The Microgenetic Changes in EFL Learners' Vocabulary Development
A Learner-Corpus-Based Study

Yurong Zheng[(✉)]

Harbin Engineering University, Harbin, China
`yurong@hrbeu.edu.cn`

Abstract. From the Complex Dynamic Systems Theory (CDST), this longitudinal study explores the Chinese EFL learners' vocabulary development. The 10 subjects are English learners at the tertiary level in a North-eastern university in China. The observation last 6 semesters, 3 school years, with each of the subjects contributing 6 timed argumentation writings. Microgenetic analysis is conducted to reveal the minute and complex changing trajectories and interactional patterns of the learners' lexical richness (sophistication and variation). The findings show that the development of lexical sophistication and variation is characterized by non-linearity and interconnection. The relationships between lexical sophistication and variation are also dynamic and individually different among the learners, whatever their language proficiency is. For the high-achievers, the relationship shows more supportive-competitive-supportive progression. However, it emerges to be more competition-oriented among the low-achievers. Further attributive analyses show that learners' developmental stage, content of writing tasks all cast impact of the developmental trajectories and interaction patterns.

Keywords: microgenetic change · Complex Dynamic Systems Theory (CDST) · Vocabulary Development · EFL learners

1 Introduction

With the introduction of complex dynamic systems theory (CDST) into the field of second language acquisition at the end of last century [1], this new perspective has brought about wide and fruitful applications in different sub-areas in SLA, second language writing being one of them [2]. Though there exist minor differences between the different strands under the umbrella term of CDST, at least four consensuses are shared: dynamicity, interconnectedness, self-organization and non-linearity [3]. In second language acquisition, that's probably the key reason for CDST becoming a suitable theory for most of the developmental studies, the focuses of which being on the dynamic changes over time [3] – [8]. Besides the dynamicity, another focus of CDST-based developmental studies is on the interconnectedness of the subcomponents of the dynamic systems [2, 9]. Both the dynamicity and the interconnectedness may explain why the outcomes of

L2 learners are seemingly nonlinear, and even chaotic [10, 11]. Variability has now been proved as one of features of second language development (e.g. [11] – [13]).

Among these developmental studies, longitudinal observation has become a norm. One of the most-cited longitudinal study is Verspoor et al.'s study [11], which observed the linguistic development of an advanced learner of English in her 18 compositions. In other study, Spoelman and Verspoor [14] conducted a three-year longitudinal study of a Dutch student learning Finnish, focusing on how the accuracy and complexity measures interacted. In their book-length publications introducing the research methods that can be used in CDST-perspective studies, Larsen-Freeman and Cameron [15] proposed microdevelopment as one of the ways.

Learner's variability has been one of the major interests in the microgenetic approach. In psychological areas, especially in areas concerning child development, especially the "microgenesis" of development, that is, the moment-to-moment change with a design of dense repeated measures [16]. It is right because of the emphasis of understanding complex phenomenon through the observation of interacting variables that incorporates microgenetic approach with CDST theories [17]. Both the two directions (CDST and microgenetic approach) cast their attention on the dynamicity of the changing individuals, which used to be regarded as "noises" (e.g. [10]). An exemplar study of this line of thought is Fogal's [18] microgenetic analysis of authorial voices among Japanese EFL learners' writing. He found that during the three-week writing course, the students' authorial voices show certain patterns of dynamicity and also subjective to different interactive factors, like training tasks and learners' awareness.

Thus, the objective of the present study is to depict some of the microlevel dynamics happening with the learners' linguistic development over the observation period, that is, 6 semesters among 10 participants with different linguistic proficiencies. In this process, the focus is first on describing the dynamics of change in learners' lexical development at the microlevel (the variability), and second, on individual learner differences in their developmental trajectories.

2   Development from the Perspective of CDST

According to Lowie and Verspoor [12], development is process-focused, rather than product-oriented. Developmental process involves variations at different time points, and the measurements at one or two points (e.g. the pretest plus the posttest) cannot create a complete picture. This incompleteness is coupled with the measurement of acquisition by accuracy in daily performance. Thus, "the observed data are not in fact orders of acquisition, but rather accuracy scores at one moment in time" [12 p. 66].

