
Linguistic complexity in second language development: variability and variation at advanced stages

Marjolijn Verspoor, Wander Lowie, Hui Ping Chan et Louisa Vahtrick



Éditeur
ACEDLE

Édition électronique

URL : <http://rdlc.revues.org/1450>

DOI : [10.4000/rdlc.1450](https://doi.org/10.4000/rdlc.1450)

ISSN : 1958-5772

Référence électronique

Marjolijn Verspoor, Wander Lowie, Hui Ping Chan et Louisa Vahtrick, « Linguistic complexity in second language development: variability and variation at advanced stages », *Recherches en didactique des langues et des cultures* [En ligne], 14-1 | 2017, mis en ligne le 30 janvier 2017, consulté le 04 juillet 2017. URL : <http://rdlc.revues.org/1450> ; DOI : [10.4000/rdlc.1450](https://doi.org/10.4000/rdlc.1450)

Ce document a été généré automatiquement le 4 juillet 2017.



Recherches en didactique des langues et des cultures is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Linguistic complexity in second language development: variability and variation at advanced stages

Marjolijn Verspoor, Wander Lowie, Hui Ping Chan et Louisa Vahtrick

Introduction

- 1 The current special issue concerns linguistic complexity in second language development. Following Bulté and Housen (2014), we will define our basic construct as follows: linguistic complexity is a quantitative property of language units. Basically, the greater the number of components a construction has and the more levels of embedding it contains, the more complex it is. Linguistic complexity can be regarded as a valid descriptor of L2 performance, as an indicator of proficiency, and as an index of language development and progress. This paper will focus on the development of linguistic complexity in three advanced L2 learners over the course of four years as they are trying to master an English academic writing style taking a dynamic perspective. Since we assume that language is a complex dynamic system¹ (CDS) and that development is a dynamic process, we expect that different linguistic complexity subsystems develop at different levels of proficiency and that learners develop through individually owned trajectories. Therefore, different linguistic complexity measures may have to be used to trace development at different stages of proficiency (cf. Norris & Ortega, 2009). Moreover, we cannot assume a priori that even if we have the most appropriate measures, we can generalize beyond the individual. Still, because some linguistic complexity measures themselves reflect averages of a large number of instances, they may be more representative than others and may be more useful in comparing different learners. The aim of this paper is twofold: to explore which measures might capture linguistic complexity best at the advanced levels across individual trajectories in development and to see to what extent these measures are stable and useful in comparing learners.

The dynamics of linguistic complexity development in academic writing

- 2 We will assume that language is intrinsically a CDS (Ellis & Larsen-Freeman, 2009), consisting of different subsystems (morphological, lexical, phrasal, syntactic, and so on) that are interconnected and affect each other (Langacker, 2000). A consequence is that if in a particular style particular constructions are preferred, it will go at the expense of other types of constructions. In the example below, the rather dense noun phrase (a determiner, pre-modifier and a noun), which may be preferred in academic English, can also be phrased as a noun phrase that contains a post-modifier realized as a finite dependent clause. The two constructions, the first arguably more sophisticated and the second linguistically more complex, express the same notion and contain similar words but have a different syntax, different phrasal composition and different morphology. In other words, there is competition in the language system itself: if a notion is expressed with one specific type of construction, it cannot at the same time be expressed with another construction:

the participant perspective *versus* the perspective that considers the participant's point of view (Biber & Gray, 2010: 9).

- 3 We will also assume that when a learner is trying to master a second language, this is a dynamic process in which different (embedded) linguistic subsystems are noticed, practiced and mastered over time. Development, which can be defined as change (progress or regress) over time, depends on the availability of resources (De Bot, Lowie & Verspoor, 2007). With respect to language learning and use, resources include external factors such as a learner's language environment (instructed or not instructed; amount of exposure and meaningful use) and internal factors such as aptitude, current level of proficiency, and attentional capacity (De Bot & Larsen-Freeman, 2011). Resources keep development going, but they are limited and may have to be distributed over the different subsystems that grow (Van Geert, 2008), sometimes resulting in competition between those sub-systems.
- 4 The idea that different sub-systems develop at different stages was found by Verspoor *et al.* (2008), who examined an advanced learner (who will also be revisited in the current study). They found amongst other things that over time the relationship between type token ratio (lexical variety) and average sentence length (general syntactic complexity) changed and developed as the language developed. In the same vein, Caspi (2010), who examined the interactions between lexical and syntactic subsystems in four advanced L2 learners (but less advanced than those in the current study) of English over the course of one academic year, found a similar interaction between the lexicon and syntax. The lexicon becomes more complex before the syntax does, which is a clear instance of different sub-systems developing at different times. Spoelman and Verspoor (2010) traced the writing development of Finnish from the beginner to intermediate L2 level and found that as the learner's language became more complex, there was a clear competitive relation between noun phrase complexity and sentence complexity, suggesting that during one phase of development, attention was focused either on the noun phrase or the sentence. Furthermore, in a cross-sectional study across five levels of proficiency from beginning to high intermediate L2 learners, Verspoor *et al.* (2012) showed that between different proficiency levels different subsystems of the language progressed. At the lower

end differences between levels were found in specific measures in the lexicon, in the middle levels the growth was found especially in specific syntactic measures, and at the end differences were found in specific lexical measures again.

- 5 Longitudinal studies from a dynamic perspective have shown that even similar learners in similar circumstances show variability and variation in their development (cf Larsen-Freeman, 2006; Vyatkina, 2012; Bulté, 2013; Chan *et al.*, 2015). However, several studies from a dynamic perspective have also pointed to the fact that some measures show less variability or variation and may be more useful in tracing general development in linguistic complexity. Verspoor *et al.* (2012) pointed out that there were five broad measures that showed almost linear progression or regression across proficiency levels: fewer simple sentences (representing sentence complexity), fewer present tenses (representing verb phrase complexity), fewer errors (representing accuracy), increased type-token ratio (representing lexical diversity) and chunk use (total number of authentic expressions). It was argued that these particular measures showed more linear development because they were all based on averages of a great number of instances. These findings are in line with Vyatkina (2012) and Bulté (2013). Vyatkina (2012) studied the longitudinal development of the writing complexity in L2 German of college-level learners. At the group level, a linear increase of average sentence length (as a general measure of syntactic complexity) and lexical variety was found throughout the study, but at the individual level there was a considerable amount of variability. The two learners who were studied in depth showed variability and variation, but less so in average sentence length scores than in other rather specific measures. Bulté (2013), who studied the development of a great number of linguistic complexity measures over time in a group of Dutch high school learners of English longitudinally, suggests that hybrid and aggregate measures are more robust and may be better indicators of general progress over time. However, he also points out that even such measures do not develop (multi-)collinearly in this group of very similar learners. In a recent single case study, Penris and Verspoor (forthcoming) traced a great number of linguistic complexity measures over 13 years and found that one particular aggregate measure, the finite verb token ratio, correlated significantly not only with all other syntactic variables, but also with all lexical variables over time.
- 6 The current study will focus on linguistic complexity measures at the very highest levels of proficiency when L1 Dutch university students majoring in English begin to master an English academic style. According to Hinkel (2002), the greatest disparities between native and non-native writers at university level is complexity of vocabulary, specifically in the use of fixed strings, collocations and idioms, and the overuse of vague nouns such as *people* or *world*. In their English corpus investigation comparing spoken versus written language, Biber and Gray (2010) found that academic writing is characterized by structurally compressed constructions (i.e. non-clausal modifiers embedded in noun phrases). Finite dependent clauses, so commonly mentioned in the L2 literature as a good measure of linguistic complexity and development (e.g. Wolfe-Quintero *et al.*, 1998), are not found as much in academic writing and are in fact more common in 19th century fiction and in spoken language. Biber and Gray (2010) conclude that in academic writing, sentences tend to be syntactically very simple and give the following sentence from a psychology research article as an example:

This may indeed be part [of the reason [for the statistical link [between schizophrenia and membership [in the lower socioeconomic classes]]]] (Biber & Gray, 2010: 7).

- 7 The sentence is simple and contains only one main clause (of 21 words) with one finite verb (*may*), but the linguistic complexity results from prepositional phrases added on to noun phrases, which make the noun phrases longer, in this case 17 words. The fact that nominalizations occur more at the advanced level of L2 learners is noted by Byrnes (2009), who examined nominalizations in the writing of English university students studying German. She studied the writing of 14 students over three levels (levels 2-4) of German classes and found a marked increase in nominalization between levels 3 and 4. Byrnes (2009) argues that this is in line with Halliday, Matthiessen and Yang (1999), who state that linguistic development proceeds from mostly parataxis (i.e. coordination) to hypotaxis (i.e. subordination), to language with much higher levels of lexical density and more complex phrases (as opposed to more clauses). Bulté and Housen (2014) found that advanced ESL writers improved over the course of only four months and found the largest effects for mean length of finite clause (our Finite Verb Ratio) suggesting that the finite clause became internally more complex.
- 8 Of course, the longer nominalizations found in German by Byrnes (2009) could also have been due to the German language itself as it is known for its longer nominalizations. Dutch also has long nominalizations and Verspoor *et al.* (2008) suspected that L1 Dutch students would probably develop such constructions before other non-finite constructions as an L1 effect. However, as Biber and Gray (2010) point out, elaborated language with longer noun phrases is also a feature of academic English styles.
- 9 To summarize, if indeed different linguistic complexity subsystems of the language develop at different levels of proficiency, and it is assumed that at the very highest levels of proficiency in academic writing, we find fewer finite dependent clauses and more simple sentences with longer noun phrases as well as more phrasal elaboration, the question is how we can best operationalize linguistic complexity at this level. In Study 1, we trace one advanced learner to establish the development of dependent clauses and to determine which complexity measures correlate best with text ratings in order to select one syntactic and one lexical complexity measure that are most characteristic. In Study 2, both dependent clauses and the two general linguistic complexity measures selected in Study 1 are traced in two other very similar learners to test to what extent similar observations can be made. The hypotheses are as follows :
1. our advanced learners are similar in their linguistic complexity development in broad measures;
 2. our advanced learners are not similar in their linguistic complexity development in specific measures;
 3. our advanced learners have individually owned trajectories in developing linguistic complexity.

