Chapter 6

Dynamic syntactic development in speaking versus writing in identical twins


6.1 Introduction

Taking a dynamic usage-based perspective on language use and language development, a number of researchers (e.g. Larsen-Freeman, 2006; de Bot, Lowie, & Verspoor, 2007, and Verspoor, de Bot, & Lowie, 2011) view L2 development as a dynamic process because all factors involved, including the amount of meaningful exposure and motivation, affect the process continuously. Within this perspective, the language itself is viewed as a complex usage-based system with many sub-systems, all of which are interconnected and may demonstrate shifts at different stages of the developmental trajectory. Such shifts can be described as the inevitable re-organization a dynamic system undergoes throughout its lifetime.

In previous L2 studies that have employed a dynamic usage-based perspective, the focus has been on the variability found within individual measures and how these may interact over time (Spoelman & Verspoor, 2010; Verspoor, Lowie, & Van Dijk, 2008), on the variation amongst learners (Larsen-Freeman, 2006), or on developmental peaks that may be demonstrative of overuse (Van Dijk, Verspoor, & Lowie, 2011). The current paper focuses on stages of development, punctuated by moments of re-organization. We will investigate at what points in the process the interaction amongst different linguistic sub-systems shifts.

Different sub-systems may show various interrelationships over time: precursor, competitive and supportive (Van Geert, 2003; Caspi, 2010). A precursor relationship occurs when a particular sub-system needs to be in place before another can begin to develop, for example in L1, one word utterances (lexicon) precede the occurrence of two or three word utterances. A competitive relation occurs when one sub-system develops at the expense of another; for example, in L1, a growth in multi-word utterances is usually accompanied with a dip in the rate of learning new single words (Robinson & Mervin, 1998). A supportive relationship occurs when sub-systems grow in tandem. For example, once the lexical and syntactic sub-systems have matured and have become automated, they no longer compete, and therefore develop synchronously. In other words, the interaction between different sub-systems may change over time, in what is called moments of self-organization (Caspi, 2010; Spoelman & Verspoor; Van Geert, 2008; Verspoor et al., 2008; Verspoor, Schmid, & Xu, 2012).

A review of previous studies that have examined the development of different measures in L2 learners reveals that it is more likely for the language to develop lexically before it develops syntactically at various stages in wave-like patterns. A cross-sectional study conducted by Verspoor et al. (2012) found the beginner (A1.1 to A1.2 as defined in CEFR) and intermediate (B1.1 to B1.2) learners in the study both developed more in the lexicon, whilst the low-intermediate learners (A1.2 to A2) were found to demonstrate more syntactic than lexical development. Similarly, in a longitudinal study with more advanced students, Caspi (2010) demonstrated that three of her four learners exhibited comparable patterns in the development of the lexicon and syntax, which developed in wave-like patterns with lexical devel-
opment preceding syntactic development. From these studies, two points can be taken. The first is that, when investigating learners of English, we should assume wave-like patterns for different sub-systems, and the second is that it is possible to distinguish shifts in wave-like patterns among the sub-systems at specific times, indicating new stages in the learner’s developmental trajectory.

Taking a usage-based perspective, we do not assume a predetermined path of development, but assume that each learner will have to discover his or her own path through trial and error. In other words, development and change are individually “owned”. This developmental trajectory is inevitably accompanied with variability, as any sub-system that has not fully matured is likely to be affected when attention is drawn to other sub-systems, and variation within this variability will occur among learners (van Dijk et al., 2011). In a longitudinal study following four Chinese learners of English, Larsen-Freeman (2006) found that although the averages of the vocabulary complexity (adjusted type-token ratio) and grammatical complexity (average number of clauses per t-unit) examined demonstrated upward trends, no two learners were alike, and all showed a great deal of variability. Moreover, Verspoor et al. (2012) found that there is more variability and variation among beginners than among more advanced learners.

To be able to investigate development in this manner, we must use dense longitudinal data of individuals, as any averaging or other smoothing of data from multiple learners will conceal the inter-individual variability and interactions between the individual’s sub-systems we are interested in studying. In the current study, we compare identical twins, who live in the same home and have been attending the same classes in school. Traditional twin studies normally compare monzygotic (MZ, or identical) twin pairs with dizygotic (DZ, or fraternal) twin pairs in order to investigate the effect that genetic factors have on language (Segal, 2010; Stromswold, 2006). The current study is not a traditional twin study in that only one pair of MZ twins is being examined. The majority of twin studies have found identical twins to perform more similarly than fraternal twins in linguistic development, validating the emphasis that we place on the identical nature of their genetic makeup in the current study (Stromwold, 2006). In stating that the participants are identical twins, we are not invoking the much-maligned equal environments assumption (Plomin, Defries, McClern, & McGuffin, 2008), which argues that MZ and DZ twins both share equal environments, so any significantly closer developmental patterns found in MZ twins must be due to genetics. Instead, we merely hypothesize that twins who share 100% of their genes and who have been raised in a similar environment are more likely than any other pair of learners to exhibit similar developmental patterns (Hayiou-Thomas, 2008). In this paper, we will investigate their development over several syntactic complexity measures in both their speaking and their writing to see (a) whether these measures develop in speaking before they do so in writing, and (b) whether the twins develop in a similar manner.

The current paper differs from previous studies that deal with dynamic development over time in that it will not focus on the variability of the separate measures,
nor on the specific interactions between these measures. Instead, this paper will focus on the moments of self-organization, i.e. the moments in the learning process where the various sub-systems reorganize and the resulting variability marks a new stage in the developmental trajectory.

