Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study

Abstract

This study explores *ab initio* development of syntactic complexity in a longitudinal corpus of learner German writing from a Dynamic Usage-Based perspective. It contributes to the research on L2 writing complexity by focusing on beginning learners of an L2 other than English (German) and on fine-grained measures of syntactic complexity, operationally defined here as syntactic modification.

The results show that not only ubiquitous global measures of syntactic complexity but also more specific measures, namely frequencies of syntactic modifiers, can serve as developmental indices at beginning L2 proficiency levels. The learners in this study modified their writing from the very onset of language study and the overall size and range of the modification system did not significantly change over four semesters. However, its composition changed continuously and reflected non-linear waxing and waning of different modifier categories. The study confirmed some results from previous cross-sectional research showing that interlanguage development is characterized by a decrease in cognitively easier (e.g., uninflected) categories and an increase in cognitively more difficult (e.g., inflected and clausal) categories. The high variability that was found along with uniform group trends demonstrates the necessity of simultaneous investigations of linguistic development in groups and individuals.

Key words

syntactic modification, L2 writing development, German, low proficiency, dynamic usage-based approach, mixed effects modeling

Highlights

- We explored the *ab initio* development of syntactic modification in a longitudinal corpus of learner German writing.
- The overall size and range of the syntactic modification system did not significantly change over four semesters of study.
- Frequency of uninflected lexical modifiers decreased and frequency of inflected and clausal modifiers increased.
- The use of modifiers varied highly among individuals yet there were several uniform group trends.

1. Introduction

Second language (L2) complexity measures have been used in Second Language Acquisition (SLA) research to investigate learner production with three main purposes: “(a) to gauge proficiency, (b) to describe performance, and (c) to benchmark development” (Ortega, 2012a, p. 128). However, as Ortega points out, in comparison with the first two, the third purpose has rarely been addressed and there is considerably less systematic knowledge of it. Manchón (2012) expresses a similar concern related to the field of L2 writing, calling the

development of L2 writing competencies "an issue of the utmost theoretical, methodological, and pedagogical relevance […] that, surprisingly, has not been systematically approached in the otherwise abundant research in the field" (p. 3). This study and this Special Issue in general aim to address these research gaps by focusing on the development of linguistic complexity in L2 writing. This focus expands and supplements the research line taken up by two other recent JSLW Special Issues: bringing together SLA research (and, in particular, L2 developmental research) and L2 writing research (Connor-Linton & Polio, 2014; Ortega, 2012b; Williams, 2012). Furthermore, this study focuses on L2 beginning German learners and uses finer-grained syntactic complexity measures, thus expanding the empirical knowledge base which has mostly encompassed English as an L2, relatively advanced proficiency levels, and global complexity measures.

This study explores the syntactic modification aspect of linguistic complexity, where modifiers are defined as optional elements describing the property of the head of a phrase (Graesser, McNamara, Louwerse, & Cai, 2004). Modifiers can be considered *par excellence* indicators of structural complexity at the sentence level because they expand the simplest possible agent-action-(object) pattern (Graesser et al., 2004; McNamara, Crossley, & McCarthy, 2010). Thus, expansion of modification in learner language would fall squarely under Foster and Skehan's (1996) definition of development in syntactic complexity as "progressively more elaborate language" and "a greater variety of syntactic patterning" (p. 303). Although many studies have used the frequency of selected modifiers as L2 complexity measures (see the overview below), investigations of the modifier system in its entirety are virtually non-existent (see, however, Hirschmann, in press, and Hirschmann, Lüdeling, Rehbein, Reznicek, & Zeldes, 2013). This study addresses this gap and explores a variety of modifiers in L2 writing data using Learner Corpus Research (LCR) methods for data extraction and coding.

The study is grounded in the Dynamic Usage-Based approach (DUB, Langacker, 2008; Verspoor, Schmid, & Xu, 2012) that is informed by the Dynamic Systems Theory (DST, van Geert, 2008). In this approach, L2 development is seen as a dynamic process, in which regular growth stages are modulated by a complex variation within and among individuals and by the continuous waxing and waning (Larsen-Freeman, 2006) of different interrelated aspects of the interlanguage system. The DST and DUB approach has been applied in a number of recent L2 writing investigations (e.g., Verspoor et al., 2012; Verspoor, Lowie, & van Dijk, 2008; Verspoor & Smiskova, 2012) and allowed researchers to show that variability drives development and thus merits a more prominent place in L2 research. The present study aims to contribute to this small but growing body of research by zooming in on the dynamics within the syntactic modification system in the interlanguage of beginning L2 German writers and by exploring both group trends and individual developmental trajectories. To account for this complex developmental picture, we use mixed effects modeling methods, which have just begun to gain popularity in SLA research (Barkaoui, 2014; Cunnings, 2012). These methods are uniquely suited for longitudinal studies because they allow the estimation of the mean developmental trajectory as well as variation and covariation among individual trajectories.