Study 1

- 10 The goal of Study 1 is to discover which measures characterize best the development of linguistic complexity in the academic writing of an advanced learner of English as a foreign language. The texts of one Dutch university student of English were collected over a period of 4 years. The texts were rated by several judges on academic proficiency and linguistic complexity. After these ratings were tested for both time and task effect, each text was coded for a great number of complexity measures and correlation tests were run

to see which measures best correlated with the text maturity ratings. Finally, two broad measures, one at the lexical and one at the syntactic level are selected to be further explored in Study 2.

Data collection

- 11 Study 1 concerns 22 texts selected from a larger pool of texts by three independent judges based on academic style written by a Dutch female university student who majored in English at the University of Groningen. She was 18 to 22 years old at the time. Given that Dutch and English are closely related languages and she had already studied English at high school, she was a rather advanced learner, estimated at the high B2 level in the Common European Framework of Reference at the beginning of her study as judged on the basis of her first writing samples. The texts concerned academic papers on different topics, ten written for courses in language, eight in literature, and three in linguistics. Seven papers were written per year in the first three years leading to the bachelor's degree. The linguistics papers were written towards the very end of the BA program as the student then specialized in linguistics and the last paper, also on a linguistic topic, was written during her fourth year for the MA program in applied linguistics. All papers were written at home, for homework without a time limit and with free access to dictionaries and other resources. From each paper a 200 word section was randomly selected after the introduction once direct quotes and references had been deleted.

Rating the texts

- 12 Six teachers in the English and Applied Linguistics department at the University of Groningen in the Netherlands, who teach English academic writing, were asked to rate each of the 22 texts. To avoid testing how the technical definition of linguistic complexity (as used in this study) corresponds to the rater's conception of complexity, the instructions were given in very general terms as follows:

The following texts were written by the same person, an English major, over the course of four years. We want to find out if her academic language has matured, in particularly her style at the linguistic level (constructions, word choice, the way ideas are combined). The texts have been put in random order. Please rate each sample holistically (You do not need look at specific language features separately but go for the all-round impression the text gives).

Assign it a score of 1 to 5

1= out of all these samples this one is at the lower end of complexity (style is rather simple).

2= somewhat better than 1, but less awkward

3= somewhat better than 2

4= slightly weaker than 5, but quite academic in style

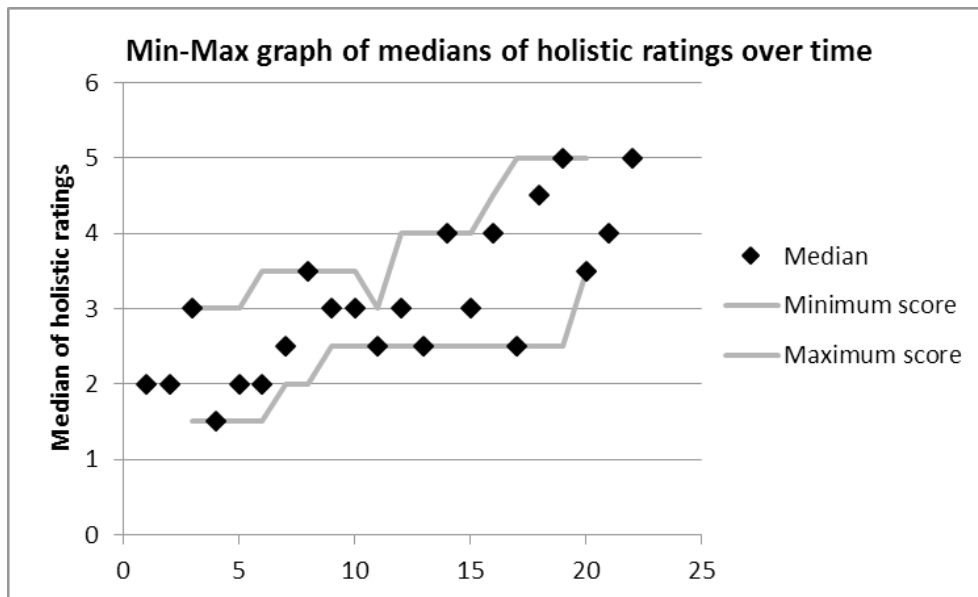
5= out of all these samples at the higher end of complexity (style is quite complex).

- 13 The inter-rater reliability with Cronbach's $\alpha = .77$ was good. Since the ratings are essentially ordinal data, the decision was made to use the median score from all six raters for subsequent analyses. For none of the texts the Cronbach's alpha was smaller than .72.

Testing for time effect

- 14 The medians of the six ratings per text correlated significantly with the sequence in which the texts were written ($r=.79$; $p<.01$), indicating that texts that were written at a later moment in time were generally rated higher. This shows a clear improvement of text quality over time. The significance of this correlation shows that the variability in these data appeared to be limited. Yet the pattern of variability in the holistic scores over time showed an interesting dynamic developmental dimension (see Fig. 1) as displayed in a min-max graph (see Van Dijk *et al.*, 2011). The bandwidth (min to max) is quite narrow for a while and then strongly increases. This suggests that around data points 11-13 there is a shift. The ratings are now generally higher, but show a greater degree of variability, which is normally associated with a developmental jump. After that moment, one would expect a trend towards relative stabilization until the next developmental jump occurs. The time series is not long enough to test this prediction but seems to develop in that direction.

Figure 1 - Min-Max graph, illustrating the changing variability of the medians over time



Testing for task effect

- 15 All texts were written in a formal, academic style, but as they were written on different topics in language, literature and linguistics classes, we checked for task effects on two broad complexity measures, the finite verb ratio and average word length. As there were only three linguistics texts that were written towards the end, they were not included in the T-test. No statistical differences were found between the finite verb ratio and the average word length in the texts written for the language ($n=10$) or literature ($n=8$) classes ($t_{FVR}(17)=-.26$; $p=.80$; $t_{AWL}(17)=1.62$; $p=.12$). Still, we cannot rule out that a degree of variability is due to task effects. This is no problem in the current dynamic analyses as the LOESS curves smooth the data over a number of instances and the algorithms in HMM modeling do not depend on single peaks and dips.

Table 1 - Testing for task effect in average word length and finite verb ratio.

Number of texts	Course	Measures	mean	sd
10	lang	FVR	6.6	6.2
		AWL	4.6	0.4
8	lit	FVR	9.9	5.4
		AWL	4.4	0.3
3	ling	FVR	4.9	2.3
		AWL	5.4	0.0

Text measures and text ratings

- 16 The medians of the ratings were subsequently used to evaluate 14 lexical and syntactic complexity measures to tap into various dimensions of complexity. Several of the measures selected were based on Verspoor *et al.* (2012), who compared English texts written by L1 Dutch learners at five different proficiency levels from beginning to high intermediate. Types of sentences and lexical diversity measures discriminated well between all five levels. The finite dependent clauses discriminated only between two subsequent levels (2-3), but were expected to decrease at the highest levels of academic writing. Noun phrase length was added because it was expected to increase at this level.
- 17 All writing samples were converted to the transcription conventions of the Codes for the Human Analysis of Transcripts (CHAT)². The 22 texts were manually coded for the syntactic complexity measures by one of the authors and carefully checked by another author. Differences were resolved through reanalyzing the sentences and discussion. The lexical measures were obtained via CLAN, except the academic words, which were obtained by means of the VP-classic analysis in Lextutor (<http://www.lex Tutor.ca/vp/eng/>). The appendix shows how one text was coded.

Table 2 - Measures, operationalization and calculation.

Measures	Operationalization	Calculation
Guiraud (Lemmatized)	The number of types divided by the square root of the number of tokens	Per text

Most frequently used words	A corpus was created of the 22 texts and frequency lists were extracted in five bands of frequency. The number of words that occur in the list of the 20% most frequently used words are counted.	Per text
Unique words	From the corpus a list of words were extracted that occurred only once. The number of words that occur on the list are counted.	Per text
Average word length	The number of characters in all words in the text are averaged.	Per text
Academic word list	The number of academic words on the Coxhead list that occur per text were counted.	Per text
Finite Verb Ratio (FVR)	The total number of words divided by the total number of finite verbs.	Per text
Simple sentence ³	One main clause (including non-finite dependent clauses) with subject and finite verb.	% of sentence types
Compound sentence	Two main clauses, each with its own subject and finite verb.	% of sentence types
Complex sentence	One main clause and one or more finite dependent clauses.	% of sentence types
Compound-complex sentence	A sentence containing both subordinate and coordinate elements (each with a finite verb)	% of sentence types
Finite adverbial clause	A clause with a finite verb functioning as adverbial	% of clause types
Finite nominal clause	A clause with a finite verb functioning as nominal	% of clause types
Finite relative clause	A clause with a finite verb functioning as post-modifier	% of clause types
Non-finite constructions	An element with a non-finite verb functioning as part of the verb phrase, adverbial, nominal or post-modifier	% of clause types
Noun Phrase length	A noun with all its modifiers, including finite and non-finite post-modifiers in embedded noun phrases (which were not calculated separately)	Average length in text

- 18 To investigate the extent to which the measures coincide with the holistic assessments of the experienced raters, we analyzed the correlations of the median of the ratings with each of the features we measured. (See Table 2. For the scores of the measures per text, please see Appendix 2)

Table 3 - Correlations of measures with the medians of the academic style ratings of the texts.

