the systems and its development, it is essential to attempt more longitudinal observations (van Dijk et al., 2011, Larsen-Freeman and Cameron, 2008, Norris and Ortega, 2009).

Following this line of suggestions, some studies have been endeavoured within the DST framework and the most recent ones are discussed here. Verspoor et al. (2008) did a longitudinal study on an advanced learner of English for a period of 3 years and collected 18 academic writing samples from this participants. They then proceeded to look for the relationship between two measures of complexity, i.e. vocabulary and sentence complexity measures. The findings of this study confirmed that the two measures are supporting each other in their development: they are, in dynamic parlance, *connected growers*.

Another longitudinal study of equal observation period length is Spoelman and Verspoor's (2010) study which looked into the interaction between accuracy rates and complexity measures in a Dutch student learning Finnish via academic setting for a period of 3 years. They concluded that accuracy rates fluctuated considerably in early stages but soon settled down as the system relaxed. However, they also pointed out that interaction between accuracy and complexity measures changed over time, confirming the DST proposition about the system's behaviour.

Larsen-Freeman's (2006) study also supported the DST claim about high intra- and inter- individual variability as she found that each participant in her study showed different developmental trajectories although the overall group average of CAF measures showed a general increasing trend over time. However, this study 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 reflected genuine development.

Similar results were also found in Vyatkina (2012) in which a general upward trend in the development of complexity measures in L2 writing was detected along with significant variability between individual and cross-sectional data. Using developmental profiling techniques, she pursed further to demonstrate how each participant followed different developmental paths and displayed different developmental patterns (Vyatkina, 2013).

Polat and Kim (2013) expanded the discussion to include two out of the three components in the CAF triad, tracing the dynamics of complexity and accuracy development of an immigrant in the USA. In this study, it was found that syntactic complexity and lexical diversity developed well in untutored situation which was the context of this study while accuracy seemed very constrained. They concluded that the participant's interlanguage was highly variable but was perhaps nested within a stable state. Polat and Kim's study was, in fact as claimed, one of the first to attempt a longitudinal observation on naturalistic learning. More such studies were called for in order to unveil the nature of L2 development, and more specifically, CAF development and interactions.

This current study addresses this need. Designed as a case study, the current study explores the dynamic unfolding of complexity and accuracy development in an advanced English learner's academic writing over one academic semester during her postgraduate study in Australia. Along with the rapid growth of global education and the corresponding demand on English academic writing skill, there is an urgent need to look into and gain more understanding of how L2 academic writing develops. This study looks for developmental transitions and identifies interactional pattern(s) between the two measures (complexity and accuracy) over time by applying variability analyses within DST framework (van Dijk and van Geert, 2007). This paper then contributes towards unveiling the dynamic relationship between these two measures and advancing our understanding of L2 writing development.

Research Questions

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

1. What is the nature of the development of complexity and accuracy measures in L2 academic writing?

2. How do the two measures interact over time during the one academic semester observation period?

Research Design

This paper is a single-case quantitative study based on longitudinal observation of a participant's written production over one academic semester. Although the most

common association of a case study label is to qualitative studies, this paper is in fact a quantitative approach.

Participant and Setting

The participant in this study is Mai (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 for about 15 years. As required for program admission, she took a standardized English test (i.e. IELTS) prior to commencing her study and achieved a score that was equivalent to B2 level on a CEFR scale, hence an advanced learner.

Data

This study adopted a *time-series* approach to follow the development of complexity and accuracy in the written output (academic essays) of the participant over one academic semester. These essays were the assignments for the courses she was enrolled in. These assignments abided by the guidelines set up by the faculty and the course coordinator. Following their submission, the copies of those assignments were sent to the researcher and served as the data for this study.

