Variability in L2 development from a dynamic systems perspective¹

**Introduction**

The main purpose of this article is to illustrate that studying intra-individual variability in Second Language Development (SLD) provides important insights into the developmental dynamics that are traditionally being ignored. In our approach, we have adopted the framework of Dynamics Systems Theory (DST) (Thelen & Smith, 1994; Van Geert, 1994). This theory aims at describing and explaining the ways in which complex systems (such as developing individuals) change over time due to a process of self-organization. It also holds that within subject variability—both systematic and free—occurs in any complex system continuously, but the degree of variability may change depending on how stable or unstable the system is at a given moment. A relatively more unstable period is often a sign that the system is changing. Therefore, by looking at the different degrees and patterns of variability in dense developmental (second language acquisition) data, we can discover how and when different sub-systems are changing and developing, and how they relate to each other.

Although variability is not analysed systematically in the SLD research tradition, several lines of research have paid attention to it in the 70’s and 80’s of the previous century. We will begin by providing a brief overview of this earlier research, whose main focus was to discover the causes of variability and the order of acquisition. Then we will report on the main findings in developmental psychology pertaining to variability. We

¹We are very grateful to Sybrine Bultena for her critical editing of the final version. Obviously, any remaining errors remain our own responsibility.
will apply the main insights into the role of variability to two SLD cases, while demonstrating several DST techniques to make the patterns of variability visible.

*Variability in SLD*

Generally, according to Ellis (1994) there are three main approaches within the field of SLD: the linguistic, sociolinguistic and psycholinguistic approach. The linguistic approach, inspired by Chomskyan linguistics, was not interested in variability very much because the main object of study was a speaker’s “competence” rather than his/her “performance”. The general claim was that one had to abstract away from individual instances to find out what learners “know” rather than “do”. To do so, researchers often used grammatical judgment tests to ascertain learners’ intuitions rather than using actual learners’ data. In contrast, a DST perspective does not make a distinction between “performance” and “competence”. According to Thelen and Smith, the distinction does not make sense in developing systems that are highly interactive in nature, such as language (1994). From a DST perspective, all development is in fact the accumulation of real-time actions, and variability is the expression of the adaptability of the system.

Also psycholinguistic and sociolinguistic approaches to SLD took interest in variability and attempted to discover the sources that were responsible for the variability and what it could tell us about the Interlanguage (IL) system; however, the IL system was often viewed as an inherently static system in which the variability could be linked to isolated factors. The psycholinguistic stream, inspired among others by Krashen’s monitoring theory, focuses on the intra-individual variability related to attention, processing and demands of short-term memory (cf. Hulstijn & Hulstijn, 1984). Within the sociolinguistic approach, the “dynamic paradigm” was interested in how the IL system
changes and drew upon language change models such as the Wave model proposed
Bailey (1973) and Bickerton’s (1975) implicational scaling model. For example,
Gatbonton’s (1987) gradual diffusion model, based on Bailey’s work, first involves the
acquisition of correct variants into all the relevant contexts and then the replacement of all
the incorrect variants there. Both the acquisitional process and the replacement process
take place gradually. However, Gatbonton’s goal is to attribute the sources of variability
to specific environmental factors, such as context. Such an approach often testifies of an
implicit suspicion towards the meaningfulness of the data. In case of the gradual diffusion
model, the supposed ‘real’ replacement process is considered to be ‘distorted’ by
environmental factors, which Gatbonton tries to link directly to known contextual
influences. DST, on the other hand, takes a radically different approach and considers
variability to be the product of the internal dynamics of a process of self-organization, not
as something that is added to it externally and therefore does not need to be explained by
contextual factors.

Another sociolinguistic group, inspired by theories such as Labov’s, looked at a
great number of factors such as interlocutor (L1 or L2 speaker), situational context
(formal, informal), task (speaking, writing), form-function relations (e.g. certain types of
article errors occur with certain types of nouns), and the like to account for the variability
observed. One strong proponent of this variationist school is Tarone (1988), whose
overview of variation in interlanguage gives a detailed account of the different theoretical
perspectives and concludes that despite the great number of studies that have been
conducted, not all causes of variability have been explained. She therefore calls for more
longitudinal studies to “show the way in which variation at a single point in time is tied to
longitudinal development of an interlanguage” (Tarone, 1988, p. 137). However, her main
focus remains on the causes of the variability rather than describing or analysing patterns of variability. Also Young (1988), who used a multi-factor analysis, found that different types of variability occurred for learners of high and low proficiency and that different linguistic contexts affect performance. However, the approach is very similar to that of Gatbonton, in the sense that much effort is put into attributing external sources of variability, instead of considering it as an internal characteristic of a developing system.