LEXICAL	Correlation		SYNTACTIC	Correlation	
Guiraud	.17		Finite verb Ratio	.59	**
Mostfreq	.08		Average sentence length	.65	**
Unique Words	.64	**	Simple	.35	
AveWord length	.66	**	Compound	-.043	*
Academic Word List	.69	**	Compound-complex	.092	
			Complex	-.2	
			Adverbial finite clauses	-.29	
			Nominal finite clauses	.21	
			Relative finite clauses	.25	
			Nonfinite constructions	.13	
			Total clauses	.13	
			Average NP length	.47	*

Level of significance is indicated by ** $p < .01$ and * $p < .05$

- 19 As far as lexical measures are concerned, the lexical diversity measures (Guiraud index; Guiraud, 1954) did not show a high correlation. The most frequently used words did not show a negative correlation. The three lexical measures that reflected degrees of lexical complexity - unique words, average word length, and words on the academic word list - showed strong and significant correlations. Unique words, academic words and average word length all correlate highly and we assume that they partially tap into the same construct as less frequent words tend to be more academic and longer. We decided that average word length is the best lexical linguistic complexity measure: it meets the defining criteria of linguistic complexity as it is a quantitative measure and it is the simplest to establish in English and in other orthographic languages. For unique words, the researcher first has to create a corpus of the texts under investigation and use software to select the unique words. To calculate the number of academic words, the process also takes extra steps and lists are available only for the most commonly studied languages.
- 20 As far as syntactic constructions are concerned, there are four measures that correlate strongly, with three that measure partially overlapping constructs (finite verb ratio, average noun phrase length and average sentence length). There is not a strong correlation with any of the measures that target other specific constructions such as the different types of finite dependent clauses, and not with the total number of finite and non-finite dependent clauses, which could also be considered an aggregated measure.

- 21 There is a negative correlation with compound constructions. The number of compound sentences is a specific sub-category of sentence types and is best pursued within its own subsystem of sentence types, which we did not focus on in the current study. As expected, noun phrase length seems to be a good general linguistic complexity measure at this stage of development, especially as it concerns a complexity measure that is supposed to emerge at this advanced level. Average sentence length and FVR are also syntactic measures that correlate well with development over time. Average sentence length has proven to be one of the most robust measures in this respect across all stages of development (see also Bulte, 2013; Bulté & Housen, 2014 and Vyatkina, 2012). Like average sentence length, the FVR does not target specific constructions but averages out over a great many instances. However, unlike average sentence length, FVR also reflects internal complexity at the clause level - including the length of noun phrases and other non-finite constructions (cf. Norris and Ortega, 2009). In addition, Penris and Verspoor (forthcoming) found that FVR is the only measure that correlated significantly with all other syntactic measures and all lexical measures over time. The FVR is selected as the preferred measure to compare the development of general linguistic complexity at advanced stages in three learners for three reasons: (1) it taps into the same constructs as average sentence length and average noun phrase length (2) the FVR shows more internal complexity than average sentence length and (3) it is easier to establish than the average length of noun phrases.

Study 2

- 22 The aim of Study 2 is to see to whether the two linguistic complexity measures found in study 1 (AWL and FVR) are more generalizable in tracing learner development than specific measures (dependent clauses), and if so, to what extent three similar learners have similar trajectories in their development. The hypotheses are :
1. broad measures will show relatively less variability than specific measures;
 2. the three learners are similar in their development of broad measures;
 3. specific measures will show relatively high levels of both variability;
 4. the three learners are different in their development of specific measures.

Data collection and selection

- 23 Texts written for various courses of the Dutch female university students (students B and C) who majored in English at the University of Groningen were collected. Students B and C were very similar to student A in age and proficiency at the beginning of their study. They took the same course of study as student A a few years later. Therefore, students B and C are more similar to each other because they took the exact same classes at the same time, and their texts often concern the same assignments. However, because not all assignments were done by both students, there is not a complete overlap of texts written over time. The procedures in selecting and coding texts were the same as for student A.

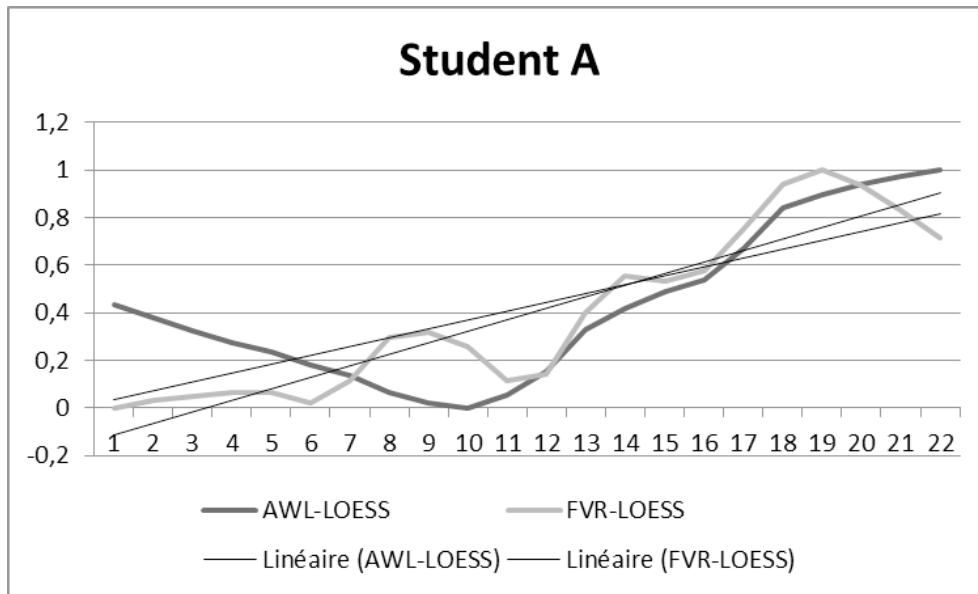
Variables to be traced

- 24 To trace general development in linguistic complexity, the two broad measures as established in Study 1 were used: AWL and FVR. For the more specific constructions to be traced, we focus on complexity measures that are expected to decline during the course of development at the most advanced stages, but which are often mentioned in the L2 literature (e.g. Wolfe-Quintero *et al.*, 1998) as good indicators of proficiency: the use of dependent clauses such as finite adverbial (FA), nominal (FN), and relative (FR) clauses and non-finite (NF) clauses, none of which showed a strong correlation over time in Study 1.
- 25 Also, in Verspoor *et al.* (2012) the number of finite dependent clauses was found to be a good discriminator at the lower levels of L2 proficiency of L1 Dutch learners. Here we want to test our assumption that at the highest levels of proficiency these types of constructions are no longer good indicators of development as learners start using more non-finite clauses and longer noun phrases. Moreover, we want to see whether indeed the increase or decrease of such specific constructions is more idiosyncratic and therefore less useful in measuring overall linguistic complexity development.

Inspecting the data with visualization techniques

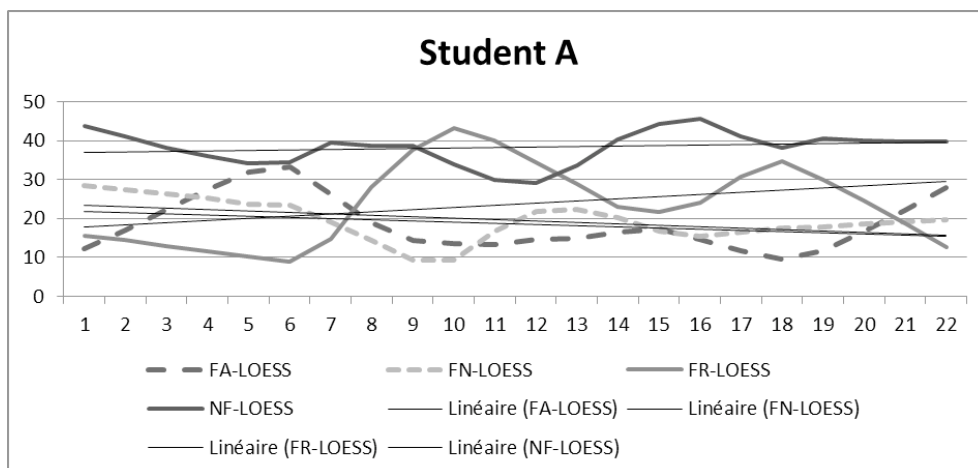
- 26 As pointed out by Van Dijk *et al.* (2011), visual inspection is first needed to get a feel for the data and to explore changes that can later be tested or explored in modelling. Since the raw data does not show clear changes over time due to the high variability within the data sets, the data is usually smoothed or trended. In the current study we use LOESS smoothing with windows of 8 or 12 data points, depending on the number of data points in the data set.
- 27 First we examine student A in detail and then compare her development with that of the other two students. Figure 2 shows student A's developmental trajectories of AWL and FVR, the two measures that correlated highly with academic style. As expected based on Study 1, the linear trend lines indicate that both the AWL and FVR increase over time. In order to capture the general tendencies and the way they relate to each other, a smoothing function, in this case a Loess function with a window of 8 data points, has been used. The figure suggests that AWL and FVR competed with each other early on in this data set, but later on, at around point 12 or 13 there is a shift: the two variables begin to develop together. In other words, early on the learner focused on either word complexity or sentence complexity, but later the learner was able to focus on them simultaneously when the subsystems became coordinated.

