The Dynamics of English Writing Development in Advanced Chinese Learners

Hou, Junping

Publication date:
2017

Chapter 6

The dynamics of L2 writing development over time: A strong learner versus a weak learner

This chapter is a slightly edited version of Hou, J., Loerts, H., & Verspoor, M. H. (under review). The dynamics of L2 writing development over time: A strong learner versus a weak learner. Submitted (April 10, 2017) to *International Journal of Applied linguistics.*

6.1 Introduction

The present paper is part of a larger project which involved 3 groups of Chinese students: a senior high school group, a university lower level group and a university higher level group. The study of Chapter 3 showed that the university higher-level group did not make any progress over the course of 18 months in holistic scores and only a bit of progress in a few isolated analytical scores obtained by means of a computational analysis on complexity. Because it was difficult to believe the group had made so little progress over 18 months of English input and exposure (total time is about 300 hours), the assumption was that the learners might have progressed in more subtle areas of language. Therefore, two subsequent studies investigated whether the learners had improved in their use of chunks (see Chapter 4) and text quality (cohesion and coherence) (see Chapter 5). These previous group studies with pre-post designs found significant overall progresses in the use of collocations and topic based coherence respectively, but showed averages with some learners making substantial progress in different areas and some learners who did not.

The current study in this chapter is an attempt to explore what the differences might be over time in developmental patterns between a strong learner (one who made progress in the holistic scores) and a weak learner (one who made no progress in the holistic scores) within a Complex Dynamic Systems Theory (CDST) framework. After explaining the CDST framework, we will present the study. The case studies are explorative in that we will try to find out retroactively what patterns within the developing L2 system may have contributed to growth for one learner and stagnation for the other.

6.2 The CDST framework

To explore developmental patterns we will make use of the CDST framework, which holds that language is a complex system, and language development is a complex dynamic process. According to one of the founding fathers of complexity studies, we can trace the meaning of complexity to its original sense in Latin: *Plexus* means braided or entwined and *complexus* means braided together. Complexity in the theoretical sense thus refers to the intricate inter-twining or inter-connectivity of elements within a system and between a system and its environment (Gell-Mann,1995,

p. 96). Language--at both a societal and individual level--is thus a complex system in that it is composed of many intertwining, interconnected elements that affect each other and will never be static as the language used in society influences the language the individual uses and the other way round. Therefore, language is not only complex but a complex adaptive system as it is able to adapt in and evolve with a changing environment. Change should thus be seen as co-evolution with all other related systems. Humans and the language they use thus co-adapt and co-evolve with their environment, which may be an instructional context in the case of L2 development. L2 development is a complex dynamic process in that it behaves like many other natural systems that evolve such as brains, immune systems, ecologies, and societies in that they are characterized by complex behaviors that emerge as a result of often nonlinear spatio-temporal interactions among a large number of component systems at different levels of organization (Chan, 2001).

Development within a CDST framework refers to either ‘the growth or increase in level of more developmentally advanced complex variables’ or ‘the decline or decrease of less developmentally advanced variables’ (Verspoor & Behrens, 2011, p. 85). According to Van Geert (1991), the human cognitive system and subsystems, including the language system, can be compared to the ecosystem consisting of species. Similar to species in the ecosystem, there are relevant subsystems such as lexicon, syntax, morphology, and general problem-solving skills in the language system. Just like species, subsystems in language may engage in various types of functional relationships with each other.

Development is a dynamic product of growth under limited resources, both internal and external ones. In the case of L2 developmental studies, limited internal resources may be due to other subsystems in the more general cognitive system such as general aptitude or affective system as degree of motivation and eagerness to learn. Another internal resource is attention, e.g. the limited amount of information one can deal with simultaneously (Kahneman, 1973; Miller, 1956) or the limited range of one’s working memory (Baddeley, 1976). The limited external resources may be environmental such as the degree of exposure or the effectiveness of an instructional approach (Fogel & Thelen, 1987; Thelen, 1989).

Depending on the point in the developmental process the system is, the subsystems may show strongly similar or dissimilar growth rates and growth-onset times. Some are supportive in that the growth in one subsystem supports the growth in another. This is a mutual support interaction in that they positively affect each other's growth. This may mean that the subsystems are coordinated in that they no longer behave independently of each other. Others are competitive in that the growth in one subsystem relates to the decline in another. This is a mutually competitive interaction in that they negatively affect each other. This may mean that the subsystems are competing for attention. Some are neutral in that the growth in one subsystem is not directly related to the growth or decline of another subsystem.