Furthermore, development is individual-oriented, not group-based. This inconsistency between individual and group performances has been found in empirical studies. For example, Derwing and Munro [19] carried out a 5-year longitudinal study on L2 accent among two L1 groups (Mandarin and Slavic) immigrants in Canada. In their findings, the Mandarin group showed no improvement, compared with the Slavic group. However, among the two groups, there existed individuals, who showed different patterns against the group means. In another earlier study [20], whether individuals progress at the same rates is the second research question, and the findings suggest comparable changes across individuals.

Thirdly, development is non-linear, following no predetermined rules. The non-linearity shown in L2 development is another point to be challenged by a dynamic systems of complexity approach [3, 15]. Non-linearity accounts for the fact that development is not rule based, therefore "language development is no longer seen as a process of acquiring abstract rules, but as the emergence of language abilities in real time" [21, p. 128].

Therefore, the focus-on-process, individual orientation and non-linear nature in development open a wide territory for the research's investigation the individual variability with a microgenetic method.

3   The Present Study

3.1   Research Questions

With the purpose of investigating the dynamic changes and variabilities in 10 Chinese EFL learners' writing, this study intends to answer the following two major questions.

Question 1: How does the learners' lexical sophistication develop over time?
Question 2: How does the learners' lexical variation develop over time?

3.2   Participants

10 Chinese EFL learners were chosen as the research subjects. They are the students majoring in English in one of the key universities in Chinese. The whole English program usually takes 4 school years, during which the students enrolled finish various courses and other academic requirements. The 10 students involve two groups: the top fives and bottom fives. The decision is made according to the results in two national tests that the students takes—TEM 4 and TEM 8 (Tests for English Majors, Band 4 and Band 8), which are both compulsory for all the English majors in China. TEM 4 is taken when they are sophomores, while the TEM 8 takes place just before their graduation. In these two tests, their language proficiency is gauged in the aspects of listening comprehension, vocabulary and grammar, reading comprehension, translation and writing. Both the validity and reliability of the tests are sound. Using the means of the two tests, the top 5 ones (among 79 students) are chosen as the high achievers, while the bottom 5 are chosen as the low achievers. Gender is not taken into account in this study (Table 1).

3.3   Research Methods

Writing tasks. The writing tasks taken as data source are the argumentations the students write for one course examination. This course lasts 6 semesters, and during the examination the students are required to write an essay with the length of 200 words to 300 words, among other exam tasks. The writing is usually argumentation, with title and/or the theme given. Since the writing takes place in examinations, reference materials, dictionaries or online resources are not available. For this research, each of the participants contribute 6 timed argumentations, all together 60 compositions, with 6 different topics, which form the data source for this study.

Table 1. The Information about the High Achievers and Low Achievers

No.	High-achievers (TEM average)	Low-achievers (TEM average)
1	Tao (77.5)	Jun (53)
2	Jiao (77.5)	Min (51.5)
3	Song (77)	Zhao (48.5)
4	Yue (76.5)	Gao (47.5)
5	Sun (76)	Yuan (41)

Table 2. Lexical Richness Indexes and Their Computations

Indexes	Contents	Computations
Lexical variation(LV)	type-token ratio	type2 ÷ token
Lexical sophistication(LS)	The proportion of 2000 high frequency words in the texts	2000 high frequency words ÷ total tokens × 100%
	The proportion of AWL(Academic World List) in the texts	AWL word families ÷ totoal tokens × 100%
	The proportion of Off-list in the texts	Off-list word families ÷ total tokens × 100%

Research Procedure. These compositions are first handwritten, then digitalized, and part of speech tagged with CLAWS 7. The lexical measurements are carried out by using the online Syntactic/Lexical Complexity Analyzer developed by Lu [22, 23].

Lexical Richness Indexes. For measuring lexical density, the formula proposed Ure [24] is employed. For lexical variation, the computation method tries to avoid the interference from the length of the texts (see [25] – [27]). Lexical Frequency Profile (LFP) [28], which is especially designed for measuring second language learners' vocabulary development, is used. Many researches have shown that LFP can differentiate learners with different language proficiency (e.g. [29]).

The contents and computations for the indexes are shown in the Table 2

4   Findings

Microgenetic analysis can reveal cognitive activities from 5 dimensions [30]: path (the sequential patterns), source (the causes), breadth (the extensions of new approaches), rate (the time bringing about the change), and variability (both inter/intra individual differences). In this study, 2 of the 5 dimensions are focused on: path and variability. The details of these two dimensions are discussed below.