Figure 2 - Trajectories over time (x axes) of student A (with LOESS curves with window of 8 data points) of average word length (AWL) and finite verb ratio (FVR).



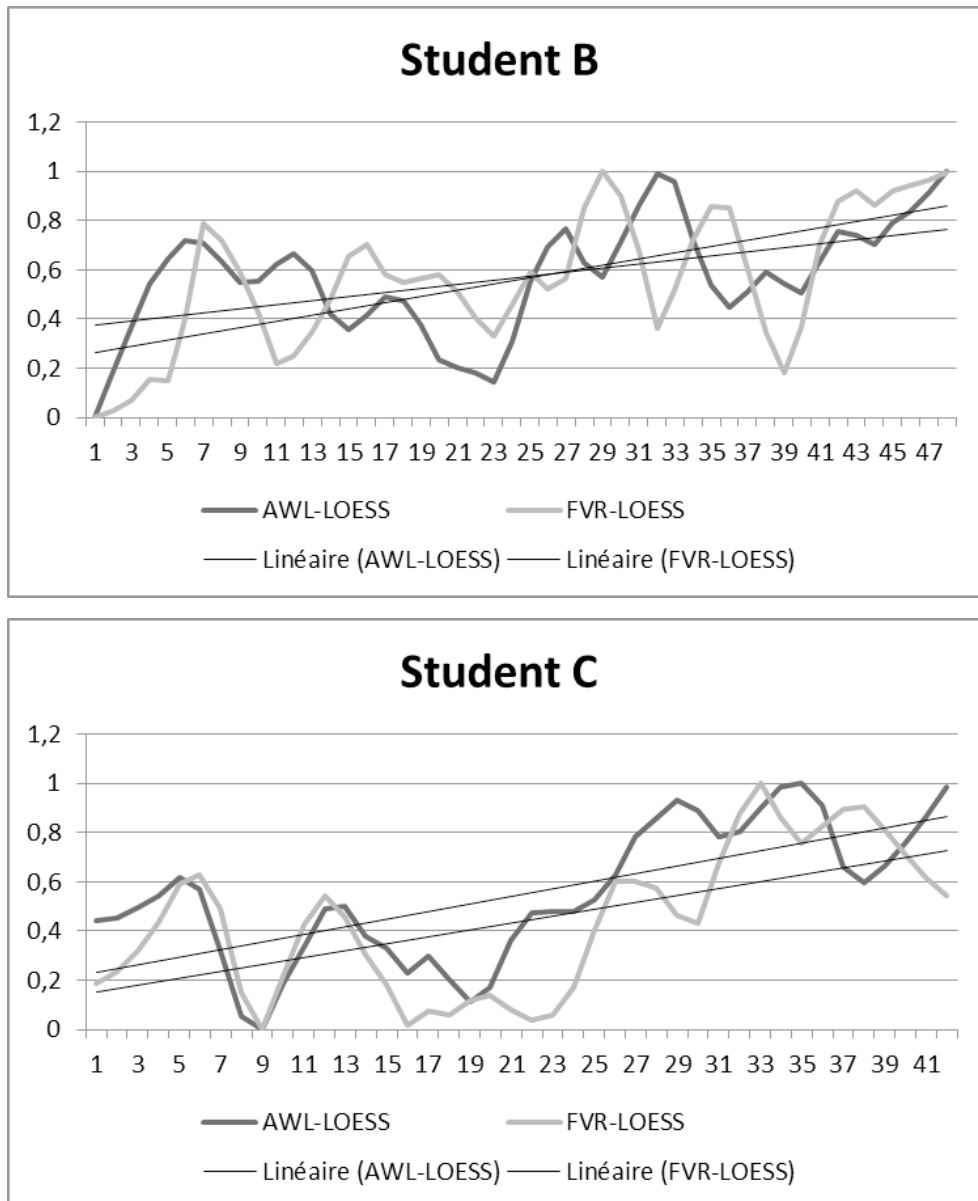
- 28 Figure 3 shows student A's development in the finite and non-finite dependent clauses. Linear trend lines show that FR and NF increase somewhat, but FA and FN decrease. The LOESS curves show a wavy pattern as suggested by Siegler (1996), where new forms emerge, but other forms do not disappear. The first small hump is seen for finite adverbial clauses. The second hump suggests an increase in relative clauses at the expense of all the other types of clauses and the third hump shows a small increase in nominal clauses. The fourth hump is for non-finite clauses, with finite adverbial and nominal clauses decreasing, and a small hump again for finite relative clauses.

Figure 3 - Trajectories over time (x axes) of student A (with LOESS curves with a window of 8 data points and linear trendlines) of finite adverbial (FA), finite nominal (FN), finite relative (FR) and non-finite (NF) dependent clauses.



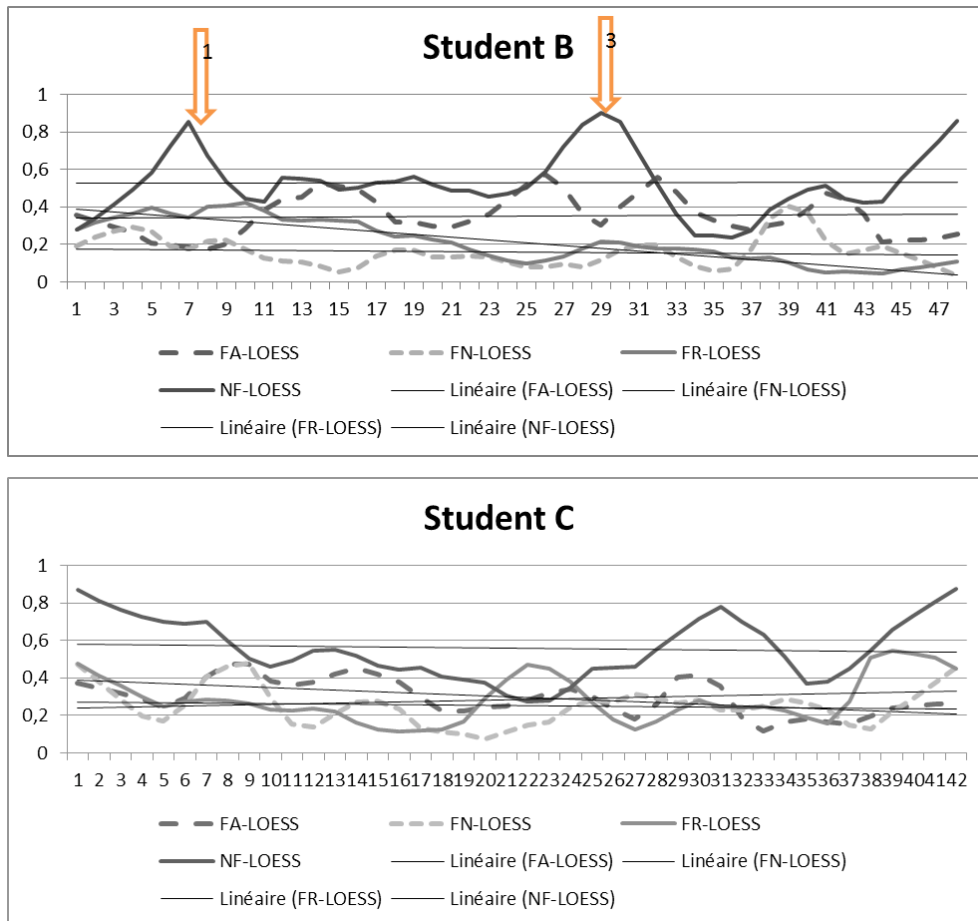
- 29 In Figures 4 and 5, we explore the development of students B and C. Given that we have more data points for these students, we use a LOESS with a window of 12 data points. Figure 4 shows the AWL and FVR and Figure 5 the dependent clauses.

Figure 4 - Trajectories over time (x axes) of students B and C (with LOESS curves with a window of 12 data points and linear trend lines) of finite verb ratio (FVR) and average word length (AWL).



- 30 Just as for student A, the linear trendlines show an increase in both the AWL and FVR for both B and C. The LOESS curves, however, show some differences. Whereas student A clearly showed some competition between FVR and AWL (Fig. 3) early on and support later on, student B (Fig. 5) shows a much more variable pattern with moments of competition and support until the end. Student C seems to have a supportive relation early on as the two variables go up and down in unison and later on show an increase, but with more competition.

Figure 5 - Trajectories over time (x axes) of students B and C (with LOESS curves with a window of 12 data points and linear trendlines) of finite adverbial (FA), finite nominal (FN), finite relative (FR) and non-finite (NF) dependent clauses.



- 31 Figure 5 shows that students B and C do not have the same patterns in the development of dependent clauses as student A. The linear trendlines differ in that there are no clear increases for FR and NF. The LOESS curves for student B show an early hump in non-finite clauses, a hump in relative clauses and then again in non-finite clauses. Student C shows small humps in finite nominal and adverbial clauses at the same time, an increase in relative clauses with a clear hump and, finally, a clear hump in non-finite clauses.