As far as we know, coordination of subsystems has not yet been explicitly addressed in longitudinal L2 studies from a CDST perspective. But as Van Geert and Verspoor (2015) point out, coordination of interrelated subsystems may suggest that some form of automatization is taking place. For example, the reading of words by a beginning reader involves many functional brain centers, motor processes, sensory processes, perception-action loops and each of these components represents a particular range of variation. If these components are not or only poorly coordinated, they may vary independently over time. However, functional behavior requires a coordination of these components, for instance between eye movements and text processing in reading, or between articulatory motor processes and processes governing word order. Coordination thus implies that the components no longer behave independent of one another, and this mutual dependency corresponds with a reduction in the number of dimensions required to specify the behavior of the system.

Before subsystems coordinate, separate subsystems may go through phases of stability and variability. Phases of stability simply refer to a phase in which the system may still fluctuate, but the variability is rather limited. Phases of variability refer to phases in a series in which the pattern and/or bandwidth change drastically compared to the remainder of the series. According to CDST, such phases of variability are central to a self-organizing, developing system in that they are "a metric of stability and a harbinger of change" (Thelen & Smith, 1996: 342). Variability is especially large during periods of rapid development because at that time the learner explores and tries out new strategies or modes of behavior that are not always successful (Thelen & Smith, 1996).

It has been shown in quite a number of L2 studies from a CDST perspective that the developmental trajectory of almost any of the separate subsystems is not linear but rather variable, which should thus be seen as part and parcel of normal development. Larsen-Freeman (2006) designed a repeated task study on five intermediate Chinese learners of English in their oral and written production, measuring the emergence of complexity, accuracy and fluency (CAF) with both broad and specific measures. She concluded that as a learner uses language her language resources also change. Moreover, none of the learners showed similar patterns in the CAF graphs. Spoelman and Verspoor (2010) tracked a Dutch learner's acquisition of written Finnish from beginner to high intermediate over three years by examining CAF measures and concluded that the interactions of variables in systems showed "classic jumps, transitions, and non-linear" (p.550) development. Verspoor et al. (2008) traced the development of an advanced learner, and showed that even at this advanced level there were different variability patterns and interactions among variables. Penris and Verspoor (2017) showed how for their advanced learner, the finite verb ratio (FVR; general measure of complexity) and noun phrase length developed almost synchronously, which suggested that the increase and decrease in FVR were mainly due to noun phrase length, and could point to coordination. However, it could also be argued in this case that the measures partially tap into the same construct and are therefore not truly two independent subsystems that become more coordinated.

Variability could be related to a great number of different factors, including particular task or motivation of the student on the particular day the task was performed. However, some of the variability may be due to reorganization taking place in the particular linguistic subsystem itself; for example, it may be the subsystem on which a learner is focusing more than normal to be able to master it or it may be competing with other subsystems, often recognized as U-shaped behavior in L1 and L2 studies (cf. Plunkett & Marchman, 1991). In CDST terms such behavior can be evident in peaks and regressions that may be tested for significance (Spoelman & Verspoor, 2010, p. 4). A significant peak indicates a developmental overuse or overgeneralization of a target construction. Van Dijk et al. (2011) analyzed the Cancino, Rosansky and Schumann's (1978) data and showed that one of the learners developed a significant peak in the *don't* construction, which was also observed in another learner of the same age. Penris and Verspoor (2017) found that their advanced learner significantly overused complex sentence and finite dependent clauses in the earlier stage of his development and unique lexical items and non-finite dependent clauses in the more advanced stage.

The current study aims to explore differences in developmental patterns of a number of linguistic subsystems between a strong learner and weak learner at rather advanced levels in L2 developmental patterns that may retroactively explain growth or stagnation. We will do so by comparing the data of the two learners in number of developmental peaks, degrees of variability, and degrees of coordination.

6.3 The study

Based on the findings in Chapter 3 which showed that an advanced group of learners had not progressed significantly in either holistic or analytical scores on their texts at all over the course of 18 months, two learners were selected: a strong one who actually made progress and a weak one who did not.

6.3.1 Subjects

From the university higher level group, two learners were selected, a strong one and a weak one. Both had a complete data set of 12 texts written over the course of 18 months. Julia was stronger than David on both the pre- and posttest scores (holistically rated texts) and made progress, whereas David did not (for detail see Table 6-1).

Both students were enrolled at the same selective university. They had similar CEE scores (college entrance examination), both had passed the proficiency test after entering this university and they were enrolled in the same class to prepare for the College English Test (CET-6). They had the same instructional setting and each contributed the same 12 writing assignments.

6.3.2 Texts

Each learner submitted 12 essays (see Appendix 8 for first and last text written by each learner) given as homework assignments on different topics (see Appendix 9) written without time pressure. The use of dictionaries and other resources were allowed during their writings. There was no word limit, which resulted in differences in text length.