4.1 Path of Change

Change is always the focus in CDST-based second language development studies. Among them, what makes microgenetic studies different is that this paradigm can "yield fine-grain as well as broad depictions of paths of change, and thus add invaluable information about them" [30 p. 488]. The fluctuations of these 10 students are depicted in the following 4 figures (Figs. 1–4).

The fluctuations displayed in Fig. 1 and Fig. 2 show the changes in 3 indices of lexical sophistication: 2K word families, AWL and Off-list word families. These indices fluctuate over time and show erratic features in different periods for different indices.

In Fig. 1, 2K words families show two high peaks, which roughly correspond with semester 2 and semester 3. The double-peaked pattern can be found to be quite homologous in 4 high achievers (Jiao, Tao, Sun and Yue), except Song. For AWL and Off-list, there exist great fluctuations all over the 6 semesters, with the later part more unstable. Roughly speaking, all of them finally arrive at the results of decrease in 2K, increase in AWL. When it comes to Off-list, the situation is a bit complicated, with one steadily decreasing (song), 4 (Jiao, Sun, Tao and Yue) achieve increase through ebbs and falls.

In Fig. 2, the situation for the low achievers is rather similar to that in Fig. 1 for the high achievers—being full of fluctuations, with the later part more erratic. And to be specific, the five students' performance in 2K word families also show rather homologous
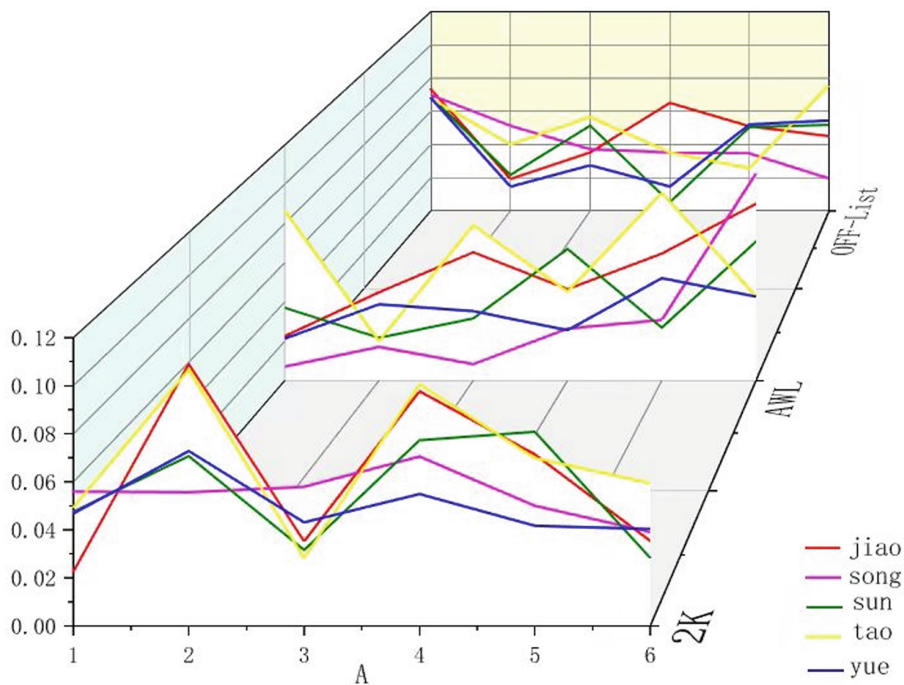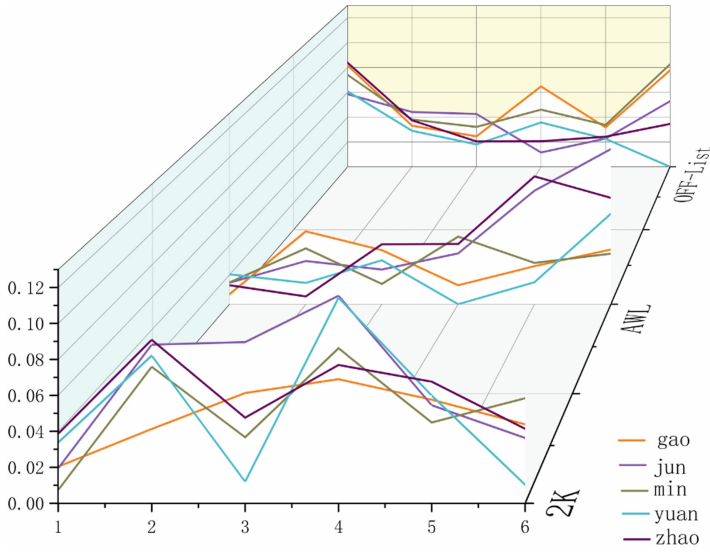