Another stream of linguists did not look at variability so much for its own sake, but dealt with IL variability in their search for (universal) developmental stages. Whereas some studies were able to show neat developmental stages by clustering several morphemes in one stage (e.g. Dulay and Burt, 1974), others such as Meisel, Clahsen, and Pienemann (1981) point out that socio-psychological factors combined with individual learning styles need to be taken into consideration when trying to discover the order of acquisition and conclude that development contains a great deal of variability and is by no means as linear as is often suggested by group studies. Even though Meisel et al.’s view of a multiplex of factors playing a role in SLD would be in line with DST, it does not look at how the variability within one individual itself plays a role in development.

Only a few applied linguists, for example Huebner (1985) and Ellis (1985, 1994), explicitly state that variability could give us insight, not only into an individual’s stages of SLD, but more specifically into the developmental process of L2 acquisition:

Not only may different acquirers follow different paths, but those paths of grammatical development may in fact be nonlinear. Huebner (1983), a microanalysis of variation within several sub-systems of one learner’s grammar over time, suggest that this in fact is the case. It also suggests that what appear to
be irregularities may be highly revealing of how functions are mapped onto target-like forms (Huebner, 1985, p. 145).

In addition, after eliminating factors that would contribute to “systematic” variation, Ellis found that there was still some degree of “free” variation, variability that could not be attributed to any known linguistic, situational or psychological factor. He cites Cancino, Rosansky, and Schumann (1978) and concludes that a general finding of a study such as this one is that “free variation occurs during an early stage of development and then disappears as learners develop better organized L2 system” (Ellis, 1994, p. 137). In another publication, Ellis points out how a particular learner’s behaviour may change over time:

First, there is a spread of rules along the interlanguage continuum, from the careful towards the vernacular style and from simple to complex linguistic environments. This process is motivated by the learner’s felt need to be socially acceptable and is helped by practice which automatizes rules that initially can be applied only when the learner is attending to his speech. Second, there is the need to make the interlanguage system more efficient by removing free variability. This involves the progressive reorganization of form-function relations and the eventual elimination of redundant forms. (Ellis, 1985, p. 96)

It is this type of variability that we wish to look at in more depth from a DST perspective. Because variability is an inherent property of a self-organising system, the degree and patterns of variability can tell us more about the developmental process. In the next
section, we will summarize the main findings concerning variability and development, and then we will show how patterns of change may be analyzed. The essence of this process-oriented approach is that we analyse patterns of variability that may help in generating intriguing hypotheses on the developmental mechanisms in SLD.

Variability and development

Whereas in SLD research, variability has not been a main focus of investigation, in the field of developmental psychology such a strand of investigation took off rapidly soon after the introduction of Dynamic Systems Theory (Fogel & Thelen, 1987; Thelen & Smith, 1994; Van Geert, 1994).

According to DST, variability is a central element, an *intrinsic property* of a self-organizing developing system (cf. De Bot, Lowie & Verspoor, 2005). Thelen and Smith state: “Variability is a metric of stability and a harbinger of change” (1994, p. 342). Variability is needed for a learner to explore and select. For example, only when learners have access to a variety of forms are they able to select those that help them develop, so the more different forms they can select from, the more likely development is to take place. A dynamic approach does not distinguish between acting, learning, and developing, but considers only change over time occurring at different scales of time. Essential to this approach is that variability is elevated, both within and between individuals, into a central element of a developing system. Thus, intra- and inter-individual variability are important features that should be treated as data and be analysed (Thelen & Smith, 1994, pp. 341-2).

In order to elevate variability and observe developmental changes, we need longitudinal studies with repeated measurements, especially during the period that strong
development in a particular sub-system may be occurring. Within this period the density of observations should be high, relative to the rate of change and should be analyzed intensively with the goal of inferring the representations and processes that gave rise to them. Different kinds of variability at different levels can be observed. For example, in SLD, we could trace how the relative distribution of different forms, such as negative constructions or different verb forms, changes over time. In writing we could also see how during each writing session the proportion of shorter versus longer sentences changes, but to get a more global view at a slightly higher level, we would also be able to observe how the average sentence length per writing session changes over time.

Variability is also of main interest in the microgenetic approach to (cognitive) developmental processes. As the name implies, this approach focuses on the “microgenesis” of development (Lavelli, Pantoja, Hsu, Messinger, & Fogel, 2004), that is on the moment-to-moment change in a dense repeated measures design. Although microgenetic studies are not by definition DST oriented, both directions share a process-oriented focus on the changing individual as the fundamental unit of study.