Analyses

- 32 Visual inspection of the graphs has given us the impression that the students are somewhat similar in the development of broad measures but quite different in the development of the more specific measures. The aim of the current analysis is to test statistically to see to what extent these Dutch learners of academic English have similar or different trajectories in development. We do so by focusing on a shift (as seen in both Figures 1 and 2), which indicate a moment of self-organization in the learners' language system. In other words, we will try to locate statistically the moment that the interaction among the measures changes and takes on a new configuration. Such moments of reorganization indicate the beginning of a new stage (not in the usual L2 developmental sense but in the sense of a new configuration of variables). The model will also provide us

with the means of each measure at each stage before and after the shift so that we can deduce the direction in which each variable moved.

- 33 In Figures 1 and 2 the shifts were quite visible because smoothed trajectories of no more than two variables were involved. But for students B and C such clear shifts were not really observable in figure 4. Moreover, with six different measures developing over time, it is impossible to visually pinpoint such changes exactly. To solve this problem, we use a dynamic model often used in the field of speech recognition (e.g. Novotney, Schwartz & Ma, 2009; Park & Glass, 2008; Zhang & Glass, 2009), called an unsupervised Hidden Markov Model (HMM). In such a probabilistic model, the raw data strings of all the measures are analyzed to detect patterns of change among the measures and then the model locates the data point where a “shift” in the complex system occurs by indicating the boundaries of stages. (For more detail see Chan *et al.*, 2015.)
- 34 With the HMM, it is possible to try out different numbers of emerging stages and explore which number of stages best reveals changes over time; however, the current study postulated only two stages for one main reason. For student A, we had only 22 data points. In HMM, there is a minimum number of data points for a stage required: the number of data points in each stage must ideally be two to three times the number of observed measures. As we worked with six measures, there should be a minimum of 12 to 18 data points in each stage, but for student A we only have around 12, so positing more than two stages was statistically not feasible.
- 35 The HMM software used is based on Chan and Lee (2013), one of whose authors programmed our software in Perl for a Linux environment specifically for the current study. The model was initialized with a linear structure where state 1 can transit to state 1 or state 2, state 2 can transit to state 2. The parameters were optimized according to the forward-backward algorithm to find the best model (Rabiner, 1989). The data training with the algorithm was discontinued when the model converged, i.e. when further iteration resulted in no significant improvement of the model and the optimal stages had been probabilistically determined. After obtaining the set of parameters, the single best stage sequence was calculated with the Viterbi algorithm (Ryan & Nudd, 1993).
- 36 Using the HMM, we trained two learning stages for each participant based on the following six variables: average word length (AWL), finite verb ratio (FVR), finite adverbial clauses (FA), finite nominal clauses (FN), finite relative clauses (FR), and non-finite constructions functioning as adverbial, nominal, or post modifying clause (NF).
- 37 The model output informs us of where stage 1 ends and stage 2 begins, and thereby locates the point at which a shift among the variables takes place. It also provides the means of the variables in each stage.

Results

- 38 Table 4 shows the output of the HMM for student A. In the third column, it shows the beginning and end data points for stage 1 and stage 2. The shift is located between the two stages. As expected for student A, based on Figs. 1 and 2, the shift took place between data point 13 and 14 (of 22), which is around the mid-point. AWL increased from 4.36 to 5.11 and FRV from 9.34 to 13.19. As assumed from the visual inspection (figure 3) the FN and FA decreased somewhat, but the FR and NF increased.

Table 4 - Output of the HMM analysis indicating between which data points a shift among the variables takes place and the means of the variables before and after the shift for Student A.

Student	Stage	Data points	AWL	FVR	FN	FA	FR	NF
A	1	1-13	4.36	9.34	.34	.34	.36	.47
A	2	14-22	5.11	13.19	.30	.26	.42	.67

- 39 Table 5 shows the output of the HMM for students B and C, who took the same courses at the same time and whose assignments were written at the same time. For student C there was also a shift around the midpoint between data point 22 and 23 (of 42), but for student B the shift took place later between data points 31 and 32 (of 42).

Table 5 - Output of the HMM analysis indicating between which data points a shift among the variables takes place and the means of the variables before and after the shift for Students B and C.

Student	Stage	Data points	AWL	FVR	FN	FA	FR	NF
B	1	1-31	4.70	10.02	.24	.34	.27	.52
B	2	32-48	4.93	11.83	.26	.24	.31	.60
C	1	1-22	4.69	10.90	.16	.37	.28	.57
C	2	23-42	4.82	11.60	.17	.32	.10	.45

- 40 As expected, the two general measures (AWL and FVR) increased for students B and C over time, but less so than for student A. Like student A, student B decreased in her use of finite adverbial clauses but remained about the same in the use of nominal clauses, and increased in finite relative and non-finite clauses. For student C, however, both the FR and NF decreased, the FN remained about the same, and FA decreased. In other words, student C is clearly different from the other two in her development of these specific measures. Table 6, summarizes the directions of change.

Table 6 - Direction of change between stage 1 and 2 Students A, B and C.

Student	AWL	FVR	FN	FA	FR	NF
A	+	+	-	-	+	+
B	+	+	+	-	+	+
C	+	+	+	-	-	-

Discussion

- 41 In our discussion we will explore the findings of study 1 and 2 separately. Our conclusion will deal with the implications for both research and teacher expectations.

Study 1

- 42 In summary, there are several interesting and useful findings in our study of one advanced writer's texts written over a period of four years. Not surprisingly, there was a strong correlation with time as our raters considered the later texts more "academic" in style on the whole. The min-max graph (Figure 1) showed variability of academic style over time. The first three texts were not the weakest; that was the fourth one. But after data point 11, the median was not below 2.5 anymore and towards the end the scores were mainly 4's and 5's. The variability could have been related to the task complexity, though it is also possible that the learner was paying more attention to or struggling with another aspect of the writing task such as content or organization at the text level, which would go at the expense of the types of linguistic constructions she used at that time. Whatever the cause, the variability shows that the learner's language system was not settled and still developing, perfectly in line with a CDS view.
- 43 The other interesting point is that according to the raters, after data point 11, there was a shift in style, suggesting some reorganization of the system that allowed the median scores to remain above 2.5 from that moment on. This point is also clearly identified in the LOESS figure 1 and by the HMM analysis in Study 2.
- 44 Subsequently, the texts were coded for a number of linguistic complexity measures known from Verspoor *et al.* (2012) to correlate well with development at the lower proficiency level and some expected to discriminate at the higher levels. The measures were correlated with the medians of the ratings. The ones that correlate positively and significantly are the ones that characterize writing development best at this level for this L2 writer.
- 45 These correlation analyses showed some surprising findings. Whereas in Verspoor *et al.* (2012) lexical measures pertaining to diversity (Guiraud index; Guiraud, 1954) were accurate indicators of progress in lower proficiency levels, lexical diversity measures showed a relatively low correlation with the rated texts in the current study. This is in line with Bulté and Housen (2014), who found that a group of ESL students at a high intermediate level did not show change in any lexical diversity measures over a 15 week time span.
- 46 Several lexical complexity measures, however, do show a high correlation with the rated texts: academic words, unique words and average word length. Of course, the use of academic words is expected to increase at the academic level. Unique words are interesting because they failed to discriminate between texts at the lower levels of proficiency (Verspoor *et al.*, 2012). Average word length is not surprising: Jarvis, Bikowski, and Ferris (2003) found that in their cluster analysis of twenty-one linguistic features related to linguistic complexity, average word length was one of the seven measures that were good predictors of essay complexity. Also Penris and Verspoor (submitted) found that average word length correlated highly with academic words. Word length is also a

good approximation of what has been referred to as lexical sophistication (Wolfe-Quintero *et al.*, 1998), because low frequency words tend to be longer than many high-frequency words. This finding for the high proficiency learner clearly contrast with both Verspoor *et al.* (2012) and Bulté (2013), who found that average word length was not a good discriminator at the lower proficiency levels they investigated. To summarize, several lexical measures that discriminate well between L2 English texts written by L1 Dutch students at lower levels of proficiency do not do so at the higher level of students with the same L1 background, and vice versa. The results convincingly show that average word length is the best broad measure to use in tracing general development in writing at the high proficiency level as it does not target a specific construction but averages out over a great number of instances and is therefore more likely to show fewer extreme peaks and dips than measures that target specific constructions.

- 47 Among the syntactic measures, the ones that discriminated well at the lower proficiency levels (simple sentences versus complex sentences) did not correlate with text ratings at all at this higher level, but the simple sentences did have the stronger correlation of the different types ($r=.35$), which suggests that indeed the sentences at the higher levels have less subordination, and complexification occurs by means of other types of constructions. Among the ones that show significant correlations, there are two related measures: the finite verb ratio and average sentence length.
- 48 The effect of average sentence length (or T-unit length) is in line with many previous studies and is one of the most robust complexity measures according to Wolfe-Quintero *et al.* (1998: 97 - 98). Penris and Verspoor (forthcoming) also showed that it continually increased over time. However, at the lower proficiency levels the average T-unit length was found to be only a moderately strong discriminator by Verspoor *et al.* (2012): there were significant differences between levels 2 - 4, and 3 - 5. In contrast, Bulté (2013), in a longitudinal study with students of about the same level as those studied by Verspoor *et al.* (2012), found the T-unit length to be one of the better discriminators. Also, Bulté and Housen (2014) found it to be a good discriminator. Assuming that average sentence length and T-unit length overlap in the complexity dimension they tap, we may conclude that these general length measures do well to trace development at a wide spectrum of proficiency levels.
- 49 The finite verb ratio was used by Verspoor *et al.* (2008) and Norris and Ortega (2009), who referred to it as Mean Length of Clause length, argued that this measure better shows internal complexity of a sentence than average sentence length. Recently, Bulté and Housen (2014) found that for their learners at a somewhat lower level the finite clause length also increased significantly over a 15 week period. Penris and Verspoor (submitted) found that it was the one measure that correlated significantly with the most (almost all) other syntactic and lexical complexity measures.
- 50 The significant negative correlation with compound sentences is striking in the light of the absence of an effect for other measures of sentence complexity. Apparently, the increasing length of sentences is not achieved by moving from simple sentences to compound sentences, but rather to either complex sentences or simple sentences with longer non-finite constructions as was shown in Fig. 2. The latter is in agreement with the significant effect of the number of words per finite verb (FVR) and shown earlier by Verspoor and Van Dijk (2011). It is also in line with Verspoor *et al.* (2012) and Bulté (2013), neither of whom found significant correlations between compound sentences and level or time. However, Bulté and Housen (2014), whose students were more advanced than those

of Verspoor *et al.* (2012) and Bulté (2013), but less advanced than the students in the current study with different L1's, did find a significant increase over time in the number of compound constructions. Apparently, at some level of proficiency or in different learner groups, compound constructions are used more frequently. In the current study, though, the use of compound sentences decreased significantly over time.