The data for this study were, therefore, the academic assignments Mai wrote and submitted for the courses she was enrolled in during the first semester. The rationale for choosing to collect the data from the very first semester was underpinned by the assumption within DST framework that sudden proliferation of both input and use may set the whole (learning) system into chaos and hence resulting in high degree of variability which benchmarks the onset of transitional stages that eventually lead to development. DST maintains that any dynamic system has the potentials to fall into chaos and then restructures; therefore, contrary to the common belief, “even for an advanced learner, the system can be far from stable” (Verspoor et al., 2008). Therefore, high degree of variability was expected in this study as it would then offer the key information to unveil the nature of development (van Dijk et al., 2011, de Bot and Larsen-Freeman, 2011).

Sampling and Coding

The data were then coded for complexity and accuracy. However, as some complexity measures, including those adopted in this study, are very sensitive to text length and correlate negatively with word count (Wolfe-Quintero et al., 1998, Spoelman and Verspoor, 2010), the text length in this study was controlled. A purposive sampling of approximately 200 words (±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 the result of this purposive sampling, a total of 10 pieces of sample texts were obtained. These sample texts were then submitted to two coders for coding.

Firstly, the data were coded for sentence types: simple (Si), compound (Co), complex (Cx) and compound-complex (CoCx) sentences. The results were then tallied. Then, instead of employing the 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 complexity, this study adopted another type of measure. As Verspoor et al. (2008) pointed out, separate indices do not capture the entire degree of complexity because "they do not bring to light additional complex construction such as longer NPs or non-finite constructions" especially in the case of advanced learners. Following their suggestion, this study employed **W/FV** (word per finite verb) ratio to calculate the overall degree of sentence complexity. Higher indices mean the more complex the sentences are.

To calculate accuracy, however, the errors detected in the sample texts were coded as global (GE: errors at grammar and style levels), local (LE: errors at word, lexical, and lexico-grammatical levels), and mechanical errors (ME: errors at form/spelling and punctuation level). Their occurrences were then tallied. Following that, the number of error free clauses (EFC) were counted and then compared to the total number of clauses (C) to obtain the **EFC/C** ratio. Such a proportion index of accurate production is highly recommended as it tells "the true story in measuring the accuracy of the learners' written production" (Jiang, 2013).

Inter-coder Reliability

As mentioned in the previous section, the sample texts were submitted to two coders (the author as coder 1, and an English teacher with more than 8 years of teaching experience as coder 2 and was given training sessions prior to doing the coding). 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. What Pearson's R value measures is whether or not the total number of confirmed cases in each sample picked by coder 1 increases (or decreases) in the same fashion as the total number picked by coder 2. It gives in a high value as long as the fashion of increase (or decrease) confirms each other although the two coders may actually pick totally different cases. Hence, it is not a genuine reflection of agreement or reliability. To measure reliability, this study therefore adopted an overlap index which reflects the percentage of overlapping positive cases confirmed by both coders (van Geert and van Dijk, 2003). It is a relatively better measure of reliability as it shows how many cases were actually picked and confirmed by both coders.

In this study, the POR value reached 95% for complexity measure and 78% for accuracy measure. As accuracy measures 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 and van Dijk, 2003). Justification of the quality of coding can be attempted

through explicitly stated procedures; the results are then considered valid and trustworthy (van Geert and van Dijk, 2003). In this study, disagreement between coders was resolved by a discussion, typically deferring to the first coder who designed both the study and the coding scheme.

Statistical Procedures

The coded data were then submitted to two stages of analyses: descriptive and correlation analyses; each stage corresponded to each research question respectively. At the descriptive stage, the data were plotted into a developmental graph and a distributional bar chart was also presented to show the portion of each type of sentences in the sample texts. This stage of analysis explored the nature of the development and hence answered the first research question. Then, a correlation analysis was conducted to explore the association between the two measures. The results of this analysis answered the second research question regarding the relationship between the two variables.

Results and Discussions

Over the one academic semester, the development of complexity and accuracy measures in Mai's writing showed a great deal of variability. The data series collected from her academic writing were analysed and the results were presented below.