In his overview article of microgenetic analyses of learning, Siegler (2006) summarizes the main findings of 105 microgenetic studies and concludes that studying within-subject variability is important to (a) predict change, (b) analyze change, and (c) understand change mechanisms (p. 481). Below, we will summarize those that are relevant for our own data.

Variability studies have shown that within-subject variability in strategy use is substantial at all ages from infancy to adulthood, during all phases of learning and at every level of analysis. For example, Adolph (1997) found in his study that each of the 5 to 15-months old infants use at least five different strategies to descend down relatively
shallow ramps and six down relatively steep slopes at different occasions. They may refuse to go down or slide on their belly or on their behind, sitting or lying, or head or feet first. For adults, an experiment by Dowker, Flood, Griffiths, Harriss, & Hook (1996) showed that even when all variables are controlled for, an adult’s use of strategies, even an expert’s, will vary, albeit to a lesser extent than the non-expert’s.

Several microgenetic studies also show that learners do not progress neatly in acquiring a skill or strategy but may regress. It tends to be greatest during periods of rapid learning, but substantial variability is present in relatively stable periods as well. It also tends to be cyclical, with periods of lesser and greater variability alternating over the course of learning. Children may use more advanced approaches on one occasion and then regress to less-advanced ones on the next, but these regressions are temporary as the general trend of change was upward in all the studies. However, at early stages, progress involves a back and forth competition between more or less advanced strategies (Siegler, 1995 in Siegler, 2006).

High initial within-subject variability tends to be positively related to subsequent learning and such learning reflects addition of new strategies, greater reliance on relatively advanced strategies already being used, improved choices among strategies, and improved execution of strategies. For example, on number conservation and sort-recall tasks, children who used more and different strategies on the pre-test used the more advanced strategies on subsequent tasks (Coyle & Bjorklund, 1997; Siegler, 1995 in Siegler, 2006).

Learning tends to progress through a regular sequence of knowledge states that parallel those that characterize untutored development. The path of learning is usually similar for learners of different ages and different intellectual levels. Also learning often
includes short-lived transitional approaches that play important roles in the acquisition of more enduring approaches. As far as the rate of development is concerned, rates of both discovery of new approaches or strategies and their uptake increase with both age and IQ. Subjects who start with more advanced strategies are likely to progress to yet more advanced approaches more rapidly than children whose initial rules are less advanced (Siegler, 2006). Interestingly, small differences in initial conditions seem to have a great effect on subsequent development. As an example from the field of first language acquisition, Tomasello (2001) found that 2.5 year olds learned the past tense ending faster when they were presented verbs with pronoun frames (e.g. He kicked it; She pushed it) than with more specific noun frames (e.g. Sam kicked the ball; Jane pushed the car). Apparently, the greater consistency in the frame made it easier to concentrate on past tense verb ending.

The general findings in variability studies mentioned so far also seem to apply to the negative construction data as presented by Cancino, Rosansky, and Schumann in 1978, which will be reinterpreted below. They describe the natural untutored acquisition of English negatives by six native Spanish speakers who had been in the US for less than three months over a period of about ten months with measurements taken every two weeks. Marta and Cheo, both aged five, Juan and Jorge aged 11 and 13 respectively, and two adult subjects, Dolores (25) and Alberto (33). The children were all from upper middle class families and attended US public schools, where they interacted with English peers on a daily basis. Dolores has a middle-class background and is exposed to English on a daily basis as she babysits English speaking children. Alberto has a lower-class background and works in a factory where some of his input is from other non-native
speakers of English. All subjects spoke Spanish at home. The subjects were interviewed twice a month in their homes for about 10 months.

Cancino and her colleagues tested whether the stages in L2 development were similar to those found earlier in L1 development by Klima and Bellugi (1966), namely:

1. **No V** constructions: e.g. *Carolina no go play* (a non-target construction)
2. **Don’t V** constructions: e.g. *I don’t hear* (sometimes a target construction, but often over-generalized to non-target contexts)
3. **Aux-neg** constructions (especially *isn’t* and *can’t*): e.g. *You can’t tell her* (A target construction)
4. **Analyzed do** constructions (*do not, doesn’t, does not, didn’t, did not*): e.g. *One night I didn’t even have the light* (A target construction)

They conclude that the L2 stages of acquisition were indeed similar to those in L1 development: each subject would go through stage one before proceeding to stage two, and so on.

The figures they use to support their conclusion capture not only the separate stages but also how the distribution of the four different negative construction changes over time. Analysing these graphs from a DST point of view yields a more refined picture of development. Their figures show that even though the learners will go through stage one before two, two before three, three before four, not all learners go through stage one and there is a high degree of both between-subject and within-subject variation. Even though all subjects have Spanish as their first language and have been in an English speaking country for about three months, each subject shows a different pattern. Some
start off with stage one and two, others with stage one, two, and three, others with stages one, two, three, and four, and others with stages two, three, and four. All the figures show that the relative frequency of each negative construction changes, with some becoming less and others more frequent, some first becoming more frequent and then less frequent, and some never becoming very frequent, but these changes take place at different rates.