2. Study background

2.1. Defining L2 complexity

The concept of "complexity" is itself "complex" due to its polysemous and

multidimensional nature, which has led to the absence of its consistent definition in SLA research. Bulté and Housen (2012) conducted a thorough analysis of this notion's use and found that, despite popularity, there has been widespread confusion between complexity and the related notion of difficulty, as well as between the formal and functional aspects of complexity. Bulté and Housen (see also Ortega, 2012a, and Pallotti, 2015, for a similar argument) argue that this confusion often leads to interchangeable use of terms representing different underlying constructs (e.g., linguistic complexity and cognitive difficulty) as well as to circular argumentation (e.g., linguistic structures are considered complex if they are used by more proficient learners, and learners are considered more proficient if they use such structures).

Responding to recent calls for clear definitions of complexity in research, we adopt Bulté and Housen's (2012) carefully structured taxonomy of complexity. First, as most of the contributions to this Special Issue, this study is concerned with linguistic complexity of learner writing, or "language complexity in objective, quantitative terms as the *number* of discrete components that a language feature or a language system consists of, and as the *number* of connections between the different components" (Bulté & Housen, 2012, p. 24, emphasis in the original).[1] More specifically, this study focuses on the grammatical component (as opposed to the lexical component) of linguistic complexity. Second, in line with Bulté and Housen, we distinguish three levels for the investigation of complexity development in L2 writing. At the theoretical level (that is, at the level of cognitive constructs), we are primarily interested in the systemic aspects of grammatical complexity, defined by Bulté and Housen as "elaboration, size, range, variation, 'breadth' of L2 grammar" (p. 27). At the observational level (that is, at the level of surface manifestations of complexity in learner writing), we focus on how this systemic complexity is realized in grammatical diversity and sophistication of sentences, clauses, and phrases. At the operational level (that is, at the level of measures), we employ frequency measures, namely ratios of syntactic modifiers as a function of total word counts or counts of the modified elements. By selecting these measures, we aim to fill a considerable gap in L2 writing complexity research that has not focused on the syntactic modification system, as evidenced in the literature review below.

2.2. L2 complexity as a developmental dimension

L2 complexity research originates from Skehan's (1989) three-part model of L2 proficiency encompassing complexity, accuracy, and fluency (CAF). Many studies have aimed to find general proficiency indicators in learner-produced oral and written texts using these three dimensions (for reviews, see Housen & Kuiken, 2009; Housen, Kuiken, & Vedder, 2012). Early studies argued that all measures increase with developing proficiency. In other words, the more proficient the learners are, the oral and written L2 texts they produce per unit of time will be longer (fluency), with more elaborate and varied syntactic and lexical units (complexity), and fewer errors (accuracy). In the field of L2 writing, Wolfe-Quintero, Inagaki, and Kim (1998) have provided the first research synthesis of CAF studies and CAF measures. They concluded that some measures correlated with L2 proficiency levels and writing quality better than others but also that differences in research design had made the comparison of results across studies problematic. They also pointed out that most of the research had been cross-sectional. In other

[1] This definition is also in line with Pallotti's (2015, p. 117) "simple view of linguistic complexity" as structural complexity.

words, developmental inferences had been made based on comparisons of different groups of learners at different proficiency levels instead of tracing the development of the same learners over time.

Since the publication of Wolfe-Quintero et al.'s (1998) seminal study, the body of research on CAF has been constantly growing, and by now it is an established fact that correlations between the length of L2 study and the three proficiency dimensions are not always positive and not always linear and that CAF indicators are interrelated with various other factors including measure type, task type, proficiency stage, individual learner differences, and many others (for overviews, see Housen & Kuiken, 2009, Housen, Kuiken, & Vedder, 2012, and Norris & Ortega, 2009). Furthermore, following Larsen-Freeman (2006), longitudinal studies exploring CAF in L2 writing started emerging (Gunnarsson, 2012; Spoelman & Verspoor, 2010; Verspoor et al., 2012; Zhang & Lu, 2013). These studies have enriched the research landscape by demonstrating that group trends need to be supplemented with illustrations of individual variation to provide a differentiated picture of L2 writing development.

