

Dynamic Paths of Complexity and Accuracy in Second Language Speech: A Longitudinal Case Study of Chinese Learners

^{1,*}HANJING YU and ²WANDER LOWIE

¹School of Foreign Languages at Dalian University of Technology, China and

²Department of Applied Linguistics at the University of Groningen, the Netherlands

*E-mail: yuhanj823@126.com

The present study investigated the developmental patterns of Chinese EFL(AQ) learners' oral language in terms of complexity and accuracy and looked into the dynamic interactions between them within the framework of Complex Dynamic Systems Theory (CDST). The data were analysed using dynamic analyses (moving min–max graphs, moving correlations and Monte Carlo Simulations). It was found that, firstly, at the group level, the general developmental trends of both complexity and accuracy showed improvements. Secondly, at the individual level, the developmental patterns were non-linear and dynamic with high degrees of variability, and individual language development was influenced by the initial states. Thirdly, the analyses revealed a complex interplay between complexity and accuracy, which gradually shifted from a clearly competitive relationship during the early stages to a supportive relationship in later stages. This shift in interaction shows that complexity goes hand in hand with accuracy, which corroborates the interconnectedness of subsystems as one of the major characteristics of CDST. The findings confirm the applicability of CDST approaches to L2 oral development and carry valuable implications for CDST theory development and oral language teaching.

INTRODUCTION

The assessment of oral language can be regarded as a useful way to measure general second language (L2) proficiency, as it shows all aspects of active language use, including, for example, productive vocabulary, use of tenses and sentence construction. What is more, oral production is more 'spontaneous' than written production due to different time constraints. However, compared to the research of L2 written proficiency, so far, a relatively small number of studies has focused on the developmental process of L2 oral production, possibly due to the difficult and time-consuming process of collecting, storing, manipulating, and analysing spoken language. In addition, most previous studies of L2 oral production have applied product-oriented methods by focusing on a dependent variable measured at one moment in time, usually in the form of averaging group scores, and regarding the variability between learners in the

data as a measurement error. General changes in group averages can indicate general trends of oral language development, but cannot capture the developmental process leading to these results. From the perspective of Complex Dynamic Systems Theory (CDST), 'variability' can be seen as a meaningful source of information on the process of language development and the differences between and within individual learners, and can therefore serve as the focal point of a study exploring the dynamicity of oral language development (Van Geert and Van Dijk 2002).

By tracing Chinese English learners' L2 oral production over four months (one semester), the present study intends to investigate the developmental patterns of oral production and compares the inter- and intra-individual variability, both at the group and individual level. In addition, we will look into the interactions between complexity and accuracy to explore the possibilities of CDST-inspired approaches towards the study of oral language development and possible implications for the teaching of oral language use to Chinese learners.

After a short discussion of some theoretical implications of CDST, we will describe the process of oral language development of two Chinese learners of English in detail and estimate the applicability of CDST approaches to researching oral language development. Finally, we will examine the implications of this study for Chinese English oral language teaching in the future.

DYNAMIC DEVELOPMENT

Dynamic Systems Theory is partly a theory of variability and has been used extensively to describe the non-linear development of systems in a wide range of scientific areas, like biology and physics. After its introduction into the fields of first language development (Van Geert and Van Dijk 2002), L2 development (Larsen-Freeman 1997; De Bot *et al.* 2007) and cognitive science and psychology (Thelen and Smith 1994; Van Gelder and Port 1995), an increasing number of researchers have acknowledged that L2 development is a dynamic, non-linear and self-organizing process. CDST-based studies have adopted a process-oriented method, investigating groups or individuals longitudinally and considering variability as an inherent property of language development (De Bot *et al.* 2007). Within a CDST framework, dynamic characteristics of language development have been investigated by studying the patterns of variability over time and by charting the dynamic interaction of related sub-systems (see Verspoor *et al.* 2011).

Variability

As an indicator of an ongoing process and even the potential driving force of development, variability over time in the data of the same learner may be a significant source of information about language development (Van Geert and Van Dijk 2002). This information may help explain how a system changes over

time and how it behaves when it is in a relatively stable or unstable state (Verspoor *et al.* 2011). For instance, a high degree of variability can be interpreted as an indication that developmental changes are taking place in one or more subsystems (Spoelman and Verspoor 2010). A lower degree of variability, on the other hand, means that the system is relatively stable, settling into what is called an attractor state before the next change in the system takes place and variability increases again. Larsen-Freeman (2009) argued that since variability is such an important source of information about the underlying language development process, variability within individuals should be a primary centre of the research focus.

A focus on individual development is further emphasized by Molenaar (2015) from a statistical point of view. Molenaar argued that there is no relation between results obtained in statistical analyses of inter- and intra-individual variation, as inter-individual variation concerns the relationships between variables for sampled subjects at one point in time, while intra-individual variation involves the analysis of time-dependent changes of an individual's performance. The former focuses on the population level, rather than on the individual level. In generalizing developmental patterns across learners, we assume subjects 'to be mere replications' (Molenaar 2015: 36) and the individuality of the subjects is considered unimportant. Molenaar explained that generalizing average group scores from studies on inter-individual variation to the analysis of intra-individual variability can only be done when the study on inter-individual variation meets the condition of ergodicity, or in other words, if the individuals in a group form an 'ergodic ensemble' (see Lowie and Verspoor (2019) for a detailed discussion). Ergodicity is clearly illustrated by Tarko (2005):

Suppose you are concerned with determining what the most visited parks in a city are. One idea is to take a momentary snapshot: to see how many people are this moment in park A, how many are in park B and so on. Another idea is to look at one individual (or few of them) and to follow him for a certain period of time, e.g. a year. Then, you observe how often the individual is going to park A, how often he is going to park B and so on. Thus, you obtain two different results: one statistical analysis over the entire ensemble of people at a certain moment in time, and one statistical analysis for one person over a certain period of time. The first one may not be representative for a longer period of time, while the second one may not be representative for all the people. The idea is that an ensemble is ergodic if the two types of statistics give the same result. Many ensembles, like the human populations, are not ergodic. (Tarko 2005)

Ergodicity has two conditions. Firstly, the process has to be stable over time, indicating that each measurement shows the same mean and variance. Secondly, there should be a homogeneous population, which means all participants follow the same dynamic patterns without individual differences (cf. Hannan 1970). In the case of language development, a process which is

characterized by extensive individual differences and a complex and dynamic interaction of a large number of factors, ergodicity is not likely to be met for most groups (Lowie and Verspoor 2019). This implies that L2 development over time can be reliably studied by dense observations in individual cases, but will be unreliable for groups of learners. Lowie (2017) therefore argued for a two-way distinction of research into L2 learning, product versus process. While product-oriented research focuses on accounting for differences between groups of learners at one or two moments in time, process-oriented research involves studying numerous observations of the dynamic development within individuals over time. Studies conducted in the CDST framework are typically process-oriented studies that closely follow individuals over time.