Transitional phases, marked by a greater degree of variation, are difficult to see in the figures as presented in Cancino et al. (1978); in the next section we will show how they can be made more visible. As an illustration, one very clear transitional phase can be seen in Juan’s data pictured in Figure 1. The graph shows that the relative distribution of the different constructions is much more variable in tape 1-7 than after tape 7 (tapes were recorded every two weeks). After Tape 7 the degree of variation stabilizes and the subsystem of negative constructions seems to remain relatively stable.
Figure 1. Development of negation in Juan showing proportion of each negating device to total negatives in each sample. From: Cancino, Rosansky, and Schumann, 1978, p. 214.

The possible regression in the developmental process can be most clearly seen in Marta’s case (see Figure 2). Within the first three tapes, she uses three different strategies, but by Tape 5 the least advanced strategy, no V peaks enormously, which then nearly disappears by Tape 7. Also, all four children keep using the non-target no V after the more target-like constructions have appeared. The no V does not disappear until the analyzed do is firmly in place.

Figure 2. Development of negation in Marta showing proportion of each negating device to total negatives in each sample. From Cancino, Rosansky, and Schumann, 1978, p. 212.
A positive relation between number of strategies and subsequent performance can be seen if we compare Juan with Alberto (see Figure 3). In the first tape, Juan uses all four negative constructions, including the target ones, to a higher degree than the others (except one adult subject who had already mastered the negative construction system when the first recording took place). He progresses through the interim stages very rapidly and by Tape 7, his negative is fully formed (Cancino et al., 1978). On the other hand, Alberto, who starts off with only two clear strategies, *no V* and *don’t V*, keeps using only these throughout the study and it may be assumed that the constructions have already fossilized.

![Graph showing development of negation in Alberto](image)

**Figure 3.** Development of negation in Alberto showing proportion of each negating device to total negatives in each sample. From Cancino, Rosansky, and Schumann, 1978, p. 216.
Finally, the conclusion Cancino and her colleagues draw is in line with findings in other microgenetic studies: learning tends to progress through a regular sequence of knowledge states. They show that even though the participants have very different learning paths, each of the participants, both the younger and older ones go through the stages of negative formation strategies in a similar order, which in turn are very similar to the ones that L1 learners of English go through (1978).

Looking at the variability in the data of Cancino et al. (1978) from a DST perspective, we can clearly see that the acquisition of the negative construction is not a linear sequence, going neatly from stage one to four. Even though learners may go through general stages, the within-subject and between-subject variability is great and even when the target constructions have been used, subjects will regress to non-target constructions, with transitional stages showing a great deal of variation. There also seems to be a relation between the degree of variability at early stages and acquisition. All of the subjects, except Alberto, used several constructions in early stages, but Alberto used only two and the lack of variability might be related to a premature attractor state, fossilization.

*Patterns in variability*

Although the figures as presented by Cancino et al. show the changes in the different constructions over time, raw data plots are difficult to interpret. To help us visualize the degree of variability in developmental data, several techniques have been developed, for example by Van Geert and Van Dijk (2002). In this section, we will illustrate various ways to present the variability, using an SLD case study of 18 academic writing samples written by an advanced learner of English over the course of three years. At the same
time, we will argue that looking at the variability using these techniques, we may discover interesting facts about the developmental process that would otherwise remain hidden.

The case study\(^2\) involves an advanced learner of English (JtB) enrolled as an English major at the University of Groningen. There are several reasons why JtB should be considered advanced, with an assumed TOEL score close to 600. Dutch is rather similar to English, most Dutch students are exposed to a great deal of English and in the media from an early age on, and most have at least nine years of English at school before they enter university. Since this advanced learner makes very few errors in her writing, we did not look at errors but focused on development in vocabulary use and sentence complexity. For vocabulary use we looked at word length (WL) and two qualitative measures to assess lexical creativity: the type-token ratio (TTR) and degree of use of words from the Academic Word List (AWL). In addition to these vocabulary measures, we included sentence length (SL) as a fluency measure because it reflects to a great extent vocabulary acquisition and the ease with which the vocabulary is used.

For sentence complexity we decided to use two measures that provide a fine-grained look at syntactic complexity. One measure is the length of the noun phrase. One way to make sentences more complex is to produce longer noun phrases. In English pre-modification is not very productive, but in Dutch premodification may be applied recursively, leading to extremely long noun phrases\(^3\). We assumed that Dutch learners of English may initially increase NP length in English in a native-like fashion, but are likely to overshoot the mark by applying L1 tendencies to their English writing. The second

\(^2\) We hereby would like to thank Jeannette ten Brink for making all her writing assignments available and Daniëlle Zijlstra for allowing us to use the data from her MA thesis.