- 51 Finally, as one would expect at the more academic level, the average NP length had a relatively strong correlation with the text ratings. This is in line with Biber and Gray (2010), who argue that one of the features of academic writing is that it is characterized by nominalizations. Both Byrnes (2009) and Bulté and Housen (2014) found higher degrees of nominalizations at the more advanced levels.
- 52 It was argued that the FVR would be the best broad measure to use in tracing general development in writing at the high proficiency level as it does not target a specific construction but averages out over a great number of instances. Unlike average sentence length, though, it also accounts for internal complexification of clauses.
- 53 To summarize, it is important to note that some of the measures that correlated well with development at this advanced level are quite different from those found by Bulté (2013) and Verspoor *et al.* (2012), who both investigated students with a similar L1 background at lower proficiency levels. In other words, we need different linguistic complexity measures to trace development at different levels of proficiency.

Study 2

- 54 In Study 2, the developmental trajectories of six measures were traced in three advanced Dutch learners of English during the same course of studies. student A's data was taken as a starting point and students B and C, who took exactly the same courses and wrote the same assignments at around the same time, were compared to Student A and each other. The main question was to see if there were common patterns in the developmental trajectories of similar learners. The two broad complexity measures (FVR and AWL) were expected to show similar patterns of development and the four specific complexity measures (finite adverbial, nominal and relative clauses and non-finite constructions) were expected to show more idiosyncratic and variable patterns.
- 55 The visual inspection of Student A's data suggested that there was a clear shift in writing style (Fig. 1) and in the interaction between lexicon (AWL) and syntactic constructions (FVR) around data point 13. The visual comparison of the other two learners in Figs. 4 and 5 suggested that the three learners had few clear similarities in their development, except that both AWL and FVR seemed to increase over time.
- 56 To statistically test for similarities and differences among the learners' trajectories we used a Hidden Markov Model, a probabilistic computer model that explores the dynamic interactions among several variables and locates stages and provides information on where in the process a shift in the relation among the six variables took place. It provides the boundaries of each stage and the mean of the variables during the stage. Comparing the means of each student in stage 1 and 2, we could see the direction of change over time: an increase or decrease. As far as shifts were concerned, students A and B were quite similar as they changed around halfway the data collection period. Surprisingly, student C (who took the exact same courses at the same time as student B) showed a shift much later in the data collection period. These differences in stages may be due to

differing initial conditions (e.g. proficiency level at the beginning of the study) or other individual differences such as learning or writing style.

- 57 As far as development in general complexity at both the lexical and syntactic level is concerned, the three students were somewhat similar. The means of the variables in each stage showed that for all three learners the FVR and AWL increased, suggesting that over time all three students used longer (more sophisticated and internally complex) words and more elaborated language with fewer finite dependent clauses, longer noun phrases and other non-finite constructions.
- 58 As far as the development of specific syntactic constructions was concerned, the three students were quite different. First of all, based again on student A's line graphs, we had expected that there was first a competition between the lexicon (AWL) and syntax (FVR), which would then change to a supportive relation. The analysis indeed confirmed this for student A, but not for students B and C. As far as the dependent clauses were concerned, all three showed a decline in the use of finite adverbial clauses, but different patterns in the other clauses. Each of these students had her own developmental path in the use of these rather specific constructions and perhaps instead of questioning this, we should accept that learners own their individual learning paths. This is in line with quite a few dynamic studies so far that show that at least one learner in the small group studied is clearly different from the others: Caspi (2010) had 4 learners, Bassano and Van Geert (2007) had 4 children, Larsen-Freeman (2006) had 4 learners, and Cancino, Rosansky and Schumann (1975) had six learners. Vyatkina (2012) found that two individuals who were followed in detail followed overall group trends but differed in the specific paths they took. However, the most telling example of variation being the norm is probably Chan *et al.* (2015), where identical twins with the same teacher and the same amount of exposure to English were traced on identical tasks for one year on both spoken and written English and showed different patterns of development even in very general measures.

Conclusion

- 59 The current study looked at the development of complexity in L2 writing from a CDS perspective. The perspective holds that language as a system is a CDS, which means that particular notions can be expressed with different constructions and as a consequence, if in a particular style, particular constructions such as longer noun phrases are preferred, it will go at the expense of other types of constructions such as finite dependent clauses. In other words, at different levels of complexity in the language, different types of constructions are used. For example, several authors (e.g. Biber & Gray, 2010; Byrnes, 2009) have pointed out that an academic writing style is different from a non-academic advanced style in that it uses more nominalizations, which are likely to include phrasal elaboration.
- 60 A CDS perspective also holds that development is a non-linear, dynamic, emergent process, where each L2 writer has to find his or her own path, often through trial and error, overusing some constructions while practicing, which may go at the expense of other constructions. However, once the learner has mastered and automated some subsystems, there are moments of self-organization when different subsystems that previously competed become coordinated and a clear shift takes place, changing the constellation of the whole system. In other words, at different moments in the developmental process (at different proficiency levels) the very make-up of the learners

L2 interlanguage system is different. For example, subordination with finite clauses has proven to be a robust measure in L2 development at many levels of proficiency (e.g. Wolfe-Quintero *et al.*, 1998), but as shown in the current paper, not in academic writing, which is characterized by nominalizations and other non-finite constructions.

- 61 Because in our approach we assume that each learner will have to find his or her own developmental path, we cannot expect individuals to behave the same. We cannot generalize from the group to the individual. If several similar individuals show similar patterns, we may tentatively assume that we have found a rather general pattern for at least these similar learners. Indeed, the current analysis has shown that the three individuals are similar in that their texts have become more complex over time, both in their lexicon and in their syntax. However, these similarities only occurred when we used the average word length and finite verb ratio as measures, both very broad, all-encompassing measures that are generalizations in themselves of many separate constructions and even these general measures do not show a linear development. Moreover, the way these variables interact over time is not the same for the learners. When we looked at the development of specific constructions, in the current study the development of finite and nonfinite dependent clauses, we saw few similarities among the three similar learners. In other words, each had her own trajectory in achieving a more academic writing style, confirming CDS principles: variation is the norm. And although developmental stages may be discerned based on group means, the developmental pattern may not be representative for any of the individuals within the group. If we are interested in how complexity develops over time, longitudinal case studies are more informative and more appropriate than group studies.
- 62 Our findings have implications for research and teaching. For research, this study has shown clearly that different levels of proficiency may demand different linguistic measures to accurately measure development. Even if we only want a global impression of development by means of global measures, we need to be aware that different ones may be called for at different levels of proficiency. Future research, hopefully in replication studies, will have to establish which measures are useful not only for different proficiency levels but also for different L1 groups, as an L1 is an initial condition and may have strong effects on L2 development as for example Marakami (2013) has shown. Moreover, even broad, general measures are quite variable over time and call for repeated testing rather than simple one-off samples in group comparisons of pre-post design studies.
- 63 For researchers who want to trace development from a dynamic perspective, this study has shown that it is interesting to trace specific constructions such as the finite and non-finite dependent clauses to see how they wax and wane, but that especially in these specific constructions each individual will find his or her own path. A corollary of this finding is that such specific constructions are not useful either in group comparison or pre-post design studies.
- 64 Finally, this study has presented a rather new statistical tool in complexity research from a dynamic perspective: the HMM, which can automatically pinpoint the moment or moments in which a shift in the configuration of a number of different variables has taken place.
- 65 For English as an L2 writing teachers, this study has shown that even students who have an L1 that is quite similar to the L2 in many respects and who are quite advanced when they enter university, still need at least three years of a great amount of input and

practice to develop a linguistic style that can be considered academic (as indicated by the expert ratings). Moreover, besides development in linguistic constructions, the student must develop an academic style at the other levels: at the discourse level to contextualize the writing in their specific fields and at the textual level in that the writing should be clearly organized and structured and contain appropriate conventionalized ways of saying things. Taking a CDS perspective, we would expect that during development, students may pay more attention to one level than to the other, and will show trade-off effects in these subsystems too. In other words, the development of an academic writing style is a dynamic process in which many different subsystems will change and interact and need to become fully coordinated and synchronized before an acceptable level of proficiency is reached.

- 66 In understanding this, the most appropriate message that teachers can take from this study is to not expect miracles, to acknowledge the relevance of instability and variability in development over time, and to have patience with their pupils!