In recent years, many CDST-inspired studies have tracked intra-individual variability to examine language development. Chan *et al.* (2015) compared the development of sentence complexity in speaking versus writing in two beginner learners of English (identical twins). The results showed that even identical twins with similar personalities and interests who were exposed to similar input within the same environment showed meaningful differences in their patterns of language development. Dong (2016) investigated the developmental patterns of a Chinese EFL learner's use of listening strategies and listening performance and found that both measures showed non-linear developmental patterns and the beginnings of new phases were accompanied by great variability. Penris and Verspoor (2017) studied academic writing development of an advanced learner of English and found that almost all variables of complexity (syntactic and lexical) and accuracy developed with high degrees of variability over time. These studies testified that the process of language development is not stable over time or identical across subjects, and is, therefore, non-ergodic, which is in accordance with Molenaar's (2015) argument that these two kinds of variation cannot be combined in the same analysis as they essentially involve different dimensions of research.

In the current study, we will focus on both inter-individual changes and intra-individual variation of oral second language development, tracing the general changes at the group level as well as the development of individual learners. The present article may therefore deepen our understanding of how oral second language proficiency in a Chinese context develops over time.

Interactions between complex dynamic subsystems

From a CDST perspective, the main system is comprised of subsystems that are nested at different levels and continuously interact with each other. Growth takes place as the complex system (or subsystem) develops over time. Van Geert (1995) defined growth within complex systems as:

A process is called growth if it is concerned with the increase or decrease (i.e. negative increase) of one or more properties, and if that increase is the effect of a mechanism intrinsic to that process.
(p. 314)

The growth process is dependent on the resources of system: internal resources (e.g. cognitive factors) and external resources (e.g. social factors and the learning context). Skehan's (1998) limited attentional resource model proposed that performing in an L2 may require a learner's attention and force them to prioritize one dimension, like accuracy, over another, such as complexity. The resources are limited but interlinked and compensatory in a dynamic system (De Bot 2008). In addition, since all subsystems are interrelated, changes in one subsystem may affect other subsystems within the same system, depending on the relative robustness of the subsystems (De Bot *et al.* 2007). Relationships between subsystems may be competitive or supportive and may show interactions between different dimensions of language proficiency (Van Geert and Steenbeek 2005). If the relationship is supportive, one subsystem will contribute to the growth of another connected subsystem. If it is competitive, improvement in one subsystem will result in another subsystem's decline. Another possible relationship between developmental measures or subsystems is a precursor relationship, which means that for some skills or knowledge the presence of a precursor is conditional for the emergence of a dependent subsystem (Van Geert 1991). The support or competition between variables can range from very strong to very weak.

Complexity, fluency, and accuracy (CAF), the three principal traits of language proficiency or representation of L2 competence (Skehan and Foster 1999), have been widely accepted as the primary foci of studies on second language development (SLD). Norris and Ortega (2009) pointed out that "CAF consists of dynamically related indices which do not all advance hand-in-hand towards an ideally complex, accurate and fluent performance" (p. 573). This may be because language development is complex, non-linear and dynamic (Larsen-Freeman 2009). From the perspective of CDST, CAF comprises the three subsystems of the language system. Competitive or supportive relationships may be found by investigating the interaction between the three measures (Van Geert and Steenbeek 2005).

Some CDST-based studies have reviewed CAF performance and the interactions between the three subsystems. Verspoor *et al.* (2008) adopted various techniques to visualize a learner's development and variability in written production. The results showed that a language system, even for an advanced learner, is not likely to remain stable, and subsystems of language tend to interact with each other. For instance, Spoelman and Verspoor (2010) investigated complexity and accuracy of a Dutch learner's acquisition of written Finnish over three years and explored the relationship between the two measures. The results showed that complexity and accuracy seemed to compete with each other in the early stages, but this changed into a supportive relationship in a later stage. Spoelman and Verspoor pointed out that the nature of the dynamic interactions could be explained when examining the learner's proficiency level, which increased over time and developed particularly quickly in the early stage.

The aforementioned studies have broadened and deepened our understanding of language development in terms of CAF and the relationship between the components of language performance. As of yet, these studies have predominantly concentrated on written language development rather than oral production. Polat and Kim (2014) traced the oral language development of a single participant who did not receive formal instruction, and showed that the untutored learner's oral language production improved in complexity but not accuracy. The learner was more motivated by communicative needs than by grammatical goals. The possible interaction between complexity and accuracy was not explored in their study.

Although previous CDST-based studies have indeed enriched our understanding of the dynamic process of L2 development and the interactions between CAF measures, there is little knowledge of the process of learners' oral language development in the context of EFL, for example in China. English learning in China is considered a requirement for both national and individual development in a new era of globalization (Dai 2013). By 2017 the number of English learners in China approximated 100 million, making it the largest English learning community in the world. However, speaking is still the weakest language skill of most Chinese learners. Further research needs to be conducted to create a comprehensive picture of learners' oral language developmental patterns, which may help Chinese English teachers refine their instructional strategies for oral language teaching (Wu 1999). Moreover, Lowie (2017) called for both inter- and intra-individual studies to be conducted in the future so as to gain a better understanding of language development in terms of the factors that may have contributed to the product of learning as well as the process of development. On the one hand, we will need to deepen our understanding of the general development of all the participants from the inter-individual research. On the other hand, the intra-individual data may enable us to study the development of each individual and the interactions between the relevant subsystems over time.

Since CDST is regarded as an effective approach to studying how a system evolves over time (Van Gelder and Port 1995), and in the wake of the interesting results of the CDST-inspired studies discussed previously, the current study, which was conducted at the tertiary EFL level in China, aims to contribute to the existing body of research in two ways. Firstly, we will examine both the inter-individual and intra-individual variation of complexity and accuracy of oral language development over time by showing the differences between them. Secondly, we aim at investigating and understanding the relationship between complexity and accuracy of language proficiency using analytical tools of CDST.

METHODS

To fill the gap in the literature on project- and process-oriented approaches to L2 oral production, we designed the current study with 10 Chinese-speaking

college learners of L2 English. The methods of intra- and inter-individual analyses will be employed to study how complexity and accuracy develop over time and explore individual differences. The study visualizes the developmental patterns of oral language and provides a new perspective on the interactions between complexity and accuracy. Our research questions are the following:

RQ1: Does Chinese learners' oral language proficiency show improvements over the whole semester?

RQ1 is a product analysis, i.e. based on group average scores rather than on dense measurements over time. We hypothesized that measures of complexity and accuracy may show a general improvement based on the results of previous studies (Larsen-Freeman 2006; Vercellotti 2017). Due to different learning contexts, our findings may be different from the results reported by Polat and Kim (2014), who found that most development occurred in learner's complexity rather than accuracy. The subject in their study was an untutored immigrant, while ours are college students who have received formal English language instruction.

RQ2: How do complexity (syntactic and lexical) and accuracy develop at the individual level over time?