\(^3\) A case in point is a noun phrase in a sentence that was used in *Jiskefet* a satirical TV show running in the 1990s: ‘De bij de drinkplaats van de immers buitengewoon schuwe toekans hinderlijk aanwezige bosnegers …’ (the at the watering place of the after all extraordinarily shy toucans disturbingly present maroons …)
sentence complexity measure concerns the complexity of the sentence as a whole. Common complexity measures usually refer to the length of the utterance, either sentence, t-unit or clause, or refer to the amount of subordination. However, for this advanced learner, these three separate measures may not entirely capture the degree of complexity because they do not bring to light additional complex constructions such as longer noun phrases or non-finite constructions. To calculate such an overall degree of sentence complexity in one single measure, the number of finite verbs in each sentence is divided by its own sentence length in words; the resulting ratio is averaged for the entire text, resulting in a single measure of language complexity that is simple to calculate: the number of words per finite verbs in a sentence (W/FV). The higher the W/FV ratio, the more complex the sentence is.

The fact that variability should not be underestimated can be illustrated with the following example. Figure 4 displays the development of the application of words from the AWL. Since the assignments were all practices of academic writing, we would expect to see an increase in the usage of these words as the writing proficiency of this student increases. Although a slight general increase in the use of academic words is clearly distinguishable, the fluctuations immediately catch the eye. While in some texts, academic words are sometimes used fairly regularly, they are hardly used at all in other texts. For instance, in assignment 10 the use of academic words ‘peak’ to 12 instances per text, but the writer regresses in the consecutive texts, where academic words are used only two or three times.
Figure 4. Development of the application of words from the AWL in writings by JtB.

Compared to the relative number of words from the AWL, the average word length increases more steadily. In Figure 5, the AWL in each text is plotted and a smoothing function is added to indicate the general trend in the data using a polynomial of the second degree\(^4\).

\(^4\) In this case we have used a polynomial function of the second degree. Other common smoothers are moving averages and regression lines. For the remainder of the discussion on smoothing techniques the exact type of smoother makes little difference.
The smoothing function illustrates a general increase, from around 4.5 letters per word to little below 5.5 letters per words. Thus, on average, this second language learner uses longer words as time progresses. The question, however, is to what extent this increase adequately characterizes the development of word length in the learning process of this student. Because the smoothing line replaces the actual data, it may conceal an essential element in the developmental process\textsuperscript{5}. However, the raw plotted data is not very revealing in itself either because we can see no clear structure around the smoothing line. To see if there are any clear patterns, such as transitional phases with a greater degree of variability at particular points of time, we have applied one of the techniques developed by Van Geert and Van Dijk (2002). In this case we have used the moving min-max graph, which shows the bandwidth of observed scores and highlights the score ranges so that subtleties in the developmental process can be distinguished.

\textsuperscript{5} Researchers who focus on these smoothing techniques, probably consider the variability as relatively uninteresting in itself, for instance because it is seen as a form of error fluctuation.
The moving min-max graph plots the *score range* for each measurement occasion by using a *moving window*, a time frame that moves up one position (= one measurement occasion) each time. Each window largely overlaps with the preceding windows, using all the same measurement occasions minus the first and plus the next. For instance, for every predetermined set of consecutive measurements, the maximum and the minimum values are calculated. This is done by way of a predetermined moving window, of for instance five positions, such that we obtain the following series\(^6\):

\[
\text{min}(t1..t5), \text{min}(t2..t6), \text{min}(t3..t7), \text{etc.}
\]
\[
\text{max}(t1..t5), \text{max}(t2..t6), \text{max}(t3..t7), \text{etc.}
\]

Once the moving minima and maxima are plotted, one can visually inspect whether there are considerable fluctuations over time that are developmentally meaningful by contrasting them with the eventual long-term changes in the minima and maxima.

Applying a min-max graph to the data of the average word length of Figure 5 shows some interesting characteristics. When we inspect the bandwidth of scores, we can see that it does not remain stable across the trajectory. Initially, the range of scores is very small (the first 5 texts are very much alike with regard to word length), but it increases from the 6\(^{th}\) text onwards. Between the 6\(^{th}\) and 12\(^{th}\) text, this bandwidth is the largest, and the average length of words per text fluctuates between around 4.2 and 5.3. Then, from observation 13 onwards, we see a much more stable performance again with regard to word length.