Whereas the abovementioned studies explored the development of complexity in L2 writing in relation to other proficiency measures (mostly accuracy), other studies have focused on interrelation of components within the system of complexity. Verspoor (and colleagues) employed dynamic research methods in a series of longitudinal studies exploring variability of grammatical and lexical complexity in learner writing. Specifically, Verspoor et al. (2008) tracked an advanced learner of English and Spoelman and Verspoor (2010) a beginning learner of Finnish, both having Dutch as their L1, over three years. The first study found that the lexical and grammatical interlanguage subsystems were competing with one another at the beginning of the observation period but developed in parallel toward the end. The second study showed a competition between the development of sentence complexity and of noun phrase complexity.

Byrnes, Maxim, and Norris (2010) expanded the L2 writing complexity research in another important direction by bringing in instructional effects. In addition to L2 German learner profiles, they described the so-called "idealized writing profiles" (Byrnes et al., 2010, p. 91) that were given to the learners along with writing prompts. The researchers then compared the development of general writing complexity in longitudinal cohorts to these models. Following the same approach, Byrnes and Sinicrope (2008) and Byrnes (2009) focused on more specific linguistic constructs (relative clauses and nominalization) and convincingly demonstrated that developmental profiles of tutored learners cannot be considered in isolation from specific instructional tasks and curricular progression. However, these studies measured longitudinal learner development in rather broad strokes, based on writing samples collected once per curricular level.

With regard to the development of syntactic complexity in L2 writing, most research has been conducted with global syntactic measures. The convergent, generally acknowledged results show that sentence or T-unit length values tend to increase linearly with growing L2 proficiency. However, this increase is associated at intermediate proficiency levels with a higher frequency of clauses per sentence but at advanced proficiency levels with increasing clause length (Byrnes et al., 2010; Ortega, 2003). What is still lacking is developmental research conducted with more specific syntactic measures, especially at the phrasal and clausal level. In addition to Spoelman and Verspoor (2010), reviewed above, two notable exceptions are Lu (2011) and Kormos (2011) who considered phrasal modification in cross-sectional studies that involved writers at different levels of English proficiency. Lu found that ratios of complex nominal phrases but not verb phrases discriminated between proficiency levels. Kormos, on the other hand, found that the

number of modifiers per noun phrase did not distinguish between L1 and L2 writers. Nevertheless, as Bulté and Housen (2014) remark, "the measurement of syntactic complexity at the clausal and phrasal level is only a fairly recent development in L1 and L2 complexity research, and the number of available measures is still limited" (p. 48).

Several studies in this special issue address this gap by including selected clausal and phrasal syntactic complexity measures, whereby the present study takes this research one step further by focusing on systemic aspects of syntactic modification at the sentential, clausal, and phrasal level. The exploration of the contribution of different modifier types to the developing L2 syntactic modification system is a novel contribution of this study not attempted in complexity research to date, although some studies have used counts of selected modifiers for developmental profiling of learner language. This type of research is reviewed below.

2.3. Developmental profiling and learner corpus research

As already mentioned above, general length-based measurement in CAF research needs to be supplemented with "more form-specific and development-sensitive measurement of L2 production" (Norris & Ortega, 2009, p. 567). One promising research direction is developmental profiling that originates in naturalistic SLA research and is based on the assumption that certain combinations of grammar forms characterize specific interlanguage stages (Clahsen, 1985; Pienemann, Johnston, & Brindley, 1988). Whereas these early studies represented longitudinal individual cases, the emergence of Natural Language Processing (NLP) and Learner Corpus Research (LCR) have enabled computer-aided tagging, frequency analyses, and statistical comparisons to be performed on large amounts of raw data. This allows the profiling of distinct levels of L2 proficiency based on multiple matrices of linguistic features. An overwhelming majority of LCR studies have explored L2 (mostly English) academic essays either to establish features that reliably distinguish between L2 proficiency levels, often defined as holistically rated writing quality, or to compare native and non-native writing. For coding purposes, parts-of-speech (POS) tags and their sequences have been used as "linguistic proxies" of surface syntactic structures (Borin & Prütz, 2004). For example, the POS "preposition" can be considered a proxy for prepositional phrases. We now turn to studies that employed such POS proxies for L2 writing complexity investigations (see also Leki, Cumming, & Silva, 2008, Chap. 14, for a research synthesis).