RQ2 is a process-oriented analysis. It can be hypothesized that the development patterns of complexity and accuracy will be non-linear with a high degree of variability over time and individual learners may show developmental paths that differ substantially from each other. Moreover, we would expect that the participants, who take tertiary-level English courses every day, will improve on their oral English production.

RQ3: What is the relationship between complexity and accuracy?

According to Skehan (1998), it can be hypothesized that due to limited attentional resources participants would take fewer risks under time pressure, and this tendency would be reflected in them choosing high-frequency words or simple syntactic structures with high accuracy at the early stage of this study. Participants can also be expected to pay attention to both dimensions (complexity and accuracy) as their oral proficiency develops over time. In other words, the interaction between complexity and accuracy is expected to shift from competitive to supportive over time, which would be in line with an increase of the coordination of the subsystems over time.

Participants

The participants in our study were 10 non-English major first-year students (five males and five females) from the same university in Shanghai. Most of these learners had been learning English as a foreign language in China for 10 years (from primary school to university), and their average score on the

National College Entrance Examination (commonly known as Gaokao) was 138 (maximum score = 150) (range = 19, SD = 6.16). None of them had ever visited an English-speaking country and they had seldom had the opportunity to speak English for communicative purposes in daily life. They had comprehensive English courses each week and oral English classes every two weeks, which were taught by Chinese English teachers. The pedagogical goal of comprehensive English courses is to improve learners' English proficiency in terms of speaking, reading and writing, by studying textbook units. In oral English class, students participate in different speaking activities. They were asked to give mini-presentations or group talks on given topics, and subsequently received teacher feedback, usually on both general oral performance and specific problems (grammatical errors).

Due to space limitations, we selected two participants to compare their developmental patterns in the current study, discussing their respective development of complexity and accuracy, and the relationship between them. For the selection of the two participants, following Baba and Nitta (2014), we used correlation tests between the week number and mean length of speech unit of each week for 10 participants. The two participants who showed the highest and the lowest correlation were selected and they were compared to examine whether the initial state influenced their oral language development or not. The highest correlation demonstrates that this participant's syntactic complexity seems to have changed more than that of the other students, and vice versa. The highest correlation was A ($r = 0.804$, $p < 0.01$), and the lowest was B ($r = -0.402$), which shows the negative correlation between week and complexity. Two participants had been learning English for six years, and A and B's National College Entrance Examination scores were 138 and 140 respectively. They came from different areas of China; A was from Nantong, a prefecture-level city in the south of Jiangsu Province, while B was from Zhuzhou, a prefecture-level city of Hunan Province. According to the Ranking List of Top 100 Chinese cities by GDP 2017, Nantong ranked 24th, whereas Zhuzhou ranked 87th, meaning that A's hometown economy develops faster than B's. Economic disparities among regions may cause differences in educational resources. Since Southern Jiangsu is the most developed economic region in China, we expect that A may perform better than B because of better educational resources, and their oral language proficiency may develop differently.

Procedures

To capture the developmental patterns of oral language use, a series of observations are needed. Twelve different topics were taken from the International English Language Testing System (IELTS) Speaking Test and were categorized into four different themes, which were social issues, locations or places, life experience and figure images (see Supplementary material Appendix A). The longitudinal study started from the fourth week of classes and lasted for the whole semester (see Figure 1). The weekly recordings were conducted in a



Figure 1: Recording dates

monologue style, meaning each participant was asked to speak on a given topic for two or three minutes and was given one minute for preparation. While preparing for the topics, the participants were not allowed to take notes or use any tools such as mobile phones or dictionaries. All speeches were recorded in a media lab on Apple Mac computers with eXtra Voice Recorder software. Only the researcher and participant were present in the media lab during recording, in order to cause as little anxiety as possible.

After recording, each participant was interviewed to gain background information regarding their English education. This included basic information such as years of English instruction, English curriculum arrangements and teaching methods from primary to senior high school, but also more social factors such as the amount of exposure to English in natural settings and learning motivation. The subjects in the current study, who had particularly strong desires to pass the Association of Chartered Certified Accountants (ACCA) exams, spent extra time learning English in natural settings. Recordings and interviews were held weekly to capture any changes in participants' oral performance.

Variables

Most previous studies have chosen T-units to measure syntactic complexity in writing (Larsen-Freeman 2006; Bulte & Housen 2014; Chan *et al.* 2015), while some have chosen analysis of speech units (AS-units) in oral language (Polat and Kim 2014). Foster *et al.* (2000) defined an AS-unit as 'a single speaker's utterance consisting of an independent clause, or sub-clause unit, together with any subordinate clauses associated with either' (p.365). Considering the focus of the current study is oral production, syntactic complexity was analysed by measuring the number of words per AS-unit.

The two variables of complexity that were studied were syntactic complexity and lexical diversity. Syntactic complexity was measured as mean length of AS-unit (MLA) in words, following Polat and Kim (2014). An increase in MLA demonstrates that the L2 learners are using increasingly more detailed and grammatically complex descriptions, including adjectives, adverbs or subordinate clauses (Norris and Ortega 2009; Pallotti 2009). In addition, we recorded the learners' lexical diversity, measured as D (see McKee *et al.* 2000) and calculated by the VOCB subprogram within Computerized Language Analysis (CLAN) (MacWhinney 2000). As a useful lexical measure for L2 data (Treffers-Daller 2009), D is especially considered a valid measure for

oral narratives of English L2 speakers (Lu 2012). Since the current study investigated oral language development over time with texts of different lengths, it seemed appropriate to assume that D is more suitable.

Researchers have used a number of different measures of accuracy, e.g. percentage of error-free clauses or number of errors per 100 words (Ellis and Barkhuizen 2005). However, doubts have arisen about the validity and reliability of these general measures. Pallotti (2009) pointed out that general accuracy measures cannot offer valid analyses of learners' L2 development due to the different kinds of errors produced by learners at different levels. Therefore, more specific measures are needed. In addition, learners have encountered difficulties in mastering the past tense, even at intermediate or advanced levels (Ellis and Larsen-Freeman 2006), and Mandarin Chinese does not mark past tense with morphological changes (Cai 2007). The use of the past tense thus remains especially difficult for Chinese EFL learners (Cai 2003; Yang and Huang 2004; Yang and Lyster 2010). As a consequence, two variables of accuracy were calculated in the current study: one general measure (number of error-free AS-units, EFA) and one specific measure (number of error-free past tenses, PTA). It was assumed that the general measure would present the holistic view of the 10 L2 learners' overall ability to use the L2 grammar, while the specific measure would evaluate efficiency of learning and using past tenses.

Coding the data

The oral data were audio-recorded and subsequently transcribed and coded for complexity and accuracy in CHAT format to be compatible with CHILDES program software (MacWhinney 2000). The data were then analysed for syntactic complexity (mean length of AS-unit), lexical diversity (D), and accuracy (general measure: number of error-free AS-units, specific measure: number of error-free past tenses). For the analysis of further developmental variables, a research assistant checked all the coding again and resolved any deviations in the data with the researcher.