6.2 Theoretical Background

6.2.1 Complexity in speaking and writing in L1 and L2

Most of the research on comparing the development of complexity in speaking and writing has been done on the lexicon rather than syntax, but as the findings are explained in terms of cost of production, we assume they may also apply to syntax. Bourdin and Fayol (1994; 1996; 2002) conducted several studies on both children and adults. In simple word-recall tasks, children from the second and fourth grades recalled fewer words in writing than in speaking, which was found to be due to less practice in handwriting and spelling (Bourdin & Fayol, 1994). Bourdin and Fayol (1996) then investigated the recall of sentences produced in speaking and writing tasks and found similar results to the first study: children performed less proficiently in writing than in speaking. In order to move beyond the level of merely recalling oral and written language, Bourdin and Fayol (2002) then conducted a follow up study which tested more complex production processes (such as sentence construction, lexical choice, and expression of meaning) in two conditions. Educated L1 adults were presented with sets of words, some semantically linked and others not semantically linked. The participants were asked to produce sentences with these sets of semantically linked or unlinked words in both the oral and written mode. The recall of the presented words was equal in both tasks, and the number of grammatically correct sentences showed no difference between the tasks. However, the two tasks differed in the conceptual domain- in oral language, there is more elaboration of ideas and more construction of a coherent framework between the semantically unlinked words than in written language. Bourdin and Fayol (2002) therefore conclude that written production is more costly than oral production, even in adults. Interpreting their results in the framework of capacity theories in production (Fayol, 1999; McCutchen, 1996; 2000), which is in line with our own dynamic view of limited resources and possible competition for attention, they suggest that the management of the writing of texts demands a certain degree of attention. In addition, writing texts with sets of weakly associated elements demands an extra capacity to maintain and organise the information in the working memory and/or the ability to adopt a strategic approach to the composition of texts. Apparently, the residual capacity needed for writing is insufficient as the cost of this elaboration must be added to that of production.

Conversely, the situation may be quite different with the L2, and learners may develop their speaking and writing skills at the same time, or in some cases the
written language is acquired before the spoken language. Milton and Hopkins (2006) found that L2 learners of English have a smaller oral English vocabulary size (M = 2260) than written English vocabulary size (M = 2655). They also found that the oral and written vocabulary size varied with English proficiency: learners of lower English proficiency levels tended to have better phonological vocabulary knowledge than orthographic vocabulary knowledge, whilst learners of higher English proficiency levels had more orthographic vocabulary knowledge. However, the sequence of developing these two types of vocabulary knowledge is related to the input. Hudelson (1984) points out that instructed learners in particular who are first exposed more to the written language than to the oral language may be able to read and write but may not be fully proficient in speaking.

In the field of L2 development, there are not many studies that have actually compared the development of oral and written language. This may be due to the fact that these skills, whilst they utilize the same linguistic resources (Levelt, 1989) are in fact very different, both in terms of how they are produced and how they appear. However, for each mode of language, many complexity indices, both lexical and syntactical, have been widely investigated and defined (Bulté & Housen, 2012; Ortega, 2003; Polio, 1997; Quinn & Nation, 2007; Wolfe-Quintero, Inagaki, & Kim, 1998).

Different complexity measures were applied in studies that compare the L2 development in oral and written language and the findings are in line with what Bourdin and Favor’s studies found. There are two studies that have focused on lexical development and one study that focused on specific types of syntactic development. Yu (2010) compared the lexical diversity with the measure D (Malvern & Richards, 2002) of the compositions and interviews of twenty-five advanced learners of English. Contrary to what was expected, the written language, which allows learners more chances to plan and organize their production, did not have a higher D than the oral language.

Chan, Lowie, & de Bot (2014) subsequently compared two lexical measures: lexical diversity with the measure “D” and lexical difficulty with “V-size” (which counts the number of words in different frequency bands and generates an estimated value of difficulty of word use) of the oral and written language of the same pair of identical twins (beginner learners of English) as in the current study. In contrast with Yu (2010), Chan et al. (2014) found that the learners had a lower lexical diversity (D) in their oral language than in their written language, but they did not have a lower lexical difficulty (V-size) in their oral language than in their written language. Chan et al. (2014) concluded that the difference in the linguistic performance between oral and written language was most likely due to the extra time allowed on task in the writing condition. In other words, the learners had more time to think during writing about which words they could use, and were therefore more likely to use a greater variety of words.

In syntactic measures, Dykstra-Pruim (2003) observed university learners of German during three tasks: an oral narrative task, a written narrative task, and a
written grammar task. Over three semesters, she compared different elements of grammar in both oral and written products. She found that the average number of attempted clauses of the more difficult type (with inversion after a preposed element or verb position at the end of a subordinate clause) per subject was higher in oral language than in written language; however, in the written mode, these clauses were significantly more accurate. Because our study does not deal with accuracy, it is important to note that the oral language was syntactically more complex than the written language. In contrast, both Ellis and Yuan (2005), who used passive phrases as an index, and Robinson (2007), who used infinitival phrases as an index, found that writing has a higher syntactic complexity than speaking, and suggest it may be related to the extra time allowed in writing.

The fact that some studies (Yu, 2010; Dyksta-Pruim, 2003) found oral language to be equally complex or more complex than written language is rather surprising given the fact that in producing written language, the learners have more time than in producing oral language to think about the manner and form in which they wish to express their thoughts. One study that shows that time may indeed be a relevant factor in the production of language is by Yuan and Ellis (2003). Focusing on oral language, they analyzed the effect of planning on complexity. They had three conditions involving both pre-task planning and online planning. Because our study did not control for pre-task planning as the learners started speaking and writing immediately when they were reminded to do their oral or written task, we will ignore the pre-task planning condition in our study. However, for the current study, the differences between the two other conditions (with or without online planning) are relevant. In the online planning condition, there was no time limit allowing for online monitoring. In the no online planning condition, there was a time limit, which forced the speakers to speed up their production. Yuan and Ellis (2003) found that the speakers who had no time limit produced oral language with greater syntactic complexity than the speakers who had a time limit. The authors concluded that the time allowed for online planning has a positive effect on the syntactic complexity of oral production.