6.3.3 Holistic scoring

In the previous study of Chapter 3, the first two (pretest) and last two (posttest) texts written by our subjects were rated holistically together with 172 other texts by a group of 9 trained raters with a Complexity, Accuracy, Fluency, Idiomaticity and Coherence (CAFIC) rubric. In the current study, all texts (including the first two and last two) were scored holistically by a new team of 3 trained raters (two of the authors and an independent rater) on the same rubric. To do so, the texts were presented in randomized order and rated independently. The level of agreement among the previous three raters' and the new three raters' in this study in rating the two pretest and posttest texts over CAFIC total measure showed excellent interrater reliability (ICC (3, 6) = .965). For the new set of 24 texts, the level of agreement over CAFIC total measure among the three new raters also showed excellent interrater reliability (ICC (2, 3) = .928), allowing us to use the average scores of the three new raters' for further analysis.

6.3.4 Operationalization of constructs

To be able to trace development, appropriate measures to operationalize constructs to represent linguistic subsystems at the advanced level had to be determined. The subsystems used in the CAFIC rubric were taken as a starting point and were operationalized as follows.

Complexity: our study in Chapter 3 showed that the advanced group did not make any progress at all in holistic CAFIC scores and only in a few isolated analytical complexity scores obtained by means of the Synlex Analyzer (Lu, 2010). As there were no clear possible measures available to operationalize complexity from this study, we decided to use average word length (AWL) (the total number of lexical items divided by the number of letters in the whole text) as it proved to be one of the better measures for advanced learners in that it correlated well with text ratings (Verspoor, Lowie, Chan & Vahtrick, 2017) and increased over time in Penris and Verspoor (2017). For sentence complexity, the finite verb ratio (FVR) (the number of finite verbs divided by all tokens in the text and multiplied by 100) was used as it was the one measure that correlated with all lexical and syntactic measures over time in Penris and Verspoor (2017) and shows more internal complexity than other measures such as average sentence length.

Accuracy: In our subsequent studies in the development of the advanced group, we did not look further at accuracy measures analytically, because counting errors or error-free clauses is difficult to do at this level as only few errors are made in grammar, morphology or syntax (which can all be counted easily) but more so in the lexicon and awkward phrasing, which may not be grammatically incorrect, but still confuse the reader. Therefore, we decided to operationalize accuracy as the holistic score given on the CAFIC rubric. In the study of Chapter 4 with all 72 texts of the advanced group, the accuracy score correlated highly with all other sub-components of the CAFIC rubric and the total CAFIC score.

Fluency: In the CAFIC rubric we asked raters to look at text length, from 1 (shortest texts) to 5 (longest texts) in the corpus. Even though our writing tasks were not timed, we agree with Wolfe-Quintero et al. (1998) that text length is a good indication of fluency in writing samples. They quote Larsen-Freeman (1978) who states that “subjects with a higher proficiency tended to write longer compositions—perhaps because of their fluency” (p, 444). In the current study, we used the common operationalization of fluency as the actual number of words in each text.

Idiomaticity: For the construct of idiomaticity in the CAFIC rubric, raters were asked to look at how idiomatically words are combined. For example, does the writer use chunks such as “stay up late” at all and if so, how many? They were asked to assign a low score (1) if \things were expressed very awkwardly and did not contain any native-like chunks and a high score (5) if many fully idiomatic collocations and longer chunks (conventionalized ways of saying things) that sound like “natural” and idiomatic English were used. The construct was later measured analytically in Chapter 4 who showed that the advanced group had improved significantly in the number of chunks they used, specifically collocations. Also chunk coverage (the total number of chunk words divided by the total number of words) increased significantly over time, indicating that the use of more chunks benefited the increased proficiency in written texts. To be able to count items efficiently and objectively, we decided to take the collocations ratio (the number of collocations divided by the number of words in the text and multiplied by 100) as operationalization for the construct of idiomaticity. Collocations were defined as chunks with collocating nouns, adjectives, verbs, and adverbs, prepositions, pronouns, such as *choose wisely, a positive attitude, pay attention,* etc. (details see Chapter 4).

Coherence: In the CAFIC rubric, raters were asked to rate coherence as follows:

> *Read the text as a whole without paying attention to the areas above [complexity, accuracy, fluency or idiomaticity], but pay attention to "flow". Do sentences have natural flow so that you can understand what is being said without having to reread sentences? Is there a focus and are all sentences in a paragraph related? If there are more paragraphs, are they connected?*

Like all other CAFIC scores, the scores on coherence correlated well with all other parts of the rubric and with the total CAFIC score (see Chapter 5). In the study of Chapter 5 cohesion and coherence were measured analytically with automated and hand-counted measures. The one measure that correlated well with almost all other measures and showed significant improvement over time was Density of Moves (DoM), originally called "the number of moves (between two related key-concepts)/10 T-units" in topic-based coherence analysis (Watson Todd et al. 2004), we applied this measure as operationalization of coherence in this current study.