Result 1: The nature of development of complexity and accuracy

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

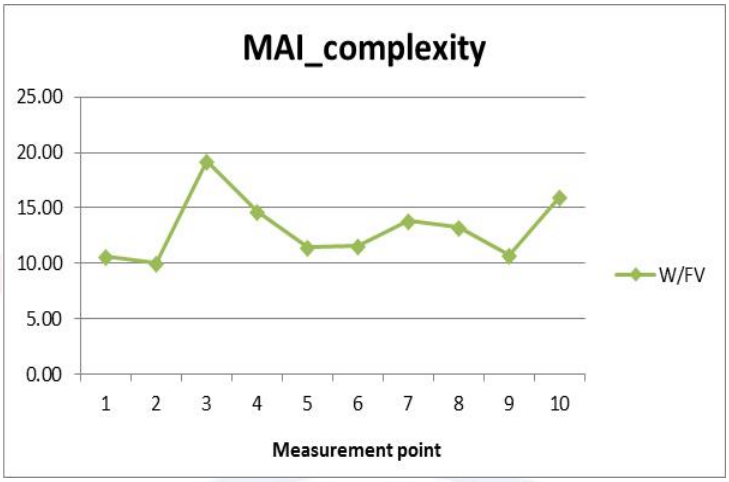

Figure 1. Complexity development over one semester

Figure 2 gives a better description of how complex the sentences were in the texts, as it shows the distribution of sentence types in each text. As can be seen in Figure 2, it was found that Mai produced only simple sentences and complex sentences in these 10 sample texts (with one exception: one occurrence of compound-complex sentence was evidenced in text number 3).

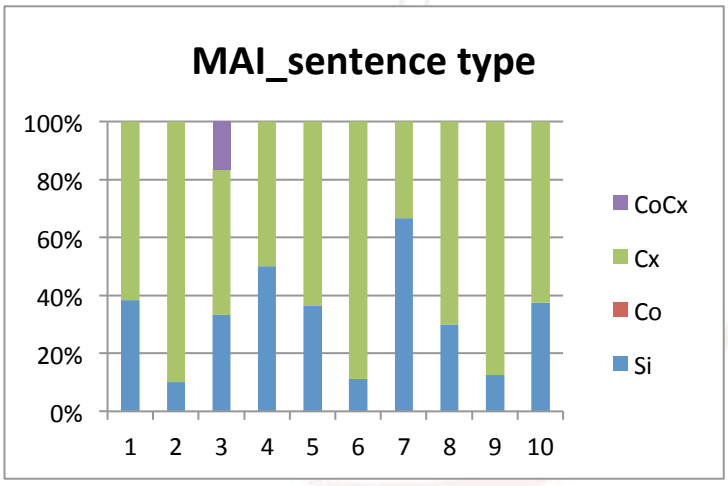

Figure 2. Sentence type distribution

There are two implications of this finding. First, the fact that the graph in Figure 1 showed a peak in measurement point no. 3 overlapped with this occasion. In a sense, it validates the measurement employed in this study (W/FV) as a sensitive, and hence suitable, measure to gauge complexity. Second, the fact that the occurrence of compound sentences was not evidenced in these sample texts does not necessarily mean that she does not produce compound sentences at all in her writings. A further examination into the rest of the texts is needed to complement this.

In terms of accuracy development, Mai's writing shows that a great degree of variability with more visible fluctuations as shown in Figure 3. It can be seen in the graph that Mai's accuracy level dropped much in measurement point 3 and raised to a perceived peak at measurement point 9. It is very interesting to see that the two occasions matched with the development of complexity measure though in an opposite direction. Whether or not a negative association can be inferred is subject to further analysis on the correlation between the two variables throughout all the measurement points instead of just two most visible occasions.