Although we might expect greater complexity in writing than in speaking because of the extra time for planning (Weigle, 2003), Biber, Gray and Poonpon (2011) suggest that oral texts may not necessarily be less complex, as the nature of written and oral complexity is different. Spoken language may actually contain more dependent clauses (which are taken as a sign of complexity in the current study) than written language, which in turn contains more elaborated noun phrases. However, Biber et al. (2011) look specifically at academic written English with elaborated noun phrases, which is of a much higher proficiency level than our subjects will use. As Byrnes, Maxim and Norris (2010) point out, the choice of appropriate complexity measures will be related to the proficiency level of the learners and will be different, especially at the advanced level.

Only a few studies (Yuan & Ellis, 2005; Robinson, 2007), have found more complexity in written than oral production, but these studies focused on isolated measures. In the other L1 and L2 studies discussed so far, the oral language has been
found to be generally more complex than the written language, despite the extra
time for online planning learners often have when producing written language.
Therefore, we will assume that, when looking at more general, holistic complexity
measures such as mean length of T-unit and the amount of subordination, we will
find a difference in favour of oral production rather than written production. We
assume that if the tasks are similar in topic and in the amount of time allowed
(no time pressure), the oral language production will be more complex. Online
planning or editing will be possible for both types of production, however writ-
ten production may be more demanding of cognitive resources (Bourdin & Fayol,
2002). In order to test this assumption, the current study investigates how L2 oral
and written language develops in syntactic complexity over time (8 months), and
whether the two learners show similar developmental paths.

6.2.2 Complexity measures in language development

To investigate the development of language production, there were calls from re-
searchers including Hakuta (1976) and Larsen-Freeman (1976) to emphasize and
investigate the construction of developmental indices. Wolfe-Quintero et al. (1998)
summarized one of the most comprehensive and thorough studies on the devel-
opmental indices, covering three major dimensions: complexity, accuracy, and
fluency. Complexity (C), accuracy (A), and fluency (F) began to be used as a
measurement of learner performance (Ortega, 2003) and served as three major
criteria in understanding the development of language production.

Complexity has been extensively investigated in several domains. Bulté and
Housen (2012) presented a taxonomic model of L2 complexity which includes three
components: propositional complexity, discourse-interactional complexity, and lin-
guistic complexity. The propositional complexity is the amount of information
(number of ideas) present in the production (Ellis & Barkhuizen, 2005); discourse
interactional complexity (which only exists in learners’ dialogic discourse) refers
to the number and the types of exchanges that learners engage in; and linguistic
complexity refers to the degree of elaboration of the productions in width and in
depth. In our study, we focus on linguistic complexity.

Within the dimensions of linguistic complexity, syntactic complexity has received a
lot of attention due to its inherent sophistication. Syntactic complexity represents
“the range of syntactic forms that surface in language production and the degree
of sophistication of such forms” (Ortega, 2003, p. 493). Several cross-sectional
studies investigated how the measures of syntactic complexity could potentially
distinguish groups of differing proficiency levels in their second language produc-
tion (e.g. Bardovi-Harlig & Bofman, 1989; Larsen-Freeman, 1978). Longitudinal
studies have investigated how to track the writing development in syntactic com-
plexity over time (e.g. Ishikawa, 1995; Norris & Ortega, 2000) and showed that
mean length of T-unit, mean length of clause, clauses per T-unit, and dependent
clauses per clause were more satisfactory than other metrics. However, as argued
by Norris and Ortega (2009) across 16 studies in task-based language learning research, these measures did not consistently show that they could distinguish adjacent English proficiency levels, due in part to different research designs, different L1s of the participants, and different formats of the measurements. It is difficult to say what metrics should be implemented for what level of English proficiency, in what type of language learning, and for which L1. In line with these findings and a dynamic usage-based perspective, Bulté (2013) summarizes the ways an L2 system may develop as follows:

The L2 system of a learner can develop (expand, grow) in many different directions (i.e. along many different dimensions and in many different subsystems). Words can be added to the lexicon (more items, more variety), different meanings or functions of words can be learned (more components within an item, and more relations), more specific words for restricted contexts and situations (higher sophistication), independent clause and simple sentence structures and word order, coordination and subordination of clauses (horizontal and hierarchical relationships), subordination within phrases, verb paradigms, ... All of these changes make the L2 system of a learner more complex, and this is (or at least can be) also reflected in a more complex L2 production. (p.100)

To thoroughly represent syntactic complexity in second language writing, the measures should sufficiently gauge different dimensions of complexity and provide distinctive characteristics with as little overlap as possible to avoid redundant measures. Therefore, we should only include metrics which represent independent traits and which do not correlate highly with other metrics. A range of distinctive measures were deployed by Norris & Ortega (2009) with five central foci: length of production units, amount of subordination, amount of coordination, sophistication and acquisitional timing of grammatical forms used in production, and total frequency of use of certain forms considered to be sophisticated. The length of production is generally calculated by dividing words by a production unit (e.g. clause, sentence, T-unit) and has been widely used in child language acquisition since Brown (1973). The amount of subordination is computed by dividing the number of instances of a subordinate clause by a production unit, for example the mean number of dependent clauses per T-unit. The amount of coordination, proposed by Bardovi-Harlig (1992), is a metric determined by dividing the number of coordinate clauses by the total number of combined clauses. However, according to Bulté and Housen (2012), this coordination index is not a pure coordination measure but rather an index of subordination.