6.3.5 Statistical analyses

To explore general similarities and differences between the two learners, the data were inspected with correlation analyses to study relationships between variables and independent sample *t*-tests to compare the two learners in SPSS 23.

To explore developmental patterns over time, first each single variable (representing a subsystem in the learner's language) was plotted in Excel graphs to visually inspect possible developmental patterns. To discover whether there are developmental peaks, whether there are differences in the degree of variability between the learners, or whether the variables showed significant growth, various Monte Carlo analyses were conducted using Poptools (Hood, 2004). A Monte Carlo analysis uses the original data in a series and then by re-shuffling the data randomly in a number of iterations (usually 5000) within a model set up with a testing criterion, the analysis calculates the chance that the same number as the testing criterion is found. If the chance is less than 5% or more than 95% it is considered a significant finding (cf. Van Dijk, et al, 2011; Lowie & Verspoor , 2015). To detect more general developmental patterns, linear trend lines or smoothers such as LOESS curves with Table Curve 2D (Systat Software Inc., 2015) are applied.

Within CDST the interactions among variables are assumed to change over time. For example, they may start out as competitors and over time become supportive, or the other way round. To find out changing relationships over time, we first smoothed (LOESS 41.7%) the data and then normalized (0-1) the data to be able to detect general interacting patterns.

6.4 Results

6.4.1 Statistical analyses

Table 6-1 provides an overview of the CAFIC scores on all 12 texts written and shows that Julia, the strong learner, started with a higher CAFIC total score and improved slightly from pre to posttest. The weaker learner David started with lower scores and did not seem to improve.

Table 6-1. CAFIC scores on 12 texts by each learner.

	Pretest scores		**Scores on other texts**								**Posttest scores**		
	1	**2**	3	4	5	6	7	8	9	10	**11**	**12**	**Average**
Julia	**18,53**	**20,33**	22,11	11,95	11,43	16,58	19,18	21,02	15,88	17,86	**23,52**	**24,37**	**18,56**
David	**16,63**	**18,13**	12,53	12,35	17,45	14,13	15,43	17,17	14,37	14,13	**15,58**	**15,36**	**15,27**

Table 6-2 gives an overview of the differences between the learners. The results of an independent sample *t*-test and Wilcoxon Signed-Rank Test using the measures of all 12 texts showed that, on average, Julia wrote significantly longer texts than David and was more advanced in most CAFIC sub-scores except accuracy. There was also a trend in average word length (AWL).

Table 6-2. The differences between the two individuals' data series in measures.

Measures	Julia		David		Differences	
	M/Median	*SD*	*M/Median*	*SD*	*T/Z*#	*Sig.*
Text length	292.58	87.61	204.92	36.12	3.205	**.006****
Complexity	4.02	.80	3.38	.51	2.314	**.030***
Accuracy	3.48	.76	3.26	.42	.873	.392
Fluency	3.47	1.24	3.26	.42	2.428	**.027***
Idiomaticity#	3.83	.73	3.07	.37	1.837	**.002****
Coherence	3.70	.78	3.03	.47	2.562	**.018***
CAFIC	18.56	4.12	15.27	1.86	2.523	**.023***
AWL	4.50	.26	4.32	.22	1.901	.070
FVR#	10.80	1.64	12.34	3.28	1.225	.100
Coll	6.53	2.19	5.71	2.37	.884	.386
DoM#	9.16	5.71	11.10	6.73	1.225	.100

As the data did not follow normal distributions (SW < .870, df = 12, $p < .065$), Wilcoxon Signed-Rank Test was used and the *Z*-Statistic is provided instead of *T*-value, median is used instead of mean.

**. Significant at p < 0.01 (2-tailed). *. Significant at $p < 0.05$ (2-tailed).

6.4.2 Correlation analyses

In our previous group studies, we showed that text length, CAFIC scores, sub-scores and analytical measures correlated well with each other. Here we look at each learner separately to see to what extent the constructs correlate with each other for that specific learner.

Table 6-3 shows that in Julia's 12 texts, the CAFIC total score and CAFIC sub-scores correlate highly with each other and text length. AWL shows very low and negative correlations with the CAFIC measures; FVR (finite verb ratio) shows strong positive correlations with all CAFIC measures; Coll (ocations) and DoM (density of moves) show no significant relations with any other measures in the table.