Fig. 1. LS for the high achievers

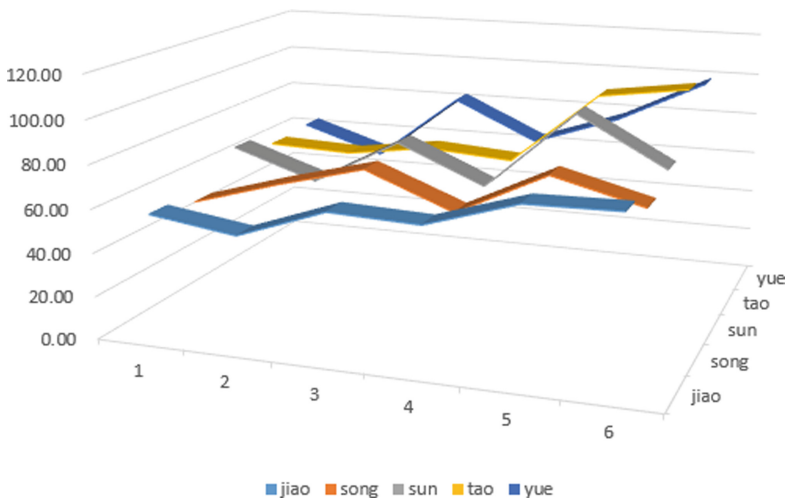**Fig. 2.** LS for the low achievers

Fig. 3. LV for the high achievers

peak-and-valley pattern of fluctuations. However, the general differences lie in the fact that the later parts are not so unstable as what we can find with the high achievers.

Figure 3 shows the changes in lexical variation (the ratio between tokens and types) for the 5 high achievers. The increase in this index means that learners can avoid the repetition in words, or word forms, and make use of more words. This time, 2 high achievers—Jiao and Tao—show rather linear increase, while the other three (Song, Sun and Yue) perform rather unstably, with Sun being the most unstable one. However, the
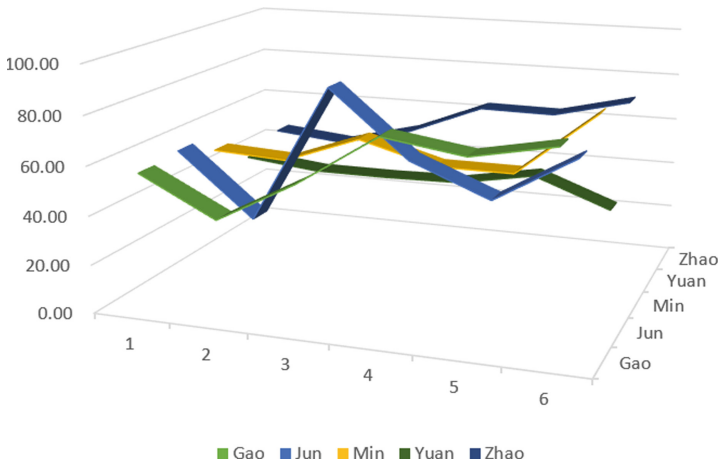
Fig. 4. LV for the low achievers

general tendency goes to the increase side, which means that all of the five high achievers arrive at better performance in lexical variation, even though they go through different trajectories.

Figure 4 shows the changes in lexical variation among the 5 low achievers. This time the fluctuations are obviously more dramatic and apparent, especially in the case of Jun. Even though most of them show general climbing in this index, Zhao, as an exception, works up rather linearly. That again, shows the different trajectories for different learners. And, the degree of fluctuations is stronger than that in the same index for high achievers.

4.2 Variability in Change

According to CDST, a complex dynamic system consists of a number of interacting subsystems, being unstable during any length of time. Thus, the observation of the different patterns of variability can reveal how the subjects' linguistic systems change over time. Due the limited space, one in the five of the two groups (high achievers and low achievers) are chosen for further discussions of the variabilities shown in the data: Song in the high achiever group and Zhao in the low achiever group.