Different metrics demonstrate different developmental trajectories at different proficiency levels of English learners. However, the mean length of the T-unit and the degree of subordination have proven to correlate strongly with different levels of proficiency over time, especially at the lower levels, and will be used in the current study. For our third dimension of complexity, we wanted to find a measure that would tap into another dimension of complexity and found that the number of coordinate phrases did not significantly correlate with either the mean length of the T unit or the degree of subordination. We will assume that coordinate
phrases are rather easy to form and may therefore occur earlier than subordinate clauses. In addition, Bulté (2013) argues that it is important to include both clausal subordination and phrasal complexity measures in order to properly assess L2 complexity, as they are both measures of syntactic complexification which do not occur together, nor develop in parallel.

In our study, we will attempt to determine in which mode (oral or written) the language of beginner learners of English (CEFR A2 to B1) demonstrates more complexity first in terms of the three different dimensions of syntactic complexity, and whether the identical twins show similar developmental patterns in these measures.

### 6.3 The study

#### 6.3.1 Participants

Gloria and Grace (not their real names) are two female identical twins, aged 15 at the time of the study. For ten years, they attended school in Taiwan in the same English class with the same English teacher, where English classes were taught in Chinese with a focus on grammar. In other words, until the current study began, they had mainly received only written input in English. At the beginning of the study, they had a very similar English proficiency level (See Table 6.1) as measured by the General English Proficiency Test (GEPT) (Wu, 2012).

<table>
<thead>
<tr>
<th></th>
<th>Grace</th>
<th>Gloria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening (120)</td>
<td>112</td>
<td>112</td>
</tr>
<tr>
<td>Speaking (100)</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Reading (120)</td>
<td>108</td>
<td>105</td>
</tr>
<tr>
<td>Writing (100)</td>
<td>88</td>
<td>82</td>
</tr>
</tbody>
</table>

Table 6.1: Score of GEPT: Gloria and Grace

According to an informal personality test, the big five test\(^1\), they also had similar personalities: they were both open to new experiences (creative, curious, and original); they were sociable, friendly, and talkative; and they were nervous, highly-strung, and worried.

\(^1\)The big five personality Test: http://www.outofservice.com/bigfive/
6.3.2 Procedure

The study lasted for eight months. During the study, the researcher asked the participants to obtain extra aural exposure through the media in three stages: low, high and low. According to the self-reports in their diaries, they obtained about 2 to 5 hours per week of extra input until data point 20; 5 to 15 hours per week until data point 56; and again 2-5 hours per week until the last data point. Frequent informal Facebook contact with the researcher about the content of the movies confirmed the statements in the self-reports. The amount of exposure was manipulated for another study on the effect of exposure on vocabulary (Chan et al., in press).

During this time, the participants produced oral and written texts approximately three times a week, which was usually on Friday, Saturday, and Sunday. For each participant, 100 oral texts and 100 written texts were gathered. The topics, selected from the list of standard TOEFL tests by the researcher, were of the same genre. All the topics were presented to the two participants at the beginning of the study. Examples of the topics for writing and speaking are as follows (See Appendix C)

Example of a speaking topic:

*Which of the following statements do you agree with? Some believe that TV programs have a positive influence on modern society. Others, however, think that the influence of TV programs is negative. What TV programs have a positive influence? Why? What TV programs have a negative influence? Why?*

Example of a writing topic:

*Do you agree or disagree with the following statement? With the help of technology, students nowadays can learn more information and learn it more quickly. Use specific reasons and examples to support your answer.*

In order to motivate and remind the participants to obtain extra exposure to English and to do the speaking and writing tasks, the researcher created a private group on Facebook for the project, which only the researcher, the participants, and the parents had access to. The researcher reminded the twins every week to record themselves and to write the texts. Recordings were sent through email, and the written texts were posted in the Facebook account. To keep the participants motivated in the study, the researcher reacted to the content of each text, but no corrective feedback on form was given on either the oral or the written texts.

All texts were prepared for automatic processing in Lu’s automatic syntactic complexity analyzer (Lu, 2010). The analyzer is designed to investigate the syntactic
complexity in writing in second language acquisition and fourteen indices of syntactic complexity are calculated (see p. 479).

Lu (2010) provides clear descriptions of sentences, clauses, dependent clauses, T-units, complex T-units, and complex nominals (p. 481-484). For our study, we used length of T-units, dependent clauses and coordinate phrases. A T-unit is one main clause plus any subordinate clause or nonclausal structure that is attached to or embedded in it (Hunt, 1970, p.4). T-units also include sentence fragments punctuated by the writer (Bardovi-Harlig & Bofman, 1989; Tapia, 1993). Because it is only possible to search parse trees one by one using Tregex, it was specified that a T-unit can only occur within a sentence punctuated by the writer (Homburg, 1984; Ishikawa, 1995). A dependent clause is defined as a finite adjective, adverbial, or nominal clause (Cooper, 1976; Hunt, 1965; Kameen, 1979). Non-finite verb phrases are excluded in the definition of clauses (e.g. Bardovi-Harlig & Bofman, 1989). As far as coordinate phrases are concerned, only adjective, adverb, noun, and verb phrases are counted in coordinate phrases (Cooper, 1976).

Among the 14 syntactic complexity measures generated from the automatic tool, there are three main categories employed extensively in second language acquisition: length of production unit, subordination, and coordination (Lu, 2010, p. 479; Norris & Ortega, 2009, p. 559).