Table 6-4 shows that in David's 12 texts, the CAFIC total score and CAFIC sub-scores do not all correlate with each other, nor with text length; AWL correlates with accuracy; FVR correlates with complexity and CAFIC. Coll shows no significant relations with any other measures; DoM correlates with complexity, accuracy, coherence and CAFIC.

Table 6-3. Julia: Correlations between holistic and analytical measures**.**

	Complexity	**Accuracy**	**Fluency**	**Idiomaticity**	**Coherence**	**CAFIC**	**AWL**	**FVR#**	**Coll**	**DoM#**
Text length	.850**	.906**	.959**	.881**	.864**	.942**	-.253	.736**	-.111	.189
Complexity	1	.926**	.831**	.926**	.846**	.946**	-.051	.746**	-.315	.196
Accuracy		1	.939**	.935**	.864**	.977**	-.176	.683*	-.390	.221
Fluency			1	.869**	.838**	.949**	-.235	.601*	-.247	.301
Idiomaticity				1	.935**	.976**	-.113	.755**	-.430	.161
Coherence					1	.932**	-.385	.734**	-.501^	.224
CAFIC						1	-.206	.748**	-.379	.189
AWL							1	-.434	.190	.217
FVR#								1	-.063	-.280
Coll									1	-.406
DoM#										1

As FVR and DoM data did not follow a normal distribution (SWs = .869, df = 12, p = .062), Spearman's rho was used.

**. Correlation is significant at the 0.01level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

^ Correlation failed to reach significance $p < 0.097$ (2-tailed).

Table 6-4. David: Correlations between holistic and analytical measures.

	Complexity	Accuracy	Fluency	Idiomaticity#	Coherence	CAFIC	AWL	FVR	Coll	DoM#
Text length	.076	-.086	.299	-.332	.275	.103	-.450	-.376	.068	.467
Complexity	1	.596*	.858**	.319	.730**	.939**	.012	.650*	-.046	.576*
Accuracy		1	.272	.413	.539^	.776**	.677*	.330	.100	.743**
Fluency			1	.027	.596*	.781**	-.317	.641*	-.084	.332
Idiomaticity#				1	.175	.497	.382	.145	.315	.167
Coherence					1	.811**	.089	.392	-.135	.606*
CAFIC						1	.231	.618*	.037	.626*
AWL							1	.066	.079	.291
FVR								1	-.051	-.056
Coll									1	-.211
DoM#										1

As the Idiomaticity and DOM data did not follow a normal distribution (SWs < .870, df = 12, *p*s < .065), Spearman's rho was used.

**. Correlation is significant at the 0.01level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

^ Correlation failed to reach significance p < 0.071 (2-tailed).

6.4.3 Variability and variation over time

Figure 6-1 shows the development over time in CAFIC total scores. Julia started with a higher score, had a dip for texts 4 and 5, but then continued to improve as the trend line shows, but the increasing slope was not significant (p = .59). David started lower, regressed and had a slight dip, and then remained rather stable as revealed by the significantly stagnant slope (p > .95).

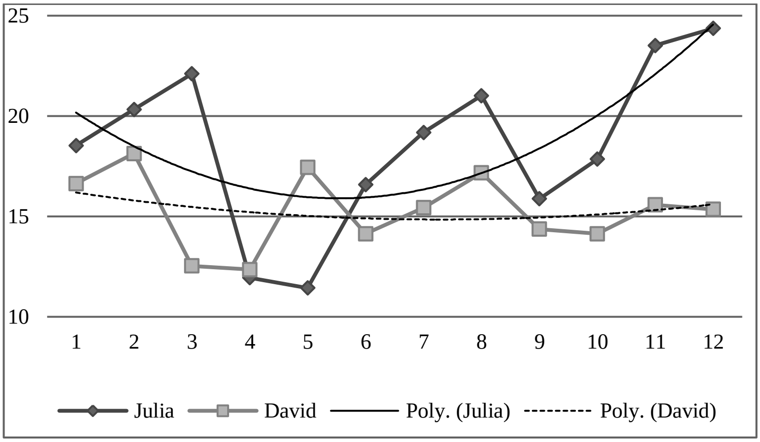

Figure 6-1. Development of the CAFIC (total) score over 18 months in 12 texts for each learner.

Figure 6-2 shows the development over time of the CAFIC sub-constructs. Comparing Julia to David, visual inspection suggests Julia increased more in all variables towards the end. Julia's FVR seems less variable all along. Her accuracy and text length seem more variable and were higher at the very beginning and increased at the end. As far as developmental peaks are concerned, Julia seems to have a peak in the AWL in Text 11, in accuracy and text length in Text 12, and in DoM in Text 9. David seems to have peak in the FVR in Text 2, in accuracy in Text 5, and in DoM in Text 1.