\(^6\) Any commercially available spreadsheet program offers functions such as MIN and MAX that can easily be computed over moving data windows.
This seems to show that the data consist of 3 steps (or maybe even 3 ‘stages’ or ‘sub-stages’). First, we see a period of relative stability (with even a slight decrease in text 6 and 7). Secondly, there is a highly variable part of the graph in which performance fluctuates from day to day, and seems to ‘jump’ to a higher level of performance. Finally, word length stabilizes again and shows only very mild fluctuations. In this sense, Figure 6 is a ‘classic’ example of a step-wise developmental process, with a period of enlarged variability as a transition phase between the two stages. The observation of enlarged variability in the vicinity of a developmental jump replicates similar findings in other developmental domains.

Figure 6. Min-max graph representing the development of average word length of JtB’s writings.

Further detailed insight into dynamic developmental processes in individual data can be achieved by looking at combinations of developmental variables. From a dynamic
systems point of view, there is a line of research that focuses on describing and modelling developmental pathways of different sub-systems and their interactions. Some connected growers, as Van Geert (1994) calls them, support each other’s growth, while others show competing relations. An example of a supportive relation could be the relation between lexical development and sentence length; with an increased command of vocabulary it is easier to make longer sentences and longer sentences require more different words. As was pointed out by de Bot, Lowie, & Verspoor (2005), a subject’s resources are limited and have to be distributed over different sub-systems that grow, but not all sub-systems require equal amounts of resources. In this way, the two connected growers need fewer resources than two growers that are unconnected. Also, the concept of carrying capacity is relevant for SLD. Since growth is resource dependent and resources are limited, growth is by definition limited. The carrying capacity refers to the state of knowledge that can be attained in a given subject’s interlinked structure of resources, referred to as the cognitive ecosystem (Van Geert, 1994). For example, as Robinson and Mervis (1998) have shown, in L1 the emergence of a multi-word sentence may coincide with a deceleration of lexical growth in the one-word phase. While in earlier phases all resources could be used to develop the lexicon through the linking of different types of sensory information, in the next phase more and maybe different resources are needed to develop the grammatical system that governs the functional distribution of information in multiple-word utterances.

It should be noted that this approach makes sense only if the variables that are studied in combination have a meaningful relation to each other. Therefore, it is reasonable to look at the interaction of the competing syntactic constructions or that of two measures of language fluency. An illustration of the occurrence of interaction
between sub-systems is the relationship between the development of the average sentence length in words and the type token ratio. As can be seen in Figure 7a, both variables show an oscillating pattern with a slight—but clearly distinguishable—upward trend (average sentence length rises a little bit steeper). However, when inspecting the oscillations in more detail, we observe a striking fact. It seems as if the ‘waves’ of the average sentence length and the type token ratio alternate almost perfectly. In most cases (at least from text 5 onwards, and with the exception of data point 14), when sentence length is going up, type token ratio is moving down, and vice versa.

![Figure 7a. Development of the average sentence length (SL) in words and the type-token ratio (TTR) in the writings of JtB.](image)
Figure 7b is based on the exact same data, but in this case both sentence length and type token ratio are *detrended*, which means that the general trend (general increase of decrease) is taken away in a graphical representation. In this case we have done this simply by using a linear regression (with regards to methodological consideration on which functions are suited to detrend developmental data, see Van Geert and Van Dijk, 2002). In this graphical representation, only the deviations from the regression line are plotted, this means that positive values indicate local peaks and negative values indicate local wells, in relation to the general linear increase in the data. Detrending the data is an important step when focusing on intra-individual variability, because otherwise the actual local variability is overestimated (or underestimated, depending on the direction of the trend) since a general trend by definition also consists of small local increases and decreases. As can be seen in Figure 7b, the observation on the basis of the raw data corresponds to what can be seen in the detrended representation: when the average sentence length goes up, the type token ratio seems to go down and the other way around.
Figure 7b. Detrended representation of the average sentence length (SL) and type-token ratio (TTR) in JtB’s writings.

On the basis of Figures 7a and 7b, we can reach the following conclusion. In those texts that have somewhat longer sentences, those sentences show somewhat less variety in word use, but when the written language is more colourful (in the sense of that there is more variety in the words used), the sentences that are made are relatively shorter. However, it should be noted that this association between sentence length and type token ratio does not result in a significant correlation, most likely because of the limited number of observations.

Nevertheless, oscillations themselves are fascinating because they suggest a rigid switching between length and variety, most likely dependent on the topic of the assignment. This indicates that these two indices of language fluency might share a competitive relation to each other. Most likely, this competitive relation extinguishes at

\[ r = -0.33, p = 0.182 \]
one point in time, because in native speakers longer sentences are not necessarily associated with a more monotonous word use. Thus, when academic writing proficiency is increasing, there initially seems to be a trade-off between a more varied word use and longer sentences which might be worthwhile investigating in a larger study.