Lu tested the reliability of the analyzer on essays written by Chinese learners of English at the university level. As these learners are advanced writers, their writings mostly contain errors of word use (e.g. collocations) rather than errors of grammatical completeness. Thus, the errors existing within the boundaries of a unit of a structure led to little misanalysis by the parser or little misrecognition of the production units. However, as Lu points out, writing samples of beginner learners which still contain errors of syntactic completeness should be carefully preprocessed. Therefore, all oral texts (each about 200 words in length) were first transcribed by the researcher. To avoid redundancy in the oral production, filled pauses (e.g. mm and er), dysfluencies (e.g. repetitions, restarts, repairs), and utterances that did not involve linguistic meaning or form (e.g. laughter) were left out. Then both the oral and written texts were preprocessed for the analyzer, mainly to be able to count the units, meaning that incorrect punctuation and incomplete sentence structures were corrected. (See Appendix for examples.) All other errors remained to keep the data as original as possible. After the preprocessing, the text files were submitted one by one to the automatic processing tool to obtain the values of the syntactic measures for observation.

For the present study, we selected the mean length of T-unit (MLT) as a general complexity measure and dependent clauses per T-unit (DC/T) as a more specific complexity measure. These two measures partly overlap as T-units tend to become longer when dependent clauses are added. Therefore, in order to find a third measure that would tap into a different, less-overlapping dimension of complexity, we ran correlations with the remaining 12 syntactic complexity measures and selected the one with the lowest correlation to the MLT, which was the number
Syntactic development in writing & speaking

of coordinate phrases per T-unit (CP/T). Based on Verspoor et al. (2012), we assumed that coordinated phrases show less development over time and are less indicative of linguistic complexity than the other two measures.

6.3.3 Analyses

To test our hypotheses, both traditional and dynamic statistical analyses are used. For overall differences, the oral and written texts were first compared with an independent two-sample T-test. To investigate the learning stages, the data was first visually inspected for general patterns. Then to find the exact moments of reorganization, the data were analyzed with the hidden Markov model (HMM).

As pointed out by Van Dijk et al. (2011), visual inspection is first needed to get a feel for the data and to stipulate hypotheses that can later be tested in modeling. Because the raw data of the measures has different numerical ranges (MLT from about 7 to 10; DC/T from about 0.2 to 3; CP/T from 0 to about 0.8), the data was first normalized to 0-1 to be able to observe the relationships between the measures on a common scale. Because the high degree of variability for each measure obscures any general, discernible pattern, a moving average trend line was added of 6 data points. These trend lines were visually inspected to see if there were shifts in the configuration of the measures. Discernible shifts point to possible boundaries to be found by the modeling.

Most longitudinal dynamic studies so far have aimed at investigating the trajectory of separate measures (cf. Larsen-Freeman, 2006; Verspoor et al., 2010; Verspoor et al. 2008) or the relations between these measures (precursor, competitive, supportive) (Caspi & Lowie, 2013; Chan, Lowie, & de Bot, in press). For example, the modeling done by Lowie et al. (2011) on such data has been done as follows: observe the empirical data by means of visual inspection, hypothesize explicit parameters for initial conditions, precursor relationships and other degrees of interaction among the measures, run simulations, and do manual adjusting of the parameters until high correlations among the observed and modeled data are found, which then can confirm the hypothesized interaction among the measures. These studies have revealed important characteristics of language development such as individual differences, the meaning of variability, and trade-off effects. The current study takes the dynamic analyses in a somewhat different direction, both in aims and type of modeling. The aim is to find moments of self-organization, the moments that the interaction among the measures shifts and takes on a new configuration, indicating the beginning of a new stage.

However, the more observed measures that are involved, the more difficult it is to visually observe the changes in the interaction of the measures and then model them as Lowie et al. (2011) suggested. Therefore, we used a dynamic model, an unsupervised Hidden Markov Model (HMM), which detects the patterns based on the data. In this model, the string of data (or the value) of all measures is first
analyzed to detect patterns of change. After that, the model finds the best stage sequence of the changes and the data point where a “shift” in the complex system occurs, indicating the boundaries of the stages.

There are two types of HMMs, supervised and unsupervised. The type of model used depends on whether there are clearly identifiable labels for the categories. The supervised HMM typically deals with correctly tagged texts such as with part of speech (PoS), where probabilities of certain sequences of PoS are calculated. For example, when you see the article “the”, there is a 40% chance that the next word is a noun. CLAWS (http://ucrel.lancs.ac.uk/claws/) pioneered the HMM-based part of speech tagging (Garside, 1987; Garside & Smith, 1997). Unsupervised models, which we use in the current study, deal with data strings that have no pre-set identifiable labels, in our case, values expressing differences in degree. In the unsupervised HMM, the values are visible but no knowledge of the probability matrices for initial, transition, and observation data is presupposed in the model. A large number of studies in the field of speech recognition use the unsupervised HMM (e.g. Novotney, Schwartz & Ma, 2009; Park & Glass, 2008; Zhang & Glass, 2009).

With the HMM, it is possible to try out different numbers of learning stages and explore which number of stages reveals changes over time best; however, the current study postulated three stages in the modelling for two reasons. We visually inspected the interaction of the measures in our charts and observed three to five stages, depending on the mode (speaking or writing) of the learner. But more importantly, there is a minimum number of data points for a stage required: the number of data points in each stage must be minimally two to three times the number of observed measures. As we worked with six measures (three spoken and three written), there had to be a minimum of 12 to 18 data points in each stage, but preferably more. Therefore, we decided on three stages.