To test if any of these observed slopes increased or decreased significantly in each individual time series of scores, we conducted Monte Carlo analyses on the global variability of each time series as testing criteria, which was determined as the average of a moving range across 3 data points. The result showed stagnant developments in David's accuracy and AWL

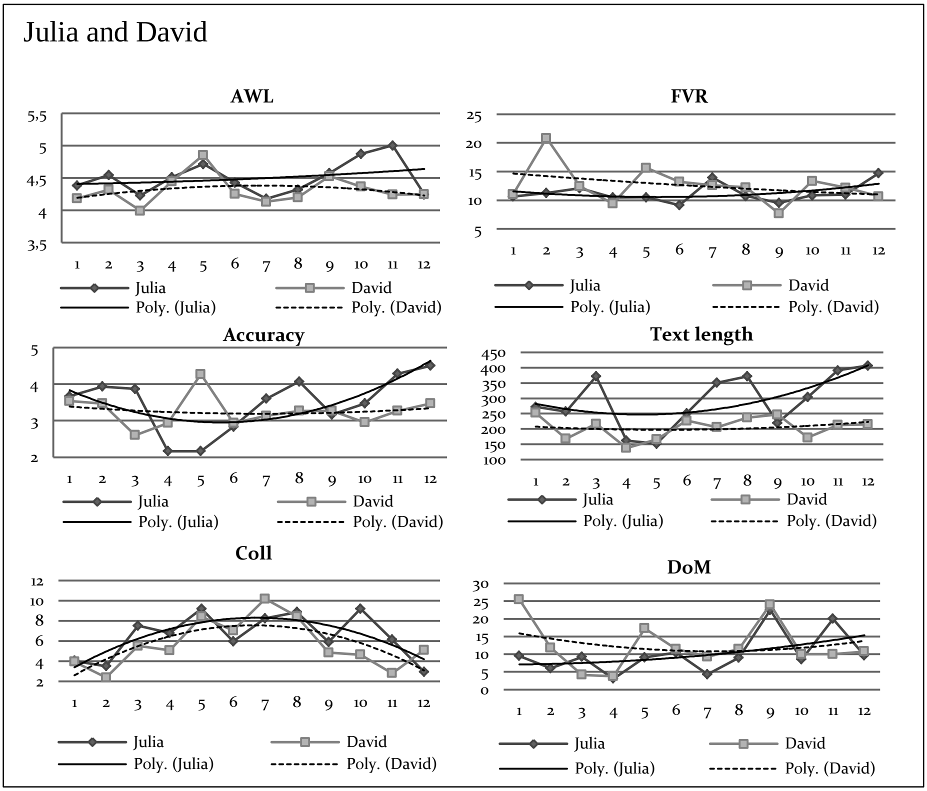

Figure 6-2. Development of the subsystems over 18 months in 12 texts for each learner.

(*ps* > .95).

To test whether there is a difference in the degrees of variability between the two individuals for all variables, we conducted Monte Carlo analyses on the slopes of each data series. The results showed the differences between their text length and AWL were significant ($p < .05$).

To see if one learner is more variable in developing of these criteria than the other, each measure was compared separately with Monte Carlo Analyses. The result showed that David is more variable in developing FVR than Julia over time ($p < .05$).

To see if there were any significant peaks or dips in the data, each measure for each learner was submitted to a Monte Carlo analysis and the results showed that Julia developed a significant peak in AWL ($p < .05$).

6.4.4 Interactions of variables over time

Even though the correlation analyses have shown how variables may correlate in general, they do not give insight into their development over time. To be able to see how for each learner variables interact over time, we curved smoothed (LOESS, 41.7%), normalized (0-1) data to see beyond the day to day variability (Figure 6-3).