Design and analysis

The current study traced the developmental patterns of oral language development over time. A moving min-max graph (Van Geert and Van Dijk 2002) was adopted to trace changes and degrees of variability in the development of oral L2 production. The moving min-max graph is a technique that visualizes the dynamic developmental process and highlights the variability. The interactions between subsystems were visualized by moving correlations. The operation principle of a moving correlation is similar to the min-max graph (Verspoor *et al.* 2011). In order to make sure that the degree of variability and interaction between variables is not distorted by the incline of the slopes (Verspoor *et al.* 2011), it is necessary to detrend all raw data. In addition, to

explore the changes in the data of the two participants on a common scale, raw scores of different numerical ranges were normalized to 0–1 values, a common method in CDST research (Polat and Kim 2014; Chan *et al.* 2015). Strong fluctuations in variability were tested for significance through resampling techniques and Monte Carlo simulations. By randomly reshuffling the data 10,000 times, a Monte Carlo analysis calculates how often a similar peak occurs in the dataset when shuffled. If a peak occurred fewer than 250 times, it was deemed significant.

RESULTS

The present study aimed at examining both inter-individual group data and intra-individual variability (cf. Molenaar 2015) to gain a deeper insight into the dynamic development of Chinese English learners' oral language performance in terms of complexity and accuracy. Therefore, measures of complexity (MLA and D) and accuracy (EFA and PTA) were studied at both the general group level and the intra-individual level.

The first question we want to answer is 'Does Chinese learners' oral language proficiency show improvements over the whole semester?'

The difference between the first measurement and the last measurement for complexity as operationalized by MLA for the entire group is shown in Figure 2. A paired-samples *t* test confirmed that on average, MLA was

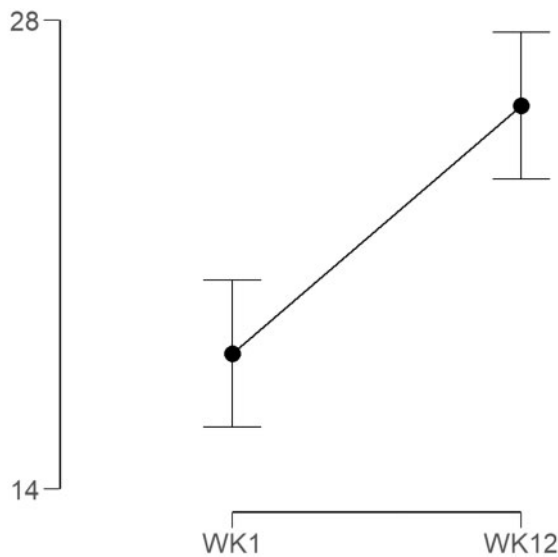


Figure 2: The group means for the MLA measure at the first (WK1) and the last (WK12) measurement

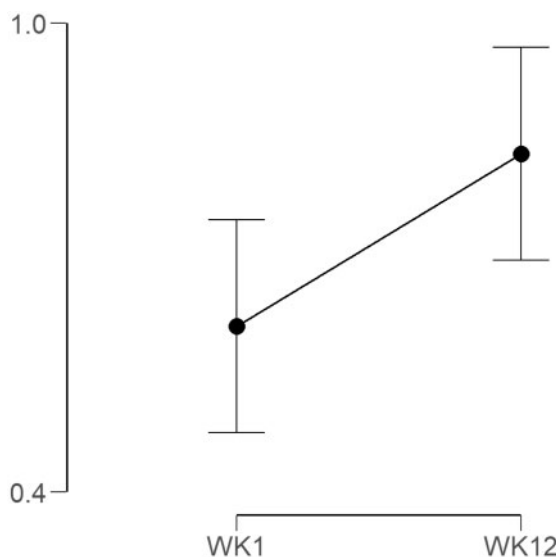


Figure 3: The group means for the EFA accuracy score at the first (WK1) and the last (WK12) measurement

significantly higher in week 12 ($M = 25.4$, $SD = 5.9$) than in week 1 ($M = 18.0$, $SD = 2.3$). $t(9) = 5.4$; $p < 0.01$, $d = 1.7$. The second complexity measure, D , was non-normally distributed in Week 12 with skewness of 1.2 ($SE = 0.7$) kurtosis of 2.1 ($SE = 1.3$). A Wilcoxon signed-rank test showed that the scores for D were significantly higher in week 12 than in week 1 ($W = 55$, $p < 0.01$, $r = 1.0$).

Accuracy also improved over time for the entire group. The difference between the first and the last measurement of general accuracy (EFA) for the entire group is shown in Figure 3.

A paired samples t -test showed a significant increase for the learners' EFA in week 12 ($M = 0.61$, $SD = 0.20$) compared to week 1 ($M = 0.83$; $SD = 0.21$), $t(9) = 0.029$, $d = 0.82$. Likewise, specific accuracy (PTA) showed higher scores the last observation ($M = 0.85$, $SD = 0.16$) than in week the first ($M = 0.4$, $SD = 0.5$), $t(9) = 2.6$, $p = 0.026$, $d = 0.84$. The data shows that oral language proficiency for these learners increased over time, though the effect was larger for complexity than for accuracy.

Our second research question was: 'How do complexity and accuracy develop at the individual level over time?'

As mentioned in the background section, analysing person-specific, non-ergodic data should be based on intra-individual variation rather than average group scores (Molenaar and Campbell 2009). Therefore, we examined the longitudinal data by adopting the process-oriented moving min-max graph technique.

COMPLEXITY

Syntactic complexity

The trajectory of both participants' syntactic complexity and the min-max values are illustrated in Figure 4. It is clear that the learners' syntactic complexity developed in a non-linear way, as the bandwidth of scores did not remain stable across the trajectory. In A's min-max graph, three periods (week 1–3, week 4–8 and week 10–12) can be distinguished throughout the study, with the bandwidth between the min and max line growing increasingly wider, indicating considerable fluctuations over time. Moreover, high degrees of variability were followed by an increase in syntactic complexity. A possible jump was spotted between week 11 and week 12. A Monte Carlo simulation (10,000 iterations) revealed that the developmental jump can be