To take another view at the same data we have also plotted sentence length (SL) in relation to the type token ratio (TTR) in a so-called *Space State Grid* (SSG, Lewis, Lamey, & Douglas, 1999). The technique of the space state grid has been developed in the late 1990s as a middle road between DST methods that are heavily mathematical (such as models of coupled growers) and DST methods that are only descriptive (such as the one just presented). Space State Grids are especially suitable for distinguishing stable patterns that emerge out of self-organization, which are called *attractors* in DST terminology. Attractors are a central concept in DST. In fact, according to DST, all developmental acquisitions can be described as attractor patterns that emerge over time, on whatever timescale (Thelen & Smith, 1994). In an SSG, an attractor is represented topographically, and behaviour is conceptualized as moving along a trajectory on this hypothetical landscape being pulled toward certain attractors, and being freed from others (Granic & Hollenstein, 2003). Much like a scatter plot, two variables are plotted in the X- and Y dimension, and each data point represents the values of the two-variable sequences. A trajectory is drawn through the successive data points to see how the scores develop across time. A typical developmental trajectory between two relevant variables start with initial low values on both axis (and thus results in an attractor in the bottom left of the space state grid) and ends with an attractor in the upper right of the grid. In between attractors, many trajectories are possible. From a grid, several different measures can be
derived that can be used for further statistical analysis. In this illustration, we limit ourselves to demonstrating only the graphical representation and its interpretation.

![Figure 8. Space State Grid of the development of type token ratio (TTR) and average sentence length (SL) in JtB’s writings.](image)

The grid in Figure 8 shows that the developmental process of the TTR in relation to SL follows a very ‘spiky’ pattern. First of all, the starting point is localized in the bottom lower quartile. It should be noted that the starting point of this study is fairly arbitrary, because writing assignments started when there already was a fairly advanced level of writing proficiency in this student. Furthermore, we see that the grid trajectory moves up immediately, in the sense that type token ratio increases at first, but then moves across the entire grid. This remains to be the case for almost the entire trajectory. It is only at the
final four data points that we see the diagonals seem to grow shorter, and the values predominantly remain in the upper quartile region. Thus, there are two main observations that result from the grid. The first is that the development sets off in a very variable fashion. Instead of a ‘smooth’ diagonal increase from the lower bottom of the graph to the upper right, we see many diagonals in the opposite direction, which might indicate a competitive relation between sentence length and type token ratio. Secondly, we observe a ‘slowing down’ of variation across the diagonals at the end of the trajectory (form data point 14 onwards), which might indicate the presence of an attractor localized around a type token ratio of 50-55 and a sentence length of 26-34 words.

Another set of variables that may have a meaningful relation to each other (connected growers) would be two different measures of complexity. It would be reasonable to assume that as noun phrases become longer, the sentences on the whole become more complex as they will have more words per finite verb. However, as we mentioned earlier, we suspected that an advanced Dutch learner may have a tendency to first overshoot the mark by making NP’s more complex rather than using other types of non-finite constructions.
Figure 9 displays the developmental trajectories of the average NP length and the W/FV. As can be seen in the graph, both variables increase across time (although average NP only slightly), and show obvious fluctuations. In order to compare these short-term fluctuations of these two variables, we detrended the data again. Figure 10 displays the same data of average NP length and W/FV without the effect of trend and shows that the development of both variables seems to follow each other almost perfectly (in the sense that when one of them increases, the other generally increases as well). The correlation between the two variables also turned out to be significant ($r=0.516$, $p=0.02$), which indicates that the local peaks of the length of noun phrases are associated with local peaks in the number of words per finite verb.
However, there is one interesting exception: around data point 6 and 7 the NP length increases while W/FV decreases. The timing of this deviation is interesting, because it is exactly at the moment of the first major peak in average NP length. Thus, although these two measures of language complexity follow each other with regard to their short-term fluctuations (which indicates the supportive relation between the two indices), there is one exception around the first outburst of NP length. Only at this point in time relatively long noun phrases do not correlate with complexity in verb use. This point in time also coincides with the period during which the alternating pattern of SL and TTR (see Figures 7a and 7b) is less obvious than in the rest of the trajectory.

These variations also seem to be a clear indicator of a developing system. When we compare this data with that of a Native Speaker of English who was enrolled in our
MA programme, we see a much lower degree of variation. Moreover, the two complexity measures do not interact.

Figure 11. Average sentence length and W/FV in four consecutive writings of a native English speaker.