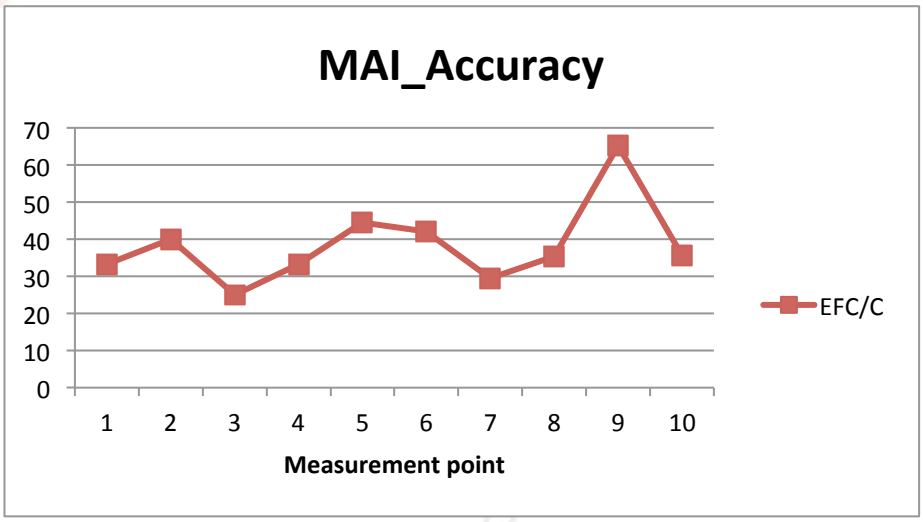

Figure 3. Accuracy development over one semester

To gain more insights into the types of errors Mai made in her writing, each error occurrence was coded and calculated and the result was mapped into a distributional bar chart. Figure 4 shows the distribution of error types in the sample texts in percentage. Evidently, Mai made many grammatical errors with the highest portion reaching 70% of the total errors in measurement point 7. As for local level of errors (including word, lexical and lexico-grammatical errors), the occurrence of such errors was also evidenced in every sample text with the highest portion of 70% of the total errors in measurement point 8. Interestingly, mechanical errors were also detected in the sample texts. The fact that the participant was an advanced learner and that the assignments were actually written with a word processor would have given the impression that spelling and punctuation errors were not to occur at all. However, the occurrence of such errors was still evidenced in 70% of the sample texts (in 7 out of the 10 samples).

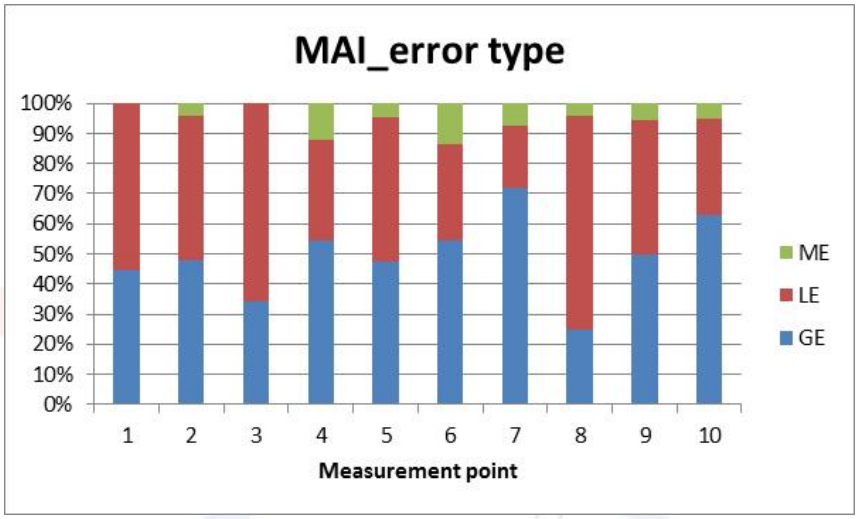

Figure 4. Error type distribution

These findings show that both complexity and accuracy development reflects the behaviour of a dynamic system, i.e. non-linear, dynamic, and displaying a high degree of variability along its trajectories. There is no single direction of development and therefore simple terms of general trend are not sufficient to do justice in describing such a dynamic process. Each measure shows its own pattern of development and is hence idiosyncratic. In fact, this micro level idiosyncrasy reflects what is going on at macro level, i.e. each learner is idiosyncratic in terms of his/her development.