The software was based on Chan and Lee (2013), and one of the authors programmed our software in Perl for a Linux environment specifically for the current study. There are two major algorithms involved in the process of looking for the highest probability. The Baum-Welch algorithm, also known as the “expectation maximization” (EM) algorithm, (Rabiner, 1989) aims to find the most likely changes of the learning stages. It finds the transitional probabilities, the emission probabilities, and the initial probability. The Viterbi algorithm (Ryan & Nudd, 1993) then determines the best stage sequence.

The six observable measures, the MLT, DC/T, and CP/T at each data point for both writing and speaking, were used to train the model. In the modelling, the raw data was used. 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 or state 3, and state 3 can only transit to state 3. The parameters were adjusted to find the best model with the Baum-Welch algorithm (Rabiner, 1989). The training was discontinued when the model converged, i.e. when further iteration resulted in no significant change in the model. After obtaining the set of parameters, the single best stage
sequence was calculated with the Viterbi algorithm (Ryan & Nudd, 1993).

The output in our software program informs us of (1) the data points at the beginning and end of each stage, (2) the means of the measures at each stage, and (3) the covariances between measures at each stage.

### 6.4 Results

The mean of Gloria’s oral language was 199.4 words (SD = 34.2) per text; the mean of Gloria’s written language was 185.3 words (SD = 48.4) per text. The mean of Grace’s oral language was 200.6 words (SD = 38.0) per text; the mean of Grace’s written language was 153.7 words (SD = 54.3) per text.

The traditional statistical analyses by means of T-tests (Table 6.2) indicate that for Gloria, all three measures of syntactic complexity in speaking were significantly higher than those in writing, but that this was not the case for Grace. In Grace’s data, MLT and CP/T in speaking were significantly higher than those in writing, but the DC/T in speaking and writing was almost the same.

<table>
<thead>
<tr>
<th></th>
<th>MLT (speaking)</th>
<th>MLT (writing)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloria</td>
<td>Mean=13.2</td>
<td>Mean=10.0</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Sd=3.1</td>
<td>Sd=1.2</td>
<td></td>
</tr>
<tr>
<td>Grace</td>
<td>Mean=14.2</td>
<td>Mean=12.9</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Sd=3.6</td>
<td>Sd=3.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>DC/T (speaking)</th>
<th>DC/T (writing)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloria</td>
<td>Mean=0.9</td>
<td>Mean=0.5</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Sd=0.4</td>
<td>Sd=0.2</td>
<td></td>
</tr>
<tr>
<td>Grace</td>
<td>Mean=0.9</td>
<td>Mean=0.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sd=0.4</td>
<td>Sd=0.4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CP/T (speaking)</th>
<th>CP/T (writing)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloria</td>
<td>Mean=0.2</td>
<td>Mean=0.1</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Sd=0.2</td>
<td>Sd=0.1</td>
<td></td>
</tr>
<tr>
<td>Grace</td>
<td>Mean=0.3</td>
<td>Mean=0.2</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Sd=0.2</td>
<td>Sd=0.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2: Mean of each variable: Gloria and Grace

To analyze change over time, the raw data was plotted in Excel with a moving
average trend line of six data points to visually inspect the data. Because it is difficult to see the interactions occurring between 6 measures, we first examined the data in each mode for each learner separately before we combined them for each learner. We added arrows where we observe a quantitatively different configuration of the measures that might suggest a shift. We looked for both changes in the measure over time (clear rise or fall) and changes in the interaction among the measures (where they cross). However, we must keep in mind that the trend lines are moving averages of 6 data points, so they cannot specify the exact data point of these changes.

In Gloria’s writing (Figure 6.1a), we observe such shifts around data point 15, 27 and 79, which mainly involves CP/T, which goes down at each of these points. In Gloria’s speaking (Figure 6.1b), we observe such moments around data points 15, 27, 59, and 79, which also involve CP/T, which goes down in each case except at data point 79, where it actually goes up. In the combined writing and speaking graph (Figure 6.1c), it is difficult to identify the specific measures, but shifts in configuration seem to occur around data points 27, 45, and 79.

To quantitatively determine the exact points of the shifts in the configuration of the six measures in Gloria’s data, we made use of the Hidden Markov Model. Table 6.3 presents the results. In the left columns are the three learning stages as found by the HMM and the next columns show the mean for each measure during that learning stage in the speaking and writing conditions.

As Table 6.3 shows, Gloria’s shifts seem to have occurred after data points 27 and 78, both of which had also been visually observed in the writing and speaking and the combined measures in the Excel graphs. Note also that Gloria’s MLT, DC/T, and CP/T in speaking were consistently higher than those in writing in all three learning stages.

In Grace’s measures (Figure 6.2), the shifts in configuration are overall visually less clear than in Gloria’s writing. In the writing measures (Figure 6.2a), we observe shifts around data point 19, where the CP/T crosses and goes down, 33

<table>
<thead>
<tr>
<th>Gloria</th>
<th>Mean of MLT (speaking)</th>
<th>Mean of MLT (writing)</th>
<th>Mean of DC/T (speaking)</th>
<th>Mean of DC/T (writing)</th>
<th>Mean of CP/T (speaking)</th>
<th>Mean of CP/T (writing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage1 (1-27)</td>
<td>13.9</td>
<td>9.6</td>
<td>1.0</td>
<td>0.5</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Stage2 (28-78)</td>
<td>12.4</td>
<td>10.0</td>
<td>0.9</td>
<td>0.6</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Stage3 (79-100)</td>
<td>14.14</td>
<td>10.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 6.3: Mean of each variable in three stages: Gloria
Syntactic development in writing & speaking

where there is a clear difference in the distance between the trend lines, from converging to diverging, 49, where the trend lines start diverging from each other again, and at 71 where they converge again. In Grace’s speaking (Figure 6.2b), we observe such moments around data points 23 where the trend lines converge and 69 where the CP/T trend line crosses again and goes up. In the combined writing and speaking graph (Figure 6.2c), it is very difficult to visually discern shifts in the configuration, but the HMM analysis (see below) points to shifts in configuration after data points 31, where the CP/T-S makes a clear dip below all the other measures, and after 71 where the trend lines seem to converge before they diverge again.