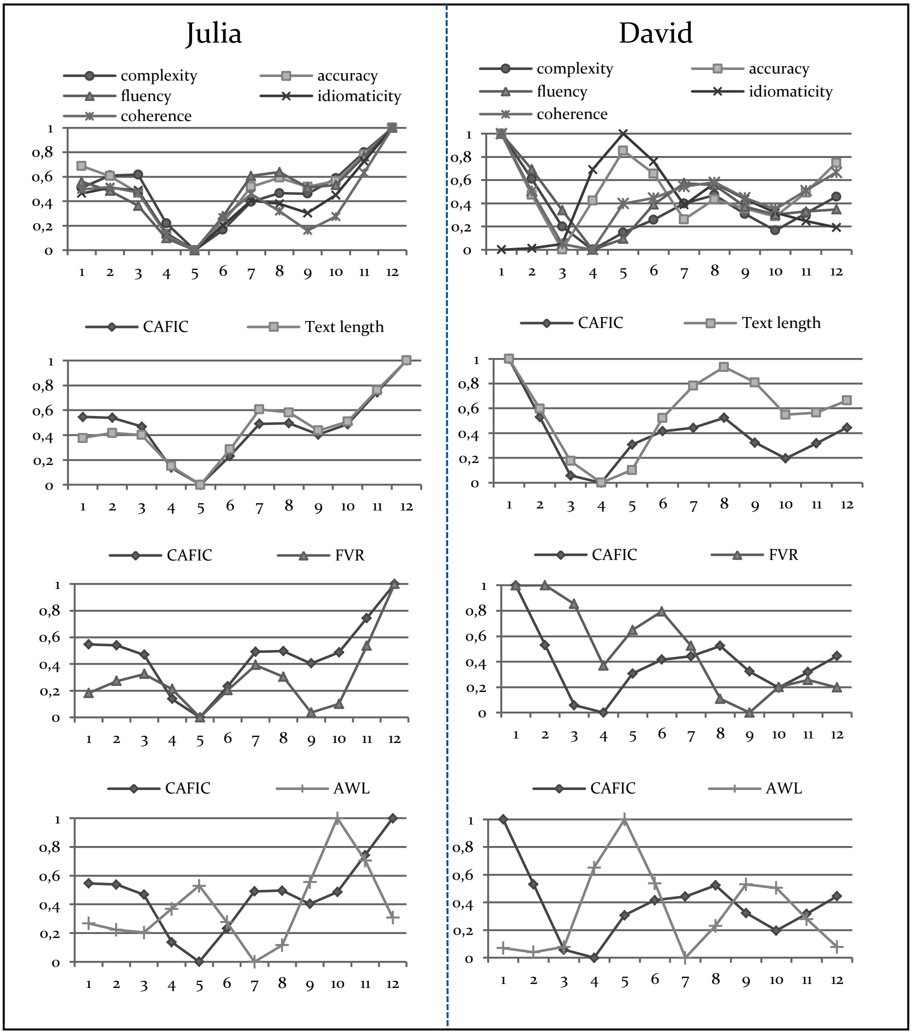

Figure 6-3. The correlations and interactions between variables within learners with smoothed (LOESS 41.7%), normalized (0-1) data.

In line with the general correlations given above, we can clearly see from Figure 6-3 that, especially in the development of CAFIC sub-scores, Julia was much more coordinated than David as the different variables developed synchronously. In other words, the subsystems were not competing with each other. Also the analytical measures (text length and FVR) developed rather synchronously with CAFIC total scores. The only exception for Julia was AWL, which is the one construct in which she made a developmental jump.

In contrast, David's development showed more competitive or random patterns. Accuracy and idiomaticity develop rather synchronously, suggesting that David may have had lower accuracy scores because of non-idiomatic language use. These two variables together show a strong competitive pattern with the other variables. CAFIC and text length do not develop synchronously either and there is a clear competitive pattern visible between CAFIC and FVR around text 6. Like Julia, David shows a competitive relation between CAFIC and AWL, but unlike Julia, the competition does not lead to progress but regress.

6.5 Discussion and conclusion

The current study compared the dynamic patterns of development of two Chinese learners of English over the course of 18 months. Based on the findings of our previous studies in Chapter 3, 4 and 5 we selected one learner who had made progress from pretest (two texts) to posttest (two texts) and one learner who had not. The goal was to see if any dynamic developmental patterns might be discovered that could clarify differences in growth versus stagnation.

Several variables were investigated. General proficiency was operationalized as the total score on the CAFIC rubric, which was a holistic evaluation of each text. Also each sub-score of the CAFIC rubric (complexity, accuracy, fluency, idiomaticity and coherence) was examined in the overall correlation analysis. There were four analytical measures: average word length (AWL) and finite verb ratio (FVR), collocation ratio (Coll) and density of moves (DOM). AWL and FVR had proven to be the most useful in other studies tracing development of English as an L2 at the more advanced stages of proficiency (Verspoor et al., 2017; Penris &

Verspoor, 2017) and were thus taken as variables to represent lexical complexity/sophistication and syntactic complexity respectively. An increase in AWL reflects longer and more advanced and academic words which in turn usually correspond to the use of less frequent lexical items. FVR reflects the ratio of each finite clause, which may become longer with elaborated noun phrases or other non-finite constructions. From the two other group studies, one on idiomaticity (Chapter 4) and the other on coherence (Chapter 5), the two measures Coll and DOM had shown significant improvement over time and were taken to operationalize the variables idiomaticity and coherence respectively. Collocations involve combinations with nouns, adjectives, verbs, and adverbs, prepositions, and pronouns. Density of moves involved the average number of moves per 10 T-units. The fewer different key concepts, the more focus there is on the topic.