Result 2: The interaction between complexity and accuracy

To explore the interactions between complexity and accuracy in the sample texts, a correlation analysis was performed. In addition, a regression line was also added to the data to visually show the association between the two measures. To accommodate the fact that both measures are dynamic and non-linear, a third degree polynomial regression was conducted as a linear regression was highly unlikely to fit such a non-linear set of data. The result is shown in Figure 5.

The plotted raw data points show that the two lines are moving in an opposite direction, i.e. when complexity increases, accuracy decreases, although not to the same extent. The superimposed polynomial trend lines confirm the trend of such behaviour in the interactions among these two variables in a more pronounced way. In fact, the result of correlation analysis adds to support the existence of negative association between complexity and accuracy in this set of data though not to a statistically significant level ($r = -.616$, $p > .05$).

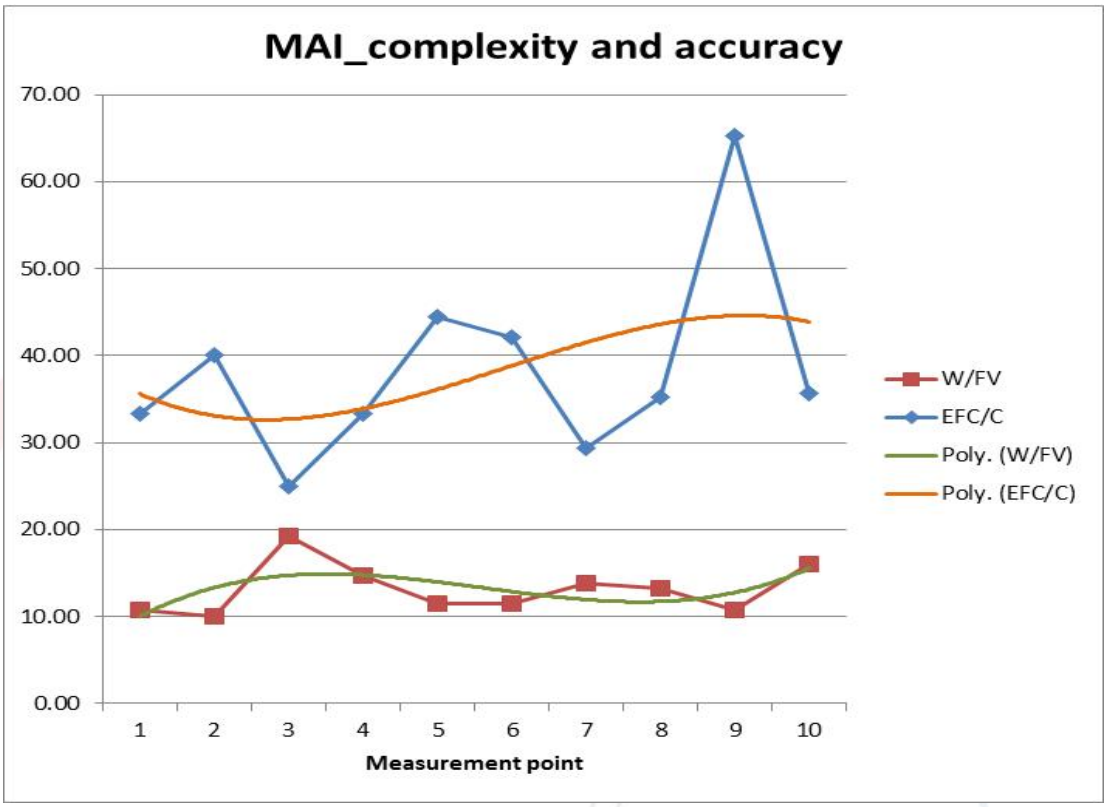

Figure 5. The interaction between complexity and accuracy

What this result suggests is that within these ten sample texts, a moderate negative association was detected between complexity and accuracy measures but the association was not statistically significant. Whether or not it translates into a trade-off relation, however, needs further examination both into specific levels of complexity and accuracy measures as well as into the rest of the texts. There are also possibilities of other factors being at play, like fluency measure, task characteristics or difficulty, etc., which affect and shape the interactions between the two variables under observation. This, too, has to be taken into consideration before any conclusion can be made.

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

Conclusions

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

Limitations and Future Directions

This study is very limited by the small number of samples and the length of observation period being only one academic semester. Not only did the data come from just one single participant, the data were also sampled through a process of purposive sampling to obtain a set of sample texts. Hence, this study is limited in both the number of participant and the amount of data analysed. Any findings, including the negative association between the two measures (complexity and accuracy) evidenced in this study 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 in the future will need more data points for further examinations.