To quantitatively determine the exact points of the shifts in the configuration of the six measures in Grace’s data, we made use of the HMM. Table 6.4 presents the results. In the left columns are the three learning stages as found by the modeling and the next columns show the mean for each measure during that learning stage in the speaking and writing conditions.
As Table 6.4 shows, Grace’s largest shifts appear to occur after data points 31 and 71, neither of which had been found exactly in the data, but close enough as our visible inspection is based on trend lines that were averaged over 6 data points. These shifts occur at different times than for Gloria and suggest that Grace’s second stage of development was shorter than Gloria’s. Note also that Grace’s MLT and DC/T were higher in speaking than in writing in the first learning stage, but lower in the second and the third learning stages. Grace’s CP/T in speaking was higher than that in writing in the first and the second learning stages, but lower in the third one. In Grace’s data, there seems to be a shift from higher syntactic complexity in speaking at the beginning to higher syntactic complexity in writing in the third learning stage.

Figure 6.2: Grace’s measures with a moving average trend line of 6 data points.
Syntactic development in writing & speaking

<table>
<thead>
<tr>
<th>Grace</th>
<th>Mean of MLT (speaking)</th>
<th>Mean of MLT (writing)</th>
<th>Mean of DC/T (speaking)</th>
<th>Mean of DC/T (writing)</th>
<th>Mean of CP/T (speaking)</th>
<th>Mean of CP/T (writing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage1 (1-31)</td>
<td>15.6</td>
<td>11.0</td>
<td>1.0</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>Stage2 (32-70)</td>
<td>13.7</td>
<td>13.9</td>
<td>0.9</td>
<td>1.0</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Stage3 (71-100)</td>
<td>13.4</td>
<td>13.5</td>
<td>0.8</td>
<td>0.9</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Table 6.4: Mean of each variable in three stages: Grace

<table>
<thead>
<tr>
<th>Gloria</th>
<th>Correlation of MLT in writing and speaking</th>
<th>Correlation of DC/T in writing and speaking</th>
<th>Correlation of CP/T in writing and speaking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage1 (1-27)</td>
<td>0.1 (p=0.8)</td>
<td>-0.1 (p=0.7)</td>
<td>-0.1 (p=0.6)</td>
</tr>
<tr>
<td>Stage2 (28-78)</td>
<td>-0.1 (p=0.6)</td>
<td>0.1 (p=0.4)</td>
<td>-0.1 (p=0.3)</td>
</tr>
<tr>
<td>Stage3 (79-100)</td>
<td>-0.3 (p=0.2)</td>
<td>0.4 (p=0.1)</td>
<td>0.5 (p=0.02*)</td>
</tr>
</tbody>
</table>

Table 6.5: Correlations of the variables in writing and speaking in three stages.

Table 6.5 shows the correlations (normalized covariances obtained from the HMM) of each measure per learning stage for speaking and writing. It will be assumed that if the correlations between speaking and writing measures are positive, the constructs develop synchronously in speaking and writing. If the correlations are negative, they develop asynchronously. If there is no meaningful relation, the relation will be regarded as neutral. As indicated by the trend lines, different developmental patterns are found for the twins. Gloria’s MLT in speaking and writing developed synchronously in the first stage and asynchronously in the last two stages, whilst Grace’s MLT in speaking and writing developed asynchronously in the first two stages but synchronously in the last stage.

Gloria’s DC/T in speaking and writing developed slightly asynchronously in the first stage but synchronously in the last two stages, whilst Grace’s DC/T in speaking and writing developed somewhat synchronously in the first stage but asynchronously in the last two stages.

Gloria’s CP/T in speaking and writing developed slightly asynchronously in the
first stage but synchronously in the last two stages whilst Grace’s CP/T in speaking and writing developed asynchronously in all three stages.

We may conclude that the directions of the correlations of the MLT and DC/T, except for MLT in stage 2, in speaking and writing were in opposite directions for the twins.

6.5 Discussion

This study explored two questions: (1) as far as syntactic complexity is concerned, does it develop first in spoken or in written language; and (2) do the two learners in our study develop in similar ways?

As far as the first question is concerned, we hypothesized that syntactic complexity would develop in speaking first and then in writing. This hypothesis was based mainly on the findings of previous studies (e.g. Dykstra-Pruim, 2003), and was formulated despite our knowledge that our learners had been exposed more to the written language than oral language before the study and that writing gives more time for online planning (Yuan & Ellis, 2005). This hypothesis proved to be correct. The T-tests reveal that the majority of means of the measures for the variable syntactic complexity are significantly higher in speaking than in writing.

Yuan and Ellis (2005) had participants with the same L1 as ours and they found that when their participants were given more time (no time pressure) to speak, they produced more complex sentences. In our study, it is possible that the participants took less time for online planning in their speaking than in their writing, but our results in terms of syntactic complexity suggest that the speaking task provided the learners with enough time to plan online.