BIBLIOGRAPHIE

- Biber, D. & Gray, B. (2010). « Challenging stereotypes about academic writing: Complexity, elaboration, explicitness ». *Journal of English for Academic Purposes*, vol. 9 n° 1. pp. 2-20.
- Bulté, B. (2013). *The development of complexity in second language acquisition: A dynamic systems approach*. (Unpublished Doctoral Dissertation). University of Brussels.
- Bulté, B. & Housen, A. (2014). « Conceptualizing and measuring short-term changes in L2 writing complexity ». *Journal of Second Language Writing*, vol. 26. pp. 42-65.
- Byrnes, H. (2009). « Emergent L2 German writing ability in a curricular context: A longitudinal study of grammatical metaphor ». *Linguistics and Education*, vol. 20, n° 1. pp. 50-66.
- Cancino, H., Rosansky, E. J. & Schumann, J. H. (1975). « The acquisition of the English auxiliary by native Spanish speakers ». *TESOL quarterly*, vol. 9, n° 4. pp. 421-430.
- Caspi, T. (2010). *A dynamic perspective on second language development*. (Unpublished Doctoral Dissertation). University of Groningen.
- Chan, H., Verspoor, M. & Vahtrick, L. (2015). « Dynamic Development in Speaking versus Writing in identical Twins ». *Language Learning*, vol. 65, n° 2. pp. 298-325.
- Chan, C. & Lee, L. (2013). « Model-based unsupervised spoken term detection with spoken queries ». *IEEE Transactions on Audio, Speech & Language Processing*, vol. 21, n° 7. pp. 1330-1342.
- De Bot, K. & Larsen-Freeman, D. (2011). « Researching second language development from a dynamic systems theory perspective ». In Verspoor, M., de Bot, K. & Lowie, W. (eds). *A dynamic approach to second language development: Methods and techniques*. Amsterdam: John Benjamins.
- De Bot, K., Lowie, W. & Verspoor, M. (2007). « A dynamic system theory approach to second language acquisition ». *Bilingualism: Language and Cognition*, vol. 10, n° 1. pp. 7-21. Doi:10.1017/S1366728906002732

- Ellis, N. & Larsen-Freeman, D. (2009). « Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage ». *Language Learning*, vol. 59, N° 1. pp. 90-125. Doi:<http://dx.doi.org/10.1111/j.1467-9922.2009.00537.x>
- Guiraud, P. (1954). *Les caractères statistiques du vocabulaire*. Paris : Presses universitaires de France.
- Halliday, M. A. K., Matthiessen, C. M. & Yang, X. (1999). *Construing experience through meaning: A language-based approach to cognition*. London: Cassell.
- Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Jarvis, S., Grant, L., Bikowski, D. & Ferris, D. (2003). « Exploring multiple profiles of highly rated learner compositions ». *Journal of Second Language Writing*, vol. 12, n° 4. pp. 377-403.
- Langacker, R. W. (2000). « A dynamic usage-based model ». In Barlow, M. & Kemmer, S. (Eds.). *Usage-based models of language*. Palo Alto, CA: CSLI.
- 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*, vol. 27, n° 4. pp. 590-619. doi:10.1093/applin/aml029
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Murakami, A. (2013). *Individual Variation and the Role of L1 in the L2 Development of English Grammatical morphemes: Insights From Learner Corpora*. (Unpublished Doctoral Dissertation). Cambridge University.
- Norris, J. & Ortega, L. (2009). « Towards an organic approach to investigating CAF in instructed SLA ». *Applied Linguistics*, vol. 30, n° 4. pp. 555-578. doi:10.1093/applin/amp044
- Novotney, S., Schwartz, R. & Ma, J. (2009). « Unsupervised acoustic and language model training with small amounts of labelled data ». in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-09)*. Taipei, Taiwan.
- Park, A. S. & Glass, J. R. (2008). « Unsupervised pattern discovery in speech ». *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, n° 1. pp. 186-197. doi:10.1109/TASL.2007.909282
- Penris, W. & Verspoor, M. (forthcoming 2017). « Academic Writing Development: a complex, dynamic process ». In Pfenninger, S. E. & Navracics, J. (eds.). *Future research directions for applied linguistics*. Bristol: Multilingual Matters.
- Rabiner, L. R. (1989). « A tutorial on hidden markov models and selected applications in speech recognition ». In *Proceedings of the IEEE*, vol. 77, n° 2. pp. 257-286.
- Ryan, M. S. & Nudd, G. R. (1993). *The viterbi algorithm*. Department of Computer Science research report 238, University of Warwick. (Unpublished).
- Siegler, R.S. (1996). *Emerging minds: The process of change in children's thinking*. New York, NY: Oxford University Press.
- Spoelman, M., & Verspoor, M. (2010). « Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish ». *Applied Linguistics*, vol. 31, n° 4. pp. 532-553. doi:10.1093/applin/amq001
- Van Dijk, M., Verspoor, M. & Lowie, W. (2011). « Variability and DST ». In Verspoor, M., De Bot, K. & Lowie, W. (eds.). *A dynamic approach to second language development*. Amsterdam: John Benjamins.

- Van Geert, P. (2008). « The dynamic systems approach in the study of L1 and L2 acquisition: An introduction ». *The Modern Language Journal*, vol. 92, n° 2. pp. 179-199.
- Verspoor, M., Lowie, W. & Van Dijk, M. (2008). « Variability in second language development from a dynamic systems perspective ». *The Modern Language Journal*, vol. 92, n° 2. pp. 214-231. doi:10.1111/j.1540-4781.2008.00715.x
- Verspoor, M. & Sauter, K. (2000). *English Sentence Analysis: An introductory course*. Amsterdam: John Benjamins.
- Verspoor, M., Schmid, M. S. & Xu, X. (2012). « A dynamic usage based perspective on L2 writing ». *Journal of Second Language Writing*, vol. 21, n° 3. pp. 239-263. doi:10.1016/j.jslw.2012.03.007
- Verspoor, M. & van Dijk, M. (2011). « Variability and DST ». In Verspoor, M., De Bot, K. & Lowie, W. (eds.). *A dynamic approach to second language development*. Amsterdam: John Benjamins. doi: 10.1075/z.100
- Vyatkina, N. (2012). « The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study ». *The Modern Language Journal*, vol. 96, n° 4. pp. 576-598. doi:10.1111/j.1540-4781.2012.01401.x
- Wolfe-Quintero, K., Inagaki, S. & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Honolulu, HI: University of Hawaii Press.
- Zhang, Y. & Glass, J. R. (2009). « Unsupervised spoken keyword spotting via segmental DTW on gaussian posteriorgrams ». In *Proceedings of IEEE Automatic Speech Recognition & Understanding Workshop (ASRU)*. pp. 398-403.

Software

Childes CLANWin software (MacWhinney, B., 2000). Disponible en ligne. <http://childes.psy.cmu.edu/clang/>.

ANNEXES

Annexe 1

First text as hand coded for items mentioned in Table 2. The definitions and operationalizations are all according to Verspoor and Sauter (2000).

[The story] **is** about [the relationship between a father and his two daughters] and about [the way (he **influenced** them ^{FRC})]. SIMPLE (not two complete main clauses with subject and predicator)

After [their father's death], [the two grown-up sisters, Constantia and Josephine] **weren't** controlled by [their father] anymore and also [they] **couldn't** rely on [him] anymore. COMPOUND

[They] **were** forced (to make more [decisions] on [their own] now ^{NFC}), (in which [they] **didn't** succeed ^{FAC}). COMPLEX

[They] **couldn't** even decide on [trivial matters] like (whether [they] **would** like [their fish] to be boiled or to be fried ^{FNC}). COMPLEX

Even [at the end of the story], [they] **couldn't** make up [their minds] about [ordinary matters]. SIMPLE

[The story] **ends** with [a dialogue between the two women about the question (whether they **should** keep Kate as a maid or fire her ^{FNC})]. COMPLEX

[The father] **had** always prevented [them] (to make [decisions] on their own ^{NFC}). SIMPLE

Shortly (after [their father] **died** ^{FAC}), [the women] talked to each other, (lying in [their beds] ^{NFC}), about [very trivial things] ([they] **felt** had to be done ^{FRC}). COMPLEX

[Constantia] even **wondered** (whether [they] **should** have [their dressing-gownes] dyed black ^{FNC}). COMPLEX

[They] **talked** to each other in [an exciting, childish way]. SIMPLE

In fact, [they] **didn't** really talk, [they] **uttered** [their opinions] by shrieking, giggling and snapping, so [they] **didn't** express [their feelings] in [a normal way]. COMPOUND (with but understood)

Average sentence length: 205 words / 11 sentences= 18.63

Finite verb ratio: 205 words / 20 finite verbs= 10.25

Types of sentences: 4 simple, 2 compound, 5 complex= 36% simple, 18% compound, 45% complex and 0% compound complex sentences

Noun phrase length: 51 noun phrases (total 104 words in noun phrases)= 2.04

Types of dependent clauses: 2 finite adverbial clauses (FAC), 3 finite nominal clauses (FNC), 2 finite relative clauses (FRC), 2 non-finite clauses (NFC)= 22% FAC, 33% FNC, 22% FRC, 22% NFC.