Among the 5 high achievers, Song appears to be unique, in that all of the three indices of lexical sophistication show relatively linear-like changing patterns. In the case 2k word families, his performance is rather stable, and begins to decrease from semester 4, which means that in his vocabulary, the high-frequently used words began to drop. This decrease can be supplemented or explained by the steady increase of his use of AWL, which is supposed to indicate his maturation in using more sophisticated words. Another interesting side of the story is the steady decline of his Off-list word families. This category, which has a wider coverage, contains all the words which are not included in 1K, 2K, and AWL. A further detailed look into his use of words in this category shows his obvious breadth in vocabulary use (see Table 3).

Since the data comes from EFL learners' timed unreferenced examination writing, there exist misspelt or ill-formed words. Here "diliver," "consistently" are misspelt, while

Table 3. High achiever Song's Use of Off-list Words

No.	Writing Length	Percentage	Words Used
1	143	7.69%	Internet (8), textbooks (1), movies (1), programmes (1)
2	253	4.68%	America (1), bothered (2), classmate (1), classmates (1), diliver (1), loser (1), optimistic (1), outgoing (1), senior (1) shy (1)
3	244	3.69%	Affection (1), affections (2), fusses (1), interpersonal (2), petty (2), subtle (1)
4	227	2.64%	Desirers (1) distrustness (1), dubiousness (1), harmonious (1), heatedly (1), indifferent (1), lovers (1), trustness (1)
5	261	3.45%	Absorb (1), absorbed (1), burden (2), Chinese (1), consistently (1), disposition (1), frank (1), homework (1)
6	256	1.95	Crisis (3), inhabitable (1), vital (1)

Table 4. Low achiever Zhao's Use of Off-list Words

No.	Writing Length	Percentage	Words Used
1	155	8.39%	addickted (1), internet (10), surf (1), there're (1)
2	163	1.84%	drawbacks (1), frustration (1), mood (1)
3	148	2.04%	affection (1), interpersonal (1), shy (1)
4	195	2.05%	adoubtful (1), dishonesty (1), human being (1), unquestionablly (1)
5	207	2.42%	civilization (1), effection (1), foundament (1), insist (1), theorical (1)
6	290	3.54	envirnment (1), etc. (1), huge (1), oneself (1), plastic (2), polluting (1), pollution (2), pollutions (1)

"distrustness" and "trustness" are ill formed from "distrust," "trust," "trustfulness". Even without taking into account of these errors, the improving tendency is still valid. The fewer use of Off-list words indicate that he is using more complicated words. Coupled with his steady decrease in 2K words and steep increase in AWL words, he is definitely making visible progress in his vocabulary performance. Of course, his trajectory differs greatly from those of the other 4 high achievers.

Zhao can be used as the counterpart of the high achiever, Song. Among the five low achievers. Being different from the other low achievers, Zhao's LV increases in a linear way and one of his LS indices—AWL also increases steadily. However, there exist fluctuations in his Off-list words. A detailed analysis of his Off-List words helps us to know the real picture (Table 4).

Compared with Song, among the Off-list words, Zhao makes more errors in spelling (adickted, adoubtful, unquestionly, theorical, envirment) and word forms (effection, foundament). Without taking these erroneous words into account, the improving tendency is rather weak. Coupled with his steep increase in AWL words and moderate increase in LV, he has also made moderate progress in his vocabulary performance. However, his 2K performance is very unsteady. This probably cannot be interpreted as his poor mastery of rather highly used words.

5 Discussion

The above-mentioned changes find sequential features explanation from DST. Thelen & Smith [31, p. 342] claim that variation is the measurement of changes, the prediction for development. Therefore, according to Verspoor et al. [11], the later parts correspond to the repeller state of interlanguage development, with bumpy fluctuations, showing that the learners' interlanguage more sensitive to both internal and external influences. That means that their linguistic proficiency improves in the variety of vocabulary, the sophistication. They can avoid repetition, by using more low frequency words.