Therefore, our findings are in line with other studies. Bourdin and Fayol (2002) demonstrated that L1 children as well as adults tend to elaborate more in oral language than in written language, mainly because writing seems to come at an extra processing cost, meaning there are fewer resources to manage the writing process. Moreover, another feature in common with Bourdin and Fayol (2002) is the amount of revision taking place in the written language. They argue that because they did not allow learners to revise the written language, the quality of the written language may have suffered due to insufficient time to manage the processes in written production. In our study, the revision was allowed in both tasks. The learners were able to go back to rephrase what they wished to say or to edit what they wrote. However, judging by their written products, the two learners did not seem to make use of the opportunity to revise their writing and they seemed to write as if they were speaking, i.e. not adjusting their production to better suit a written form. The written texts had many grammatical errors, and the focus seemed to be mainly on using the right words instead of composing
complex and accurate sentences.

This leads to another possible explanation of our findings. The language proficiency of these two learners is rather low so that in speaking they have difficulty retrieving the words they seek to use on time. Therefore, in order to fully express what they wish to explain, they circumscribe the item, resulting in longer sentences, especially with dependent clauses and coordinate phrases. This is in line with Chan et al. (2014), who found greater lexical diversity in writing than in speaking. During writing, there may be more time for learners to search for the right lexical item so they can express one idea more concisely. However, this idea is not quite in line with Milton and Hopkins (2006), who found that learners of lower English proficiency levels tend to have better phonological vocabulary knowledge than orthographic vocabulary knowledge.

The idea that the low intermediate learners’ lexical use rather than their syntactic complexity benefits more from the extra time allowed for revision in writing is in line with Verspoor et al. (2012), who found that her low proficiency learners tend to focus more on acquiring the lexicon than on using complex sentence constructions. Caspi (2010) also found that the lexicon seems to develop before syntax.

Our second research question concerned the developmental patterns of our learners. As expected from a dynamic usage-based perspective, we saw a great deal of variability within each learner; however, since we controlled for as many variables as we could by tracing identical twins in very similar conditions, we expected less variation in general developmental patterns. General patterns have been found before among less similar subjects. For example, Caspi (2010) found that three of the four learners showed similar sequences in the development of their lexicon and syntax. Van Dijk et al. (2011) showed how different Spanish learners of English went through similar sequences in their acquisition of English negative verb phrases. However, when we tested the development of the measures over time with the HMM, there were some similarities but also clear differences between the learners. Both showed more syntactic complexity in speaking early on, and Gloria’s syntactic complexity in speaking remained stable during the study, but Grace showed a shift to more complexity in writing rather than speaking in the second and the third stage. This finding is also in line with the finding by Chan et al. (2014), where the two girls demonstrate contrasting lexical developmental patterns. A possible explanation for the disparate development of written syntactic complexity between the twins may be found in the competition between lexical and syntactic measures. Whereas one twin focused more on the lexicon, the other focused more on the syntax. Further study will have to examine this relationship.

We may conclude that even identical twins with similar personalities and interests who are exposed to similar input within the same environment may demonstrate different developmental paths. Larsen-Freeman (2006) argues that whilst some variation between learners is attributable to external social factors, variation must also come from the internal restructuring that occurs within the language learning system. From a usage-based perspective, one expects not only that input plays
an important role in shaping a learner's language, but also that each individual has to find his/her own way in figuring out the meaning and use of words or constructions. Identical twins are no exception to this.

6.6 Conclusions

Language development is dynamic, with different sub systems of the language developing at different rates and interacting continuously, which may result in different relations between these sub-systems over time. One of the main contributions of the current study is to have shown a way of modeling the development of the separate measures and interactions among them to discover moments of reorganization of these measures and the emergence of new stages in the developmental process. These stages, when confirmed with the HMM, can be seen to occur when there is an ascertainable overall difference in the ways in which the measures interact with each other, and in the measures themselves. The HMM allowed us to objectively observe and chart the changes of several measures at a time, rather than restricting us to just subjectively observing the overall pattern (i.e. linear growth or decline) of the development.

In using the HMM, the current study establishes a new line of confirming subjective findings within the dynamic usage-based perspective. If we had simply wished to investigate which modality of production was more complex, we would not have needed to use the HMM. However, what the HMM has allowed us to do is to look at the way in which the complexity in writing and speaking changes over time, thereby establishing the study as process-oriented rather than product-oriented.

The analyses demonstrated that within the time of the study, the syntactic complexity in oral language developed sooner than in written language, a pattern also generally found in the L1. This study also found that the time limit advantage of written language over oral language did not seem to result in higher syntactic complexity. The low intermediate learners in fact elaborated their ideas and expressions more in their oral production than in their written production, which resulted in greater syntactic complexity in three dimensions: mean length of T-unit, dependent clause per T-unit, and coordinate phrase per T-unit. However, these syntactic complexity measures do not reveal the depth of the complexity—such as the number of ideas expressed or the coherence of the production.

The analyses also showed that these two very similar learners, identical twins living in the same household with the same schooling and very similar exposure to English, showed contrasting developmental patterns of syntactic complexity. An important point to note here comes from Stromswold (2006):

Researchers who study language acquisition often implicitly assume that when one refers to the role of environmental factors on language development, one
is primarily referring to postnatal, psychosocial factors such as the quantity or quality of adult linguistic input that children receive. If psychosocial factors have a large impact on language development, this would support theories that argue that language development is largely the result of children’s social and language environments (empiricist/emergentist theories). It would also call into question nativist/biological theories that argue that language acquisition is largely the result of children’s innate, biological endowment (nativist/biological theories). (p. 341)

Therefore, we can conclude that, despite being monozygotic twins, the sisters, in displaying contrasting developmental patterns, provide further support to empiricist/emergentist theories of language development. This observation makes us wonder to what extent there actually exist “average” learners who develop in similar manners.