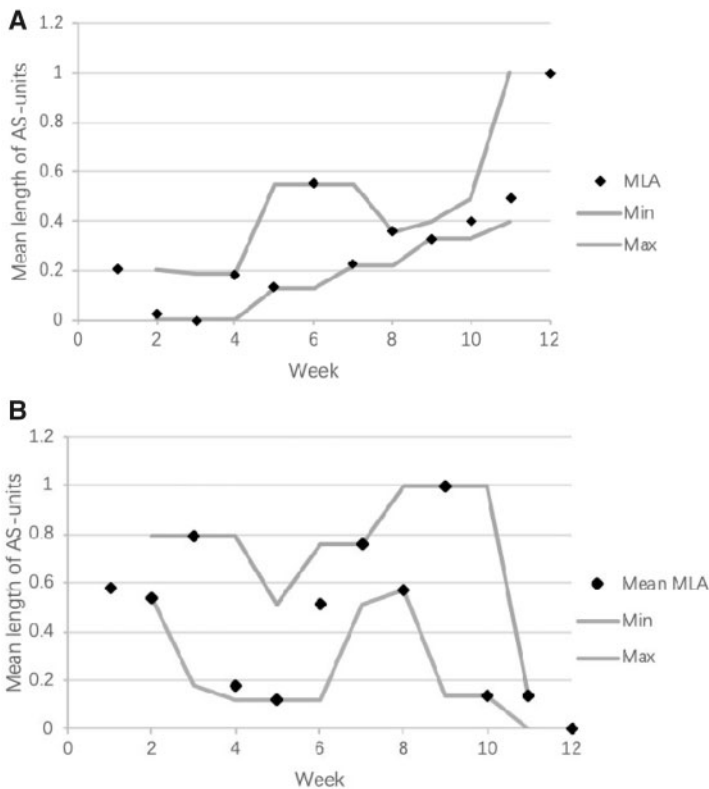


Figure 4: Moving min-max graph of syntactic complexity for both (A) and (B)

considered significant ($p < 0.05$). In contrast, for B, constant fluctuations (week 3–6, week 8–10) and high degrees of variability were followed by a decline in syntactic complexity (from 0.6–0). Compared to A, B's developmental pattern of syntactic complexity showed much higher degrees of variability over the whole process, with several peaks and dips. A Monte Carlo simulation (10,000 iterations) revealed that the resulting p value was 0.49, which means that these peaks and dips were likely the result of coincidental fluctuations. In addition, another Monte Carlo simulation was run to see whether A's syntactic complexity was more variable than B's. The p value was 0.048 with 10,000 iterations, meaning A was indeed more variable than B.

Lexical complexity

When looking at both A's and B's moving min–max graphs of lexical diversity (Figure 5), higher degrees of fluctuations in lexical complexity can be seen than in the syntactic complexity graphs, especially during the last few weeks (week 7–12). Resampling and Monte Carlo simulations (10,000 iterations) showed that there were no significant peaks and dips in development of lexical diversity of either A ($p = 0.30$) or B ($p = 0.41$).

ACCURACY

General accuracy (EFA)

The trajectory of both participants' EFA and the min–max values are illustrated in Figure 6. It showed that the bandwidth between the min and max lines was rather large over the whole process, especially during the beginning of the study, representing high degrees of variability. A Monte Carlo analysis indicated that the peak in A's EFA development (week 6–7) was not significant ($p = 0.39$ with 10,000 simulations). Moreover, the difference between the two participants in the degree of variability was not significant, either ($p = 0.33$).

Specific accuracy (PTA)

In Figure 7, the moving min–max graphs displaying the participants' PTA have been plotted. A's PTA shows three stages (week 1–week 3, week 4–week 8, week 9–week 12), with a period of high variability as a transition phase between the first and last stage, and includes a sharp increase in the third stage. Surprisingly, contrary to our expectation, only two values (0 and 1) occurred in B's PTA moving min–max graph. We rechecked each transcribed data of B and found that no form of past tense was used in week 1, 6, and 10, which explains why the value of PTA was zero during these weeks. In addition, the fact that PTA was 1 in the other weeks seems to indicate that B has mastered the use of the past tense.

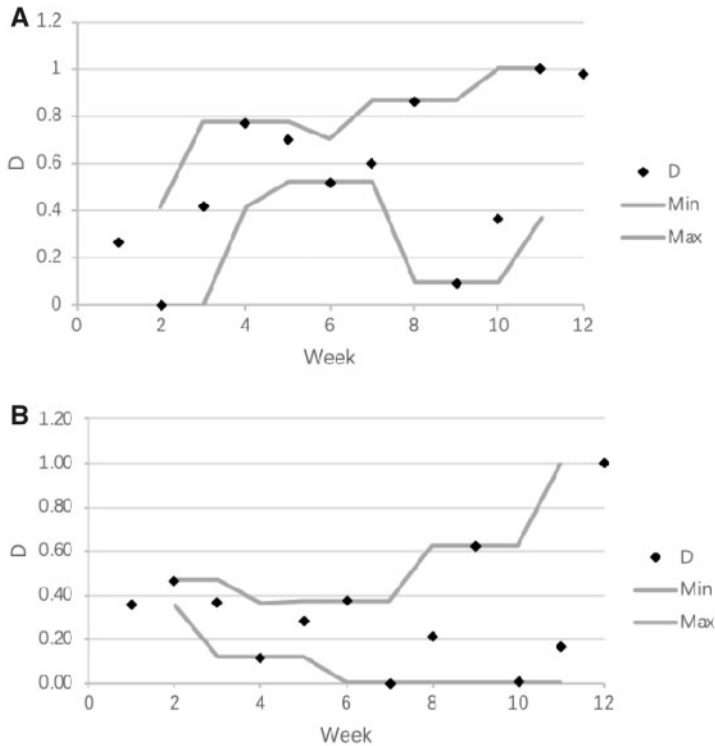


Figure 5: Moving min-max graph of lexical diversity for both (A) and (B)

Interactions between complexity and accuracy

Our final research question concerns the relationship between complexity and accuracy. Here, we limited ourselves to examining the general relation between syntactic complexity (MLA), lexical diversity (D), and accuracy (EFA), as Skehan and Foster (1999) pointed out that general measures are more effective for analysing the relationship.

We will first consider the relationship between complexity and accuracy at the group level. A Pearson correlation revealed that there was a strong positive relationship between syntactic complexity and accuracy ($r = 0.58$, $p < 0.05$). The relationship between lexical complexity and accuracy, on the other hand, was very weak and insignificant ($r = 0.04$, $p = 0.89$).

The interaction between syntactic complexity and accuracy was plotted in Figure 8. The plot indicates that the correlation shifts between negative and positive coefficient values over time. In Figure 8, a negative correlation between syntactic complexity and accuracy can be seen in the first half of this study of participant B, while a positive correlation was found of A. After that,