Annexe 2

Raw scores measures in Study 1

Text	Types	LEX						SYNTAX													
		Tokens	TTR	D	Mostfreq	UniqWrds	AvWrtL	#FinVerb	FinVerbTot	AvSL	Simple	Compd	SIM+CD	Compd-spr	Complex	COMPLEX	Adverbial	Nominal	Nonfinite	Relative	AvNPL
1	111	207	0.54	67.52	25.00	26.67	5.22	24	8.63	18.64	27.27	27.27	54.55	0.00	45.45	45.45	8.33	25.00	50.00	16.67	2.16
2	100	200	0.50	67.45	44.44	13.89	5.31	25	8.00	18.18	36.36	0.00	36.36	18.18	45.45	63.64	33.33	33.33	26.67	6.67	2.38
3	107	185	0.58	93.52	25.35	23.94	5.58	19	9.74	20.00	55.56	0.00	55.56	33.33	11.11	44.44	7.69	23.08	53.85	15.38	2.89
4	115	129	0.58	84.44	27.63	13.16	5.87	20	9.95	16.42	50.00	0.00	50.00	0.00	50.00	50.00	18.18	27.27	27.27	27.27	3.03
5	97	193	0.50	51.90	20.97	14.52	4.32	22	8.77	19.30	20.00	20.00	40.00	40.00	20.00	60.00	50.00	25.00	25.00	0.00	2.41
6	116	207	0.56	73.06	25.76	22.73	5.68	25	8.28	18.64	20.00	40.00	60.00	10.00	30.00	40.00	35.71	14.29	50.00	0.00	2.26
7	116	198	0.59	75.25	25.00	22.22	5.10	23	8.61	27.86	0.00	14.29	14.29	14.29	71.43	85.71	29.41	25.29	17.65	17.65	2.87
8	98	214	0.46	57.06	21.79	14.10	5.26	23	9.30	26.25	12.50	0.00	12.50	50.00	37.50	87.50	7.69	11.54	65.38	15.38	2.75
9	94	203	0.46	43.60	27.14	14.29	4.70	14	14.50	28.86	14.29	0.00	14.29	20.57	57.14	85.71	16.67	0.00	25.00	58.33	4.41
10	102	208	0.49	56.28	28.00	26.67	4.99	24	8.67	20.80	10.00	10.00	20.00	20.00	60.00	80.00	11.76	5.88	41.19	41.19	2.30
11	96	215	0.45	58.29	52.70	12.16	4.66	27	7.96	26.89	12.50	0.00	12.50	37.50	50.00	87.50	17.65	11.76	23.53	47.06	4.82
12	125	218	0.57	79.34	35.06	24.68	5.04	24	9.08	19.73	36.36	0.00	36.36	9.09	54.55	63.64	5.00	25.00	30.00	30.00	2.78
13	112	209	0.54	63.17	33.33	24.64	5.25	21	9.95	23.22	22.22	0.00	22.22	0.00	77.78	77.78	23.53	29.41	23.53	23.53	3.02
14	105	208	0.50	74.64	51.95	7.79	6.16	15	13.87	34.33	33.33	0.00	33.33	33.33	33.33	66.67	14.29	7.14	42.86	35.71	4.45
15	124	203	0.61	90.92	17.24	41.38	5.55	15	13.53	25.38	37.50	12.50	50.00	0.00	50.00	50.00	13.33	20.00	60.00	6.67	3.47
16	117	205	0.57	85.73	27.59	39.66	5.26	25	8.20	34.67	16.67	0.00	16.67	66.67	16.67	83.33	25.00	20.00	35.00	20.00	3.53
17	142	216	0.66	103.78	19.75	29.63	5.54	20	10.80	26.25	12.50	0.00	12.50	25.00	62.50	87.50	10.53	10.53	47.37	31.58	4.19
18	119	209	0.57	57.18	30.00	40.00	7.74	11	19.00	29.57	57.14	0.00	57.14	0.00	42.86	42.86	0.00	12.50	37.50	50.00	6.19
19	109	198	0.55	66.43	38.96	16.88	7.26	14	14.14	26.00	62.50	0.00	62.50	0.00	37.50	37.50	10.00	30.00	30.00	30.00	3.23
20	116	208	0.56	85.00	14.47	22.37	6.20	17	12.24	34.17	14.29	14.29	28.57	28.57	42.86	71.43	16.67	16.67	41.67	25.00	3.26
21	118	200	0.59	80.92	12.17	25.22	5.63	12	16.67	28.14	71.43	0.00	71.43	14.29	14.29	28.57	14.29	14.29	57.14	14.29	3.90
22	103	186	0.55	81.00	32.50	22.50	6.58	18	10.33	26.57	28.57	0.00	28.57	14.29	57.14	71.43	35.71	21.43	28.57	14.29	3.49

NOTES

1. There are many terms used for basically the same theory such as *Complex Adaptive System* or *Dynamic Systems Theory*. At a recent conference in Nottingham on the dynamics of motivation, several experts agreed that the term *Complex Dynamic System* is a nice compromise.

2. CHAT and CLAN were developed as part of the Child Language Data Exchange System (CHILDES) project (MacWhinney, 2000).

3. The classification of sentence types is according to definitions given in Verspoor and Sauter (2000). A simple sentence is considered simple if it has one main clause, which may include non-finite constructions. This decision was made because it is often difficult to distinguish semi-modal constructions such as "wants to do something", "supposed to do something" from "decided to do something".

RÉSUMÉS

De plus en plus de recherches montrent que le langage est un système dynamique et que le développement langagier est un processus dynamique qui se caractérise par de la variabilité (changements intra-individuels) et de la variation (différences inter-individuelles). Le développement de la complexité linguistique serait donc un processus individuel et il conviendrait de ne pas tenir pour acquis qu'une généralisation, au-delà de l'individu, est possible. Cependant, si nous voulons explorer des schémas communs chez les apprenants, il serait plus pertinent de retracer l'évolution du développement dans des études de cas. Cet article explore les mesures de la complexité qui semblent caractériser le mieux le développement à des niveaux avancés en L2. L'étude 1 est une étude de cas unique qui questionne ces mesures de la complexité afin de déterminer lesquelles peuvent capturer le mieux un développement général pour ce cas précis. La longueur moyenne des mots et le ratio de verbes personnels ont montré une forte corrélation avec à la fois le développement dans le temps et l'évaluation des productions écrites. L'étude 2 reprend ces deux mesures avec en plus celle des propositions subordonnées dans les productions de deux autres apprenants. Une analyse basée sur les modèles de Markov cachés a également montré que les apprenants étaient à des stades très différents et cela confirme l'hypothèse dynamique des trajectoires particulières individuelles.

As shown by a growing body of research, language is a dynamic system and language development is a dynamic process characterized by variability (intra-individual changes) and variation (inter-individual differences). This implies that the development of linguistic complexity is an individually owned process and we should not assume a priori that generalization beyond the individual is possible. Therefore, if we want to explore common patterns in learners, it is best to trace development in individual cases. This paper explores which linguistic complexity measures most convincingly characterize development at advanced L2 stages. Study 1 is a single case study that explores which linguistic complexity measures capture overall development best for this individual and average word length and finite verb ratio are found to correlate strongly with both development over time and text ratings. Study 2 traces these two measures and dependent clauses in two other, similar learners. The three learners were indeed somewhat similar in the development of the two general measures, but not in the development of dependent clauses. A Hidden Markov Modeling analysis also showed that the learners developed in different stages, confirming the dynamic hypothesis of individually owned trajectories.

INDEX

Mots-clés : développement dynamique, niveau académique, mesures de la complexité linguistique, longueur moyenne de mots, ratio de verbes finis, proposition subordonnée finie
Keywords : dynamic development, academic level, linguistic complexity measures, average word length, finite verb ratio, finite dependent clauses

AUTEURS

MARJOLIJN VERSPOOR

Marjolijn Verspoor est professeure en langue anglaise et en anglais langue seconde à l'université de Groningen. Ses travaux visent principalement la compréhension du développement d'une langue seconde dans une perspective de complexité dynamique pour la mesure de l'efficacité des méthodes d'enseignements, et plus particulièrement celles qui s'inspirent d'une perspective basée sur l'usage.

Courriel: m.h.verspoor[at]rug.nl

Adresse: Department of English Language and Culture, University of Groningen, PO Box 716, 9700 AS, Groningen.

WANDER LOWIE

Wander Lowie est docteur en linguistique à l'université de Groningen. Il est directeur du département de Linguistique Appliquée et membre du comité de direction du Centre du langage et de la cognition à l'université de Groningen. Actuellement, ses recherches portent principalement sur les applications de la théorie des systèmes dynamiques au développement d'une langue seconde (enseignement-apprentissage).

Courriel: w.m.lowie[at]rug.nl

Adresse: Dept. of Applied Linguistics, University of Groningen. PO Box 716, 9700 AS, Groningen.

HUI PING CHAN

Hui Ping Chan a obtenu son doctorat à l'université de Groningen au département de Linguistique Appliquée en 2015. Ses travaux questionnent le développement des compétences orales et écrites, l'apprentissage du vocabulaire et la modélisation du développement langagier dans une perspective de théorie des systèmes dynamiques.

Courriel: huipingchan1981[at]gmail.com.

Adresse: Dept. of Applied Linguistics, University of Groningen. PO Box 716, 9700 AS, Groningen.

LOUISA VAHTRICK

Louisa Vahtrick est éditrice d'ouvrages portant sur le domaine de l'anthropologie à Routledge/Taylor and Francis Group. Elle a participé à la publication de monographies de recherche, de collections d'ouvrages et de manuels dans plusieurs disciplines dont l'anthropologie linguistique.

Courriel: lmb.vahtrick[at]gmail.com.

Adresse: Dept. of Applied Linguistics, University of Groningen. PO Box 716, 9700 AS, Groningen.