Besides the trajectories of each individual sub-systems, DST also stresses the interactions among these elements. The peaks of 2K in Figs. 1–2 and the peaks of LV in Figs. 3–4 appear alternatively, which shows that LS and LV are competitors. That means the overuse of 2000 high-frequency words will result in the decline in lexical variety. This finding corresponds with Lu's findings [23], with different interpretations, however. In Lu's study, the low significance can be is regarded in the way of the different constructs of LV and AWL in testing. In this study, they constitute common growers (see Verspoor et al. [11]). If learners are able to use the low frequency words, they hold more options in expressing themselves. The increase in options offers more opportunities to develop.

Wang and Wang's study [32] explores 17 students during one semester, each of them contributing 6 writings. This study also finds different trajectories of development among high, intermediate and low achievers. Besides, among the interactions between LS and LV, this study finds that the high achievers show the pattern of attractor state-repellor state-attractor state, rather regularly. However, the intermediate achievers remain to be rather in low stability in performance, and the low achievers keep fluctuating. What appears in Wang and Wang's study [32] does not appear here in this study. Probably the causes can be traced to different writing conditions (in Wang and Wang [32], the writing is untimed off-line writing), different topics.

This study, once again, proves that the genesis paradigm can really shed light on the things which could go unnoticed in previous group-means studies. Both the trajectory-tracking and the detailed analyses on their output can portray a fine-grained picture of the development that individual learners go through in their language proficiency, here lexical proficiency focused.

6 Conclusion

Through the quantitative and qualitative exploration of the vocabulary richness growth from 2 aspects—LS and LV, on the basis of learner longitudinal corpus, the two research

questions have been answered. The discoveries indicate the existence of the "attracting states," "repelling states," the common grower and the competitors in the process of EFL learners' written development. This result confirmed the dynamicity of the growth of interlanguage. At present, besides the methods that are commonly used in CDST research, more techniques are needed to reveal the fine-grained interconnections in this complexity.

There are certain limitations in this study in the aspects of the relatively limited amount of the corpus data, the confined scope of vocabulary in writing and the single task of argumentation in writing. Another restraint comes from the technology applied in corpus analysis, which is only capable of dealing with a certain specific variable, without integrating the interrelated factors. Despite the above imperfections, it is hoped that this research will trigger further explorations with more light shed on the development of learners' interlanguage.

References

1. Larsen-Freeman, D. (1997). Chaos/complexity Science and Second Language Acquisition. Applied Linguistics, 18(2), 141-165. https://doi.org/10.1093/applin/18.2.141
2. de Bot, K. & Larsen-Freeman, D. (2011). Researching Second Language Development from a Dynamic Systems Theory Perspective. In M. H.Verspoor, K. de Bot & W. Lowie (Eds.), A Dynamic Approach to Second Language Development: Methods and Techniques (pp. 5-24). Amsterdam: John Benjamins.
3. de Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory Approach to Second Language Acquisition. Bilingualism: Language and Cognition, 10, 7–21. https://doi.org/10.1017/S1366728906002732
4. de Bot, K., Verspoor, M., & Lowie, W. (2005). Dynamic Systems Theory and Applied Linguistics: the ultimate "so what"? International Journal of Applied Linguistics, 15(1), 116-118. https://doi.org/10.1111/j.1473-4192.2005.0083b.x
5. 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. https://doi.org/10.1093/applin/aml029
6. van Geert, P. (2008). The Dynamic Systems Approach in the Study of L1 and L2 Acquisition: An Introduction. The Modern Language Journal, 92(2), 179–199. https://doi.org/10.1111/j.1540-4781.2008.00713.x
7. Zheng, Y. (2012). Exploring Long-term Productive Vocabulary Development in an EFL Context: The Role of Motivation. System, 40(1), 104-119. https://doi.org/10.1016/j.system.2012.01.007
8. Zheng, Y. (2016). The Complex, Dynamic Development of L2 Lexical Use: A Longitudinal Study on Chinese Learners of English. System, 56, 40-53. https://doi.org/10.1016/j.system.2015.11.007
9. Caspi, T. & Lowie, W. (2010). A Dynamic Perspective on Academic English L2 Lexical Development. In R. C. Beltrán, C. Abello-Contesse & M. M. Torreblanca-López (Eds.), Insights into Non-native Vocabulary Teaching and Learning (pp. 41-60). Bristol: Multilingual Matters.
10. van Dijk, M., Verspoor, M. & Lowie, W. (2011). Variability and DST. In M. Verspoor, K. de Bot & W. Lowie(Eds.), A Dynamic Approach to Second Language Development: Methods and Techniques (pp. 55-84). Amsterdam: John Benjamins.