Potential remedies to the limitations in this study include attempting a more in-depth study with more participants and in a longer observation period. Ideally, the whole texts are to be analysed instead of sample texts. Expanding the measures to include both global and specific measures of each construct will have the potentials to unveil the dynamism underpinning the behaviour of the constructs. Lastly, a touch of qualitative analysis on the quality of the writing to complement the quantitative findings will offer more insights into the development of second language academic writing.

References

DE BOT, K. 2008. Introduction: Second language development as a dynamic process. *The Modern Language Journal,* 92**,** 166-178.

DE BOT, K. & LARSEN-FREEMAN, D. 2011. Researching second language development from a dynamic systems theory perspective. *In:* VERSPOOR, M. H., DE BOT, K. & LOWIE, W. (eds.) *A dynamic systems approach to second language development.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

DE BOT, K., LOWIE, W. & VERSPOOR, M. H. 2005a. *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**,** 7-21.

DE BOT, K., VERSPOOR, M. H. & LOWIE, W. 2005b. Dynamic systems theory and applied linguistics: The ultimate "so what"? *International Journal of Applied Linguistics,* 15**,** 116-118.

ELLIS, R. 2003. *Task-based language learning and teaching,* Oxford, Oxford University Press.

FERRARI, S. 2012. A longitudinal study of complexity, accuracy, and fluency variation in second language development. *In:* HOUSEN, A., KUIKEN, F. & VEDDER, I. (eds.) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

GUNNARSON, C. 2012. The development of complexity, accuracy and fluency in the written production of L2 French. *In:* HOUSEN, A., KUIKEN, F. & VEDDER, I. (eds.) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

HOUSEN, A. & KUIKEN, F. 2009. Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics,* 30**,** 461-473.

HOUSEN, A., KUIKEN, F. & VEDDER, I. 2012. Complexity, accuracy, and fluency: Definitions, measurement and research. *In:* HOUSEN, A., KUIKEN, F. & VEDDER, I. (eds.) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

JIANG, W. 2013. Measurements of development in L2 written production: The case of L2 Chinese. *Applied Linguistics,* 34**,** 1-24.

KELLERT, S. H. 2008. *Borrowed knowledge: Chaos theory and the challenge of learning across disciplines,* Chicago, The University of Chicago Press.

KUIKEN, F. & VEDDER, I. 2012. Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. *In:* HOUSEN, A., KUIKEN, F. & VEDDER, I. (eds.)

Dimensions of L2 Performance and Proficiency: Complexity, accuracy and fluency in SLA. Amsterdam/Philadelphia: John Benjamins Publishing Company.

LARSEN-FREEMAN, D. 1976. Evidence of the need for a second language acquisition index of development. *In:* RITCHIE, W. (ed.) *Second language acquisition research: Issues and implications.* New York: Academic Press, Inc.

LARSEN-FREEMAN, D. 1977. The construction of a second language acquisition index of development. *Language Learning,* 27**,** 123.