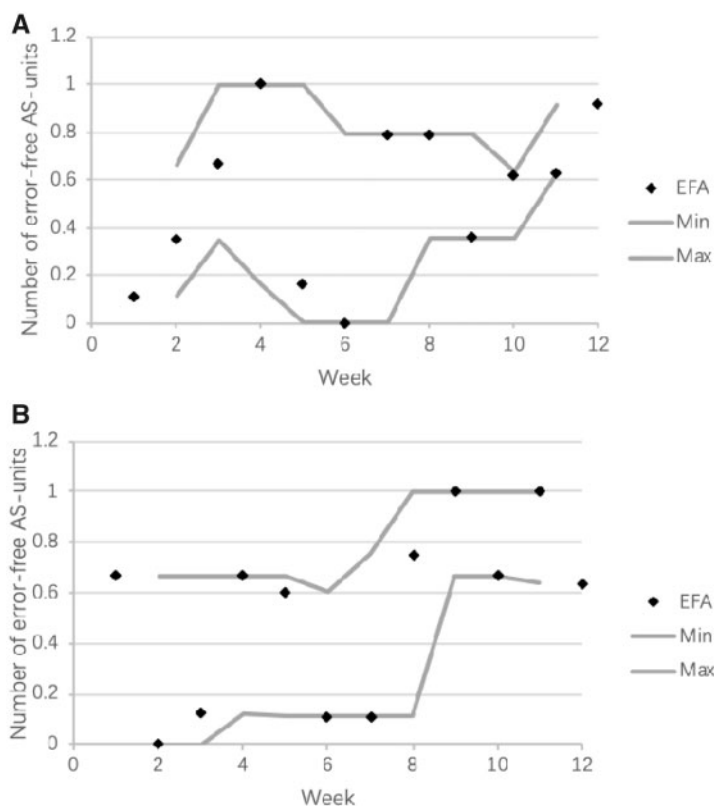


Figure 6: Moving min-max graph of EFA for both (A) and (B)

the interaction between syntactic complexity and accuracy changed from negative to positive for both A and B, which means syntactic complexity and accuracy improved simultaneously as the study went by.

Figure 9 shows that correlations between lexical diversity and accuracy were alternately negative and positive over time. A and B displayed different interaction patterns. A's correlation showed continuous fluctuations, which turned into a positive correlation at the end of this study. In contrast, B's correlation was negative during the first half of this study, but turned into a positive one in the second half and remained positive until the end of this study, if not for the final measurement.

DISCUSSION

The purpose of the current study was to evaluate the developmental patterns of the L2 oral production of Chinese English learners over a period of four months. Both product- and process-oriented methods were adopted.

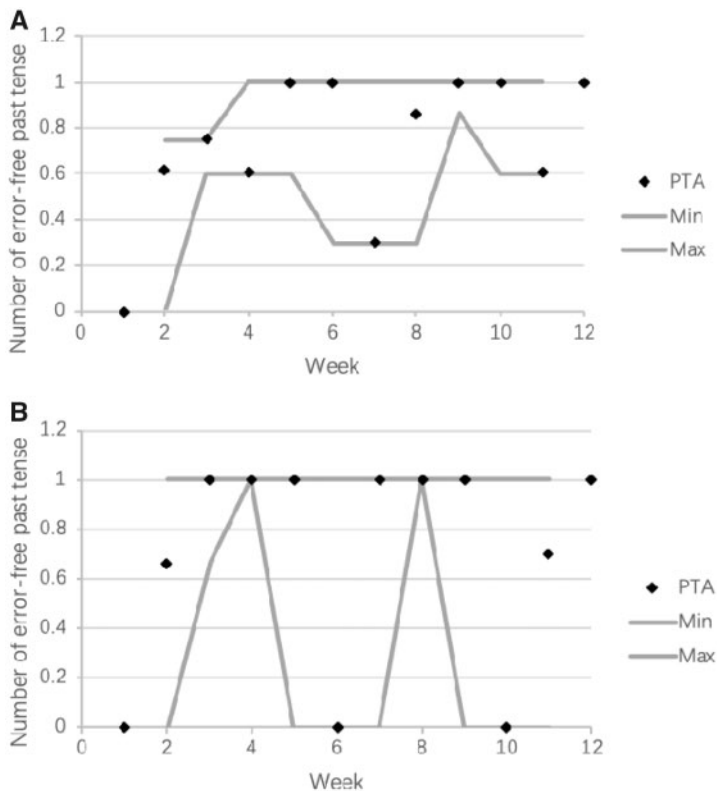


Figure 7: Moving min-max graph of PTA for both (A) and (B)

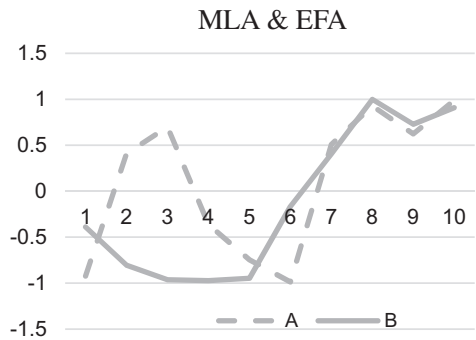


Figure 8: Moving correlation between syntactic complexity and accuracy for (A) and (B)

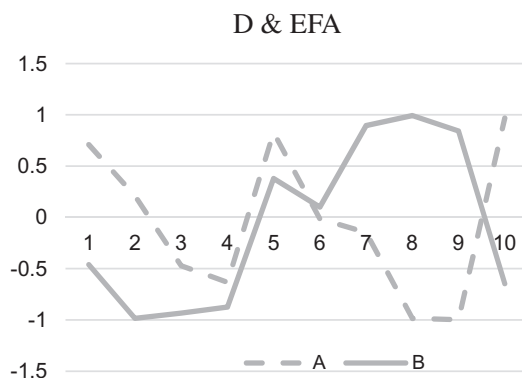


Figure 9: Moving correlation between lexical complexity and accuracy for (A) and (B)

Complexity and accuracy showed general improvement at the group level, which contrasts with results found by Polat and Kim (2014), who traced development in speaking proficiency of an untutored learner. They found that the participant made progress on complexity, but not on accuracy. The difference could be due to the different learning contexts of the studies. The subject, in their study, did not receive formal instruction or feedback (from teachers or tests) and improved his oral proficiency through successful social communication. In other words, the focus of his oral language performance was more on communicating meaning than on linguistic form. However, the subjects in this study were college students who received formal instruction in a classroom setting in which a focus on grammatical knowledge may have supported their development through tertiary-level classes over the course of an academic semester. Chinese teachers tend to give much more feedback on grammatical errors than other aspects of both written and spoken second language use, which may lead to over-awareness of language accuracy. This may explain the difference between Polat and Kim (2014) and this study.

When the process-oriented methods had been applied to individual learners' data, numerous differences were found between the group and the individuals within this group. At the group level, both complexity and accuracy showed improvement. At the individual level, however, high degrees of variability could be seen in the non-linear, dynamic developmental patterns of oral production. The variable pattern of development is consistent with the assumption that variability is an inherent quality of a complex dynamic system (Larsen-Freeman and Cameron 2008). Moreover, each participant showed a unique pattern of development, suggesting that their oral production development may be influenced by different 'internal or external factors', meaning it would probably be meaningful to follow this with one example of an internal factor and one example of an external factor.

The two participants were selected based on their overall degree of change, with A showing a relatively high degree of improvement and B showing the lowest degree of change over time. As we mentioned above, A and B come from different provinces which belong to different economic regions of China. Economic disparities among regions may cause differences in educational resources, associated with the observation that learners in more developed areas have more advanced English proficiency than their peers elsewhere (Yan and Horwitz 2008). Moreover, the interviews revealed that A had been learning English since primary school and received formal instruction in four language skills (listening, speaking, reading, and writing) in junior and senior school, and his speaking courses were taught by native speakers. In addition, A told us that his senior school English teacher had taught in English and had required students to answer in English as well. This allowed them to gradually develop their language skills, eventually leading to a higher level of English. B, in contrast, had been learning English since junior school without any listening or speaking courses. Moreover, all the English courses he had were taught in Chinese, which means that his exposure to English was rather limited compared to A.