The deviating pattern of JtB around point 6 and 7 also emerges when we plot an SSG of these data. This representation clearly shows a fairly “smooth” diagonal across the cells, which indicates a positive relation between the length of noun phrases and the ratio of finite verbs over the number of words in that sentence. However, there is one outlier (indicated by the circle), and that is data point 7.

\[\text{Note that the values had to be rounded to whole numbers, which --in this case-- can distort the trajectory to a certain degree (that is within each cell, there is no distortion across the trajectory)}\]
This outlier seems to confirm our previous assumption that an advanced Dutch learner of English, when developing towards more complex language use, may first increase the length of noun phrases before using other complex constructions such as non-finite constructions.

The data in this case study show that a detailed focus on learner variability can reveal interesting patterns and interactions in the learner’s development that would remain concealed in a means analysis. Similar to a general means analysis, the data show an overall increase over time of almost all of the variables examined. In addition to this, the variability analysis we carried out shows that some of the measures, like the use of words from the Academic Word List, are very unstable and keep displaying approximately the same amount of variation throughout the trajectory. As we are still
recording the data of this learner, future analyses may reveal if a decrease of variation, possibly with an expected jump in development will indeed occur. Applying a min-max graph to the data of average word length has shown that an increase in variation often precedes a jump in development. Although this pattern is not obvious for all the measures, the occurrence in this case confirms the assumptions based on the microgenetic variability studies referred to earlier. A very interesting pattern emerged when comparing the development of type-token ratio and sentence length. The possibly competitive relationship between these two measures, particularly clear in the detrended data, seems to point to some form of competition between them; the lengthening of sentences coincides with a decrease in what can be regarded as creativity. This seems to indicate a lack of resources, limiting the learner’s capacity to concentrate on both of these measures.

Another coinciding pattern became apparent for the two complexity measures included. The positive correlation between the scores for these measures and the coinciding local highs and lows support the idea of connected growers, both being measures of complexity. However, the peak in NP length at data point 7 shows that sentences became more complex at this point because of longer noun phrases.

Clearly, we have focused on the occurrence of patterns that confirm our assumption that variability appearing in L2 development behaves according principles of dynamic systems. At the same time, it has not been very difficult to find these patterns. The occurrence of attractors, the increase of variability preceding a developmental jump and the interaction of the variability of logically related measures was obvious and pervasive in the data of this case study.
Conclusion

Traditionally, variability studies in SLD have always concentrated on attempting to account for variability as an external source. Conversely, in a DST approach variability must be seen as an intrinsic and central property of a self-organising dynamic system. Starting from an initial state, the system develops through iterations on the basis of the resources available. In such a system, there can be no development without variability and the amount and type of variability can reveal the actual developmental process. Consequently, studying intra-individual variability in Second Language Development provides important insights into the developmental dynamics that are traditionally being ignored. In this article we have applied the insights from microgenetic general learning studies to SLD. These microgenetic analyses have shown that the amount of variability constantly changes and that progress and regress follow each other, showing nonlinear patterns of development. The studies also showed that children who initially used a wide variety of strategies, used more advanced strategies in subsequent tasks. This pattern was also found in a reinterpretation of a variation study by Cancino et al. (1978) on the use of negation. While the authors concentrated on accounting for variability as an external source, their data clearly show the relevance of regarding internal variability as a source of information in itself. Learners who initially used more and different strategies, used more advanced strategies at subsequent stages. The relevance of studying internal variability per se becomes even more apparent when applying more advanced visualisation techniques to dense longitudinal SLD data. The case study we reported on reveals that even for an advanced learner the system can be far from stable. Although a general increase over time is apparent for the correlates included, the development is
nonlinear, showing moments of progress and regress. The case study also brings to light an interesting dynamic interaction of sub-systems. In accordance with the assumption of a limitation of resources, the learner shows a variable development for some related measurements in the course of the trajectory.

We are not claiming that a DST approach should replace other approaches to variability; the visualization of variability is by no means the final step in the study of second language development, nor should it be the ultimate goal of our studies. However, we hope to have shown that it is a crucial and indispensable step towards our understanding of the process of the development of a second language as a dynamic system and that a close look at individual variability from a DST perspective may help us discover developmental patterns that would otherwise remain hidden. Only through longitudinal studies with dense recordings of a wide variety of well motivated measurements can we gain insight into the actual developmental processes and formulate relevant hypotheses about second language development. However, as Thelen and Smith point out, adopting a DST approach does not necessarily mean we have to start our research endeavors all over again:

We encourage our readers to reach into their file cabinets where they store the studies they did not publish because their ANOVAs did not detect significant effects. If errors of design or execution are not at fault, think dynamically and use the variability as data! (Thelen & Smith, 1994, p. 341).
References