LARSEN-FREEMAN, D. 1978. An ESL index of development. *TESOL Quarterly,* 12**,** 439-448.

LARSEN-FREEMAN, D. 1997. Chaos/complexity science and second language acquisition. *Applied Linguistics,* 18**,** 141-165.

LARSEN-FREEMAN, D. 2002. Language acquisition and language use from a Chaos/Complexity theory perspective. *In:* KRAMSCH, C. (ed.) *Language acquisition and language socialization.* 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**,** 590-619.

LARSEN-FREEMAN, D. 2009. Adjusting expectations: The study of complexity, accuracy and fluency in second language acquisition. *Applied Linguistics,* 30**,** 579-589.

LARSEN-FREEMAN, D. & CAMERON, L. 2008. Research methodology on language development from a complex systems perspective. *The Modern Language Journal,* 92**,** 200-213.

LEVKINA, M. & GILABERT, R. 2012. The effects of cognitive task complexity on L2 oral production. *In:* HOUSEN, A., KUIKEN, F. & VEDDER, I. (eds.) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

MYLES, F. 2012. Complexity, accuracy, and fluency: The role played by formulaic sequences in early interlanguage development. *In:* HOUSEN, A., KUIKEN, F. & VEDDER, I. (eds.) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA.* 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**,** 555-578.

ORTEGA, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics,* 24**,** 492-518.

PALLOTTI, G. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics,* 30**,** 590-601.

POLAT, B. & KIM, Y. 2013. Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics*, 1-25.

ROBINSON, P. 2001. Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. *In:* ROBINSON, P. (ed.) *Cognition and second language instruction.* Cambridge: Combridge University Press.

ROBINSON, P. 2003. Attention and memory during SLA. *In:* DOUGHTY, C. & LONG, M. H. (eds.) *The handbook of second language acquisition.* Malden, Ma: Blackwell.

SKEHAN, P. 2009. Modeling second language development: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics,* 30**,** 510-532.

SKEHAN, P. & FOSTER, P. 1999. The influence of task structure and processing conditions on narrative retellings. *Language Learning,* 49**,** 93-120.

SKEHAN, P. & FOSTER, P. 2007. Complexity, accuracy, fluency and lexis in task-based performances: A meta-analysis of the Ealing research. *In:* DAELE, S. V., HOUSEN, A., KUIKEN, F., PIERRARD, M. & VEDDER, I. (eds.) *Complexity, accuracy and fluency in second language use, learning and teaching.* Brussels: Contactforum.

SKEHAN, P. & FOSTER, P. 2012. Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research. *In:* HOUSEN, A., KUIKEN, F. & VEDDER, I. (eds.) *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

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**,** 1-22.

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**,** 7-33.

VAN DIJK, M., VERSPOOR, M. H. & LOWIE, W. 2011. Variability and DST. *In:* VERSPOOR, M. H., DE BOT, K. & LOWIE, W. (eds.) *A dynamic approach to second language development: Methods and techniques.* 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**,** 179-199.

VAN GEERT, P. & VAN DIJK, M. 2003. Ambiguity in child language: The problem of inter-observer reliability in ambiguous observation data. *First Language,* 23**,** 259-284.

VERSPOOR, M. H. & BEHRENS, H. 2011. Dynamic systems theory and a usage-based approach to second language development. *In:* VERSPOOR, M. H., DE BOT, K. & LOWIE, W. (eds.) *A dynamic approach to second language*

development: Methods and techniques. Amsterdam/Philadelphia: John Benjamins Publishing Company.

VERSPOOR, M. H., LOWIE, W. & VAN DIJK, M. 2008. Variability in second language development from a dynamic systems perspective. *The Modern Language Journal,* 92**,** 214-231.

VYATKINA, N. 2012. The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal,* 96**,** 576-598.

VYATKINA, N. 2013. Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal,* 97**,** 11-30.

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.

iafor