This may imply that prior knowledge plays an important role in language development at later stage, which corroborates De Bot's (2008) statement that initial state is an influential characteristic of a complex dynamic system. Lowie *et al.* (2010) pointed out that 'language development is dependent on the initial condition and shaped by a wide range of interacted factors in a dynamic way' (p. 135). This may also justify why A's syntactic complexity had a significant change and A's syntactic complexity was more variable than B's.

The results differ from other CDST-inspired studies, like Spoelman and Verspoor (2010) and Baba and Nitta (2014), who found several significant peaks or dips in their subjects' written production. This may be due to the different modality of language production we chose. Written and oral language both require learners to actively use all aspects of L2, including vocabulary, verb tenses and sentence construction. However, writing tasks may allow learners to produce high-quality texts with ample preparation (Chan *et al.* 2015). In other words, the oral production of L2 learners who are under processing and communication pressure may not be on a par with their written one with the influence of their native language. Therefore, the variability in the development of oral production may differ from variability in written production. In addition, the 12 observations from the present study were likely insufficient to track development of oral production; generally, an extended time period (often years) is needed to develop L2 proficiency (Fogal 2017). Although the variability observed in the current study may not be systematic or significant, we still expect oral production in general to be more variable than written production, due to the different online processing capacity that is required for oral production.

In addition, two participants' specific accuracy scores (past tense accuracy) reached high values in the present study, indicating they may have mastered this grammatical structure. When we re-examined all of A and B's oral data (24 data points), however, we found that most of the used L2 past tenses were high-frequency words, rather than low-frequency words. It is likely that the participants preferred to maintain high accuracy at the cost of using low-frequency words. This is exemplified in Excerpt 1 and 2 below.

Excerpt 1:

*A: and at age about 12 his father and mother realize[*pt] that this guy has [*pt] talent to be a top basketball player, but they never thought he could be a national team player.

*A: And so then he went to the sports school and abandoned[/] abandoned academic[/] academic[*p] works.

*A: <In a>[/]. . . At age of 18, he become[*pt] the youngest player of Guangdong Province[*p] team of basketball.

Excerpt 2:

*B: And I was so excited and but a little nervous because I was just before the audience and finally I conquered the <fear>[/] fears and <and do>[/] <and did>[/] and did at the host said.

*B: It was adapted by a novel and er <I saw it >[/]I saw it in a cinema <near my house>[/] near my home.

These excerpts show that the majority of past tense verbs used by A and B are high-frequency words, which could account for the high accuracy scores in this study. The students' preference for high-frequency words over low-frequency words may partly be related to the teaching methods commonly adopted by Chinese teachers in oral English courses. In China, teachers tend to give much more feedback on grammatical mistakes than other aspects of their students' language use (Zhuang 2012), leading to over-awareness of language accuracy compared to language complexity. In addition, the participants' first language may also influence their preference, as there is no past tense form in Chinese, making it difficult for learners to learn past tense verbs; occasionally, the participants confused past tense with past participle.

Regarding the interaction between complexity and accuracy, moving correlation graphs indicated changes in developmental patterns over time, something which could not have been revealed by correlations of measurements at one point in time. In A and B's moving correlation graphs, correlations between complexity and accuracy were alternately negative and positive, demonstrating that the relationship between the measures alternated between competitive and supportive at different points in time. For instance, the moving correlations between syntactic complexity and accuracy for both participants were mostly negative in the early stages, indicating that the two measures were in a competitive relationship. In other words, learners'

syntactic complexity improved at the expense of accuracy at the beginning of the semester. Later, however, the correlation became positive, meaning that the measures entered a more supportive relationship for the remainder of the semester, and the increase in e.g. syntactic complexity also resulted in increase in accuracy. This is in line with Vercellotti's (2017) study. However, the interactions between lexical complexity and accuracy for both participant A and B revealed several differences. Skehan (2009) proposed that 'For non-native speakers, lexical diversity tends to be positively related to accuracy: the less recycling of vocabulary there is, the higher the accuracy that is achieved' (p.116). Only A's result supported Skehan's statement. Considering the fact that A and B come from different regions in China, their oral language proficiency may vary depending on formal instructions and amount of exposure to English. Therefore, we assume that the different interactions are mainly due to the individual differences and the oral proficiency level of these two participants, which developed over time.

CONCLUSION

The current study is an answer to Wu's (1999), Lowie's (2017) and Polat and Kim's (2014) call for further research on developmental patterns of oral language. The general developmental trends of both complexity and accuracy showed improvements, while the developmental traces of individual learners were rather non-linear and dynamic with high degrees of variability. Moreover, a complex interplay existed between complexity and accuracy, and the relationship between them changed from a clearly competitive relation during the early stage to a supportive relation at later stage. This shows that complexity and accuracy influence each other greatly, which is in line with CDST claiming that all factors within a dynamic system are connected. The results mentioned above also appear to confirm the possibility and effectiveness of CDST-inspired approaches towards the study of L2 development.

Far more longitudinal studies are needed in order to broaden the scope of CDST-based studies on L2 oral production, and it is important to include more learners from different sociocultural contexts. Meanwhile, fluency, a third component of oral language proficiency next to complexity and accuracy, needs to be addressed from a CDST perspective as well. In addition, the length of the data collection period should also be extended, which may allow us to trace oral language developmental patterns more comprehensively. With the study presented here, we hope to have contributed to the body of research showing the relevance of unravelling the dynamic process of second language development.

SUPPLEMENTARY DATA

Supplementary material is available at *Applied Linguistics* online.