Based on the CDST literature and in particular the literature concerning longitudinal writing studies at the more advanced levels, we conducted several analyses. We looked for overall correlations among variables, the assumption being that the more correlated they are, the more coordinated they are. What we found is that overall (taking all 12 texts) Julia's CAFIC sub-scores were all highly coordinated with each other. The CAFIC scores also had a strong correlation with text length and FVR. When looking at development over time, Julia was relatively more coordinated in developing variables and showed improvement in most measures. The only variable that showed clear competition with CAFIC was AWL (see Figure 6-3) and a developmental jump in AWL. Analyses on the difference between their degrees of variability also showed significance in text length and AWL. These observations suggest that overall Julia's English language system was a coordinated, automated whole (the mutual dependency of the subsystems corresponds with a reduction in the number of dimensions required to specify the behavior of the system). As most of the system was coordinated and automated, the various subsystems did not require extra attention and just as in Verspoor et al. (2017) and Penris and Verspoor (2017), this advanced learner was able to become more sophisticated in the lexicon.

For David the overall pattern of development is much more scattered. First of all, his FVR was found to be significantly more variable than Julia's, and this variability probably affected all the other correlations found. The variability in David's FVR indicates that his sentence complexity was not stable at all and may have competed at different times with other variables he was focusing on. Often degrees of variability suggest change, usually for the better, but in David's case, for the worse. The LOESS curves in Figure 6-3 suggest that accuracy and idiomaticity developed synchronously, suggesting that they were interdependent, the decreasing in the variability of his CAFIC, accuracy and AWL measures showed significant stagnation. His unidiomatic language may have been reason for the low accuracy scores.

It is surprising that we found few meaningful patterns with collocations and density of moves for either learner. These measures were chosen because they showed significant increase over time in the group studies, but one could argue that these analytical scores focus only on one specific sub-aspect of the whole construct. Idiomaticity is more than just the number of collocations used and coherence is more than the number of turns made in a writing. When operationalized as a holistic score, there were strong correlations for Julia (Table 6-3) and few correlations for David (Table 6-4) with other measures, suggesting that holistic scores may do the construct as a whole more justice.

The variability we found in both learners is in line with all other previous longitudinal studies from a CDST perspective. Writers do not remain stable, not even on such global measures as overall proficiency. This variability may be due to various factors. In addition to the effect of writing topic, we assume that the learner's internal resources (e.g., focusing on different aspects of the language; different degrees of motivation at different times) and external resources (e.g., the amount of authentic input that was received before time of writing) all interact and affect the product. However, high degrees of variability may drive the process of development (Van Geert & Van Dijk, 2002), which may be either progress or regress. This was evident in the developmental jump the stronger writer made in AWL.

The original question was whether we could detect any patterns that would clarify why one learner was stronger than the other in their developing linguistic system. We can point to two main differences between the learners. First of all, initial conditions were different. Julia apparently was already more advanced in almost all of our measures than David to begin with, and may thus have been in a more advanced phase of L2 development in which her linguistic system no longer consisted of independent subsystems that go their own way but a coordinated whole, giving her room to make a developmental jump in AWL. David's linguistic system, on the other hand, was much less coordinated, especially evident in the fact that his fluency (text length) did not correlate with any other measure. David's data showed a few correlations with complexity and some other measures, but in one of the main complexity measures (FVR) he showed significantly more variability than in Julia's series, but no significant jump. After data point 5 (Figure 6-3), Julia showed growth in proficiency (CAFIC score) and David not only stagnated but even regressed in some of the measures.

Unfortunately, we did not collect much personal data on these learners at the beginning of the study and we do not know what made Julia progress more than David. However, as has become clear from what we have found in Chapter 2, learners from the same programs differed greatly in their motivation for studying English and the amount of extra exposure to English they received. It would not surprise us if Julia was like a minority of the students who actually liked English, was not just driven to study because of the tests, and sought extra input. David, on the other hand, must have been like the majority of students, who study English mainly to pass the required test and are frustrated because it takes a lot of time and effort without clear results.

Our findings are difficult to compare with those of other studies as this is the first time an attempt has been made to compare a strong learner with a weaker learner, and this is also the first longitudinal L2 study from a CDST perspective that has suggested that coordination of subsystems may be a factor in further development.

The study has its limitations, of course. For one thing it would have been better to have more learners, more data points and better control of topics. However, a CDST framework by definition looks at single cases, all assumed to be unique with individual learning paths and they are not meant to be generalized to other learners. For another thing, it would be interesting to explore further whether coordination among different subsystems has to occur at the advanced levels for learners to make further progress in the language.