11. Verspoor, M., Lowie, W. & van Dijk, M. (2008). Variability in Second Language Development from a Dynamic Systems Perspective. The Modern Language Journal, 92(2), 214–231. https://doi.org/10.1111/j.1540-4781.2008.00715.x

12. Lowie, W. & Verspoor, M. (2015). Variability and Variation in Second Language Acquisition Orders: A Dynamic Reevaluation. Language Learning, 65(1), 63–88. https://doi.org/10.1111/lang.12093

13. Zhang, J. & Lu, X. (2013). Variability in Chinese as a Foreign Language Learners' Development of the Chinese Numeral Classifier System. The Modern Language Journal, 97, S1, 46-60. https://doi.org/10.1111/j.1540-4781.2012.01423.x

14. 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(4), 532–553. https://doi.org/10.1093/applin/amq001

15. Larsen-Freeman, D. & Cameron, L. (2008). Research Methodology on Language Development from a Complex Systems Perspective. The Modern Language Journal, 92(2), 200-213. https://doi.org/10.1111/j.1540-4781.2008.00714.x

16. Lavelli, M., Pantoja, A. P. F., Hsu, H., Messinger, D., & Fogel, A. (2004). Using microgenetic designs to study change processes. In Teti, D. M. (Ed.), Handbook of research methods in developmental psychology (pp. 40–65). Oxford, UK: Blackwell.

17. Hiver, P. & Al-Hoorie, A. H. (2016). A Dynamic Ensemble for Second Language Research: Putting Complexity Theory into Practice. The Modern Language Journal, 100(4), https://doi.org/10.1111/modl.12347

18. Fogal, G. G. (2019). Tracking Microgenetic Changes in Authorial Voice Development from a Complexity Theory Perspective. Applied Linguistics, 40, 432–455. https://doi.org/10.1093/applin/amx031

19. Derwing, T. & Munro, M. J. (2013). The Development of L2 Oral Language Skills in Two L1 Groups: A 7-year Study. Language Learning, 63(2), 163-185. https://doi.org/10.1111/lang.12000

20. Derwing, T., Thomson, R. I. & Munro, M. J. (2006). English Pronunciation and Fluency Development in Mandarin and Slavic Speakers. System, 34(2), 183-193. https://doi.org/10.1016/j.system.2006.01.005

21. Evans, J. (2007). The Emergence of Language: A Dynamic Account. In E. Hoff & M. Shatz (Eds.), Blackwell handbook of language development (pp. 128–148). Malden, MA:Blackwell.

22. Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics, 15(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu

23. Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. TESOL Quarterly, 45(1), 36-62. https://doi.org/10.5054/tq.2011.240859

24. Ure, J. (1971). Lexical Density and Register Differentiation. In G. E. Perren & I. L. M. Trim (Eds.), Applications of Linguistics (pp. 443-452). Cambridge: Cambridge University Press.

25. Larsen-Freeman, D. (2009). Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. Applied Linguistics, 30(4): 579–589. https://doi.org/10.1093/applin/amp043

26. Wen, Q. (2006). The Differences between Oral and Written Lexicon Use among English Majors. Foreign Languages and Their Teaching, 208(3), 9–13. https://doi.org/10.13458/j.cnki.flatt.003241

27. Bao, G. (2010). A comparative study of EFL learners' linguistic complexity. Modern Foreign Languages, 33(2), 166-176.

28. Laufer, B. & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in Written Production. Applied Linguistics, 16(3): 307-322.

29. Laufer, B. & Nation, P. (1999). A Vocabulary-size Test of Controlled Predictive Ability. Language Testing, 16(1), 33-51.
30. Siegler, R. S. (2006). Microgenetic Analyses of Learning. In D. Kuhn & R. S. Siegler (Eds.), Handbook of Child Psychology, Volume 2: Cognition, Perception, and Language (6th ed.)(pp. 464–510). New York: Wiley.
31. Thelen, E. & Smith, L. B. (1994). A Dynamic Systems Approach to the Development of Cognition and Action. Cambridge, MA: Bradford Books/MIT Press.
32. Wang, Y. & Wang, Y. (2020). Exploring Microgenetic Changes in L2 Productive Vocabulary Development from a Dynamic Systems Perspective. Foreign Language Education in China, 3(1), 44-52.