REFERENCES

- Baba, K. and R. Nitta.** 2014. 'Phase transitions in development of writing fluency from a complex dynamic systems perspective,' *Language Learning* 64/1: 1–35.
- Bulté, B. and A. Housen.** 2014. 'Conceptualizing and measuring short-term changes in L2 writing complexity,' *Journal of Second Language Writing* 26: 42–65.
- Cai, J.** 2003. 'The relationship between narrative structure and simple past variation in English interlanguage—A test of discourse hypothesis,' *Modern Foreign Language* 26/1: 59–68.
- Cai, J.** 2007. *The Effects of Multiple Linguistic Factors on the Simple past Use in English Interlanguage*. Beijing: Foreign Language Teaching and Research Press.
- Chan, H., M. Verspoor, and L. Vahtrick.** 2015. 'Dynamic development in speaking versus writing in identical twins,' *Language Learning* 65/2: 298–325.
- Dai, W.** 2013. *The Orientation, Layout and Development of China Foreign Language Major Education*. *Contemporary Foreign Languages Studies*, 7.
- De Bot, K.** 2008. 'Introduction: Second language development as a dynamic process,' *The Modern Language Journal* 92/2: 166–78.
- De Bot, K., W. Lowie and M. Verspoor** 2007. 'A dynamic systems theory approach to second language acquisition,' *Bilingualism: Language and Cognition* 10/1: 7–21.
- Dong, J.** 2016. 'A dynamic systems theory approach to development of listening strategy use and listening performance,' *System* 63: 149–65.
- Ellis, R. and G. Barkhuizen.** 2005. *Analyzing Learner Language*. Oxford University Press.
- Ellis, N. C. and D. Larsen-Freeman.** 2006. 'Language emergence: Implications for applied linguistics—Introduction to the special issue' *Applied linguistics* 27/4: 558–89.
- Fogal, G. G.** 2017. 'Tracking microgenetic changes in authorial voice development from a complexity theory perspective,' *Applied Linguistics*; doi: 101093/applin/amx031.
- Foster, P., A. Tonkyn and G. Wigglesworth.** 2000. 'Measuring spoken language: A unit for all reasons,' *Applied Linguistics* 21/3: 354–75.
- Hannan, E.** 1970. *Multiple Time Series*. Wiley.
- Larsen-Freeman, D.** 1997. 'Chaos/complexity science and second language acquisition,' *Applied Linguistics* 18/2: 141–65.
- 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.
- Larsen-Freeman, D.** 2009. 'Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition,' *Applied Linguistics* 30/4: 579–89.
- Larsen-Freeman, D. and L. Cameron.** 2008. 'Research methodology on language development from a complex systems perspective,' *The Modern Language Journal* 92/2: 200–13.
- Lowie, W.** 2017. 'Lost in state space? Methodological considerations in complex dynamic theory approaches to second language development research,' in L. Ortega and Z. Han (eds): *Complexity Theory and Language Development: In Celebration of Diane Larsen-Freeman*. John Benjamins Publishing Company, pp. 123–41.
- Lowie, W. M. and M. H. Verspoor.** 2019. 'Individual differences and the ergodicity problem,' *Language Learning* 69 (S1): 184–206.
- Lowie, W., M. Verspoor and K. De Bot.** 2010. 'A dynamic view of second language development across the lifespan,' in K. de Bot and R. W. Schrauf (eds): *Language Development over the Lifespan*. Routledge, pp. 125–146.
- Lu, X.** 2012. 'The relationship of lexical richness to the quality of ESL learners' oral narratives,' *The Modern Language Journal* 96/2: 190–208.
- MacWhinney, B.** 2000. *The CHILDES Project: Tools for Analyzing Talk*, 3rd edn. Lawrence Erlbaum Associates.
- McKee, G., D. Malvern and B. Richards.** 2000. 'Measuring vocabulary diversity using dedicated software,' *Literary and Linguistic Computing* 15/3: 323–38.
- Molenaar, P. C.** 2015. 'On the relation between person-oriented and subject-specific approaches,' *Journal of Person-Oriented Research* 1/1–2: 34–41.
- Molenaar, P. C. and C. G. Campbell.** 2009. 'The new person-specific paradigm in psychology,' *Current Directions in Psychological Science* 18/2: 112–17.
- Norris, J. M. and L. Ortega.** 2009. 'Towards an organic approach to investigating CAF in

- instructed SLA: The case of complexity,' *Applied Linguistics* 30/4: 555–78.
- Pallotti, G.** 2009. 'CAF: Defining, refining and differentiating constructs,' *Applied Linguistics* 30/4: 590–601.
- Penris, W. and M. Verspoor.** 2017. 'Academic writing development: A complex, dynamic process' in S. E. Pfenninger and J. Navracscs (eds): *Future Research Directions for Applied Linguistics*. Multilingual Matters, pp. 215–42.
- Polat, B. and Y. Kim.** 2014. 'Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development,' *Applied Linguistics* 35/2: 184–207.
- Skehan, P.** 1998. *A Cognitive Approach to Language Learning*. Oxford University Press.
- Skehan, P. and P. Foster.** 1999. 'The influence of task structure and processing conditions on narrative retellings,' *Language Learning* 49/1: 93–120.
- Skehan, P.** 2009. 'Lexical performance by native and non-native speakers on language-learning tasks,' in B. Richards *et al.* (eds): *Vocabulary Studies in First and Second Language Acquisition*. Palgrave Macmillan, pp. 107–24.
- Spelman, M. and M. Verspoor.** 2010. 'Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish,' *Applied Linguistics* 31/4: 532–53.
- Tarko, V.** (2005, December 29). What is ergodicity? Individual behavior and ensembles. Softpedia News. Available at: <https://news.softpedia.com/news/What-is-ergodicity-15686.shtml>. Accessed 12 March 2019.
- Thelen, E. and L. Smith.** 1994. *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press.
- Treffers-Daller, J.** 2009. *Code-Switching and Transfer: An Exploration of Similarities and Differences*. Cambridge University Press.
- Van Geert, P.** 1991. 'A dynamic systems model of cognitive and language growth,' *Psychological Review* 98/1: 3.
- Van Geert, P.** 1995. 'Growth dynamics in development,' in R. F. Port and T. van Gelder (eds): *Mind as Motion: Explorations in the Dynamics of Cognition*. The MIT Press, pp. 313–37.
- Van Geert, P. and H. Steenbeek.** 2005. 'Explaining after by before: Basic aspects of a dynamic systems approach to the study of development,' *Developmental Review* 25/3–4: 408–42.
- Van Geert, P. and M. Van Dijk.** 2002. 'Focus on variability: New tools to study intra-individual variability in developmental data,' *Infant Behavior and Development* 25/4: 340–74.
- Van Gelder, T. and R. F. Port.** (eds) 1995. *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press.
- Vercellotti, M. L.** 2017. 'The development of complexity, accuracy, and fluency in second language performance: A longitudinal study,' *Applied Linguistics*. 38: 90–111.
- Verspoor, M., K. De Bot and W. Lowie** (eds) 2011. *A Dynamic Approach to Second Language Development: Methods and Techniques*. John Benjamins Publishing Company.
- Verspoor, M., W. Lowie and M. Van Dijk.** 2008. 'Variability in second language development from a dynamic systems perspective,' *The Modern Language Journal* 92/2: 214–31.
- Wu, X.** 1999. 'Non-linguistic development of oral ability in the FL classroom setting,' *Modern Foreign Languages* 1: 1–33.
- Yan, J. X. and E. K. Horwitz.** 2008. 'Learners' perceptions of how anxiety interacts with personal and instructional factors to influence their achievement in English: A qualitative analysis of EFL learners in China,' *Language Learning* 58/1: 151–83.
- Yang, S. and Y. Y. Huang.** 2004. 'The impact of the absence of grammatical tense in L1 on the acquisition of the tense-aspect system in L2,' *IRAL* 42/1: 49–70.
- Yang, Y. and R. Lyster.** 2010. 'Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms,' *Studies in Second Language Acquisition* 32/2: 235–63.
- Zhuang, X.** 2012. 'The study of corrective feedback mechanism in classroom setting,' *Foreign Language Education* 33/2: 48–52.