The most consistent results across different research contexts relate to prepositions (and/or prepositional phrases). Their frequency has been shown to increase with increasing L2 proficiency (for L2 English see Connor, 1990; Ferris, 1994; Grant & Ginther, 2000; Hawkins & Buttery, 2010; Tono, 2000 ; for L2 French see Granfeldt & Nugues, 2007). Furthermore, even advanced L2 writers have been shown to underuse prepositional phrases in comparison with native speakers (for L2 English see Aarts & Granger, 1998; Granger & Rayson, 1998; Kormos, 2011; Reid, 1992 ; for L2 German see Hirschmann, in press).

Another frequently investigated syntactic complexity phenomenon is subordination, often measured in frequencies of *subordinating conjunctions*. Results for this focal construct are mixed: whereas some L2 English studies (Aarts & Granger, 1998; Granger & Rayson, 1998; Grant & Ginther, 2000; Verspoor et al., 2012) found a positive correlation between the frequency of subordination and proficiency level (or nativeness), others found no relationship (Ferris, 1994; Kormos, 2011; Reid, 1992) or a negative correlation (Lu, 2011; Reid, 1992). There are several possible reasons for these inconsistent results. First, subordinate clauses in a text constitute a

subset of all clauses. Their frequencies are expected to correlate with total clause counts (a general complexity measure), which have been shown to increase in L2 writing at intermediate proficiency levels but to plateau or decline at advanced levels (Ortega, 2003). Furthermore, there is evidence that the frequency of subordination depends on the writing genre (Kormos, 2011; Lu, 2011). Next, there are indications of an L1 effect: for example, L2 English writers with the L1 Chinese background were found to use more subordinators in opening positions than L1 English writers and learners with other L1s (Reid, 1992) as well as to use fewer subordinate clauses as their L2 proficiency developed (Lu, 2011). Finally, and most importantly for this study, subordinate clauses are a heterogeneous category because they encompass verb complements (complement clauses), verb modifiers (adverbial clauses), noun modifiers (relative clauses), and even these subordinate clause types can be further divided into subtypes. It is revealing that studies that explore specific subordination types show different developmental profiles for different types (e.g., Byrnes & Sinicrope, 2008; Grant & Ginther, 2000; Hirschmann, in press; Mellow, 2006; Verspoor et al., 2012). All of the above may be reasons for the mixed results across studies based on differing proficiency stages, L1s, genres, and subordination types.

Research on other word class proxies for syntactic modifiers has been rare, although some researchers have considered adjectives and adverbs. Based on L2 English essays, Grant and Ginther (2000) found a positive linear correlation between the use of both adjectives (as noun modifiers) and adverbs (verb modifiers) and overall essay quality. Ferris's (1994) findings corroborate this result for L2 English adverbials, as do Granfeldt and Nugues (2007) for L2 French attributive adjectives. Hinkel (2002) has shown that L2 writers overuse predicative adjectives in comparison to L1 writers. Hirschmann (in press), in a comprehensive study of syntactic modification in German, found that adverbs but not adjectives were significantly underused in advanced L2 German essays when compared to L1 essays. On the other hand, Aarts and Granger (1998) found that advanced L2 English (L1 French) writers overused adverbs compared to L1 writers, whereas adjectives were used with a similar frequency. Some inconsistencies in results are certainly due to differences in research contexts and definition of variables, as mentioned above.

More recently, great strides in developmental profiling have been made with the design of NLP tools that involve syntactic parsing and allow the annotation of textual data for a wide variety of global and specific syntactic, lexical, and discourse measures (e.g., Graesser et al., 2004; Lu, 2011). These innovative tools allow researchers to arrive at rich L2 writing profiles for different proficiency levels as evidenced by studies in this special issue as well as many other studies (Friginal, Li, & Weigle, 2014; Hawkins & Buttery, 2010; Jarvis & Crossley, 2012; Kormos, 2012; Lu, 2011; Mazgutova & Kormos, this volume). Nevertheless, as aptly noted by Ortega and Sinicrope (2008), developmental profiling with specific complexity measures had not been applied to learner language below upper-intermediate proficiency levels (p. 4). Ortega and Sinicrope addressed this gap by analyzing a corpus of very short oral performance samples collected from beginning learners of Spanish and German using POS as complexity measures. The study showed how learner developmental profiles differed between the Novice Low, Novice Mid, and Novice High levels (American Council on the Teaching of Foreign Languages, 2012). In particular, ca. 80% of all content words used by Novice Low learners were nouns, whereas Novice High learners already used a more balanced POS variety in which verbs constituted up to 30-40% of content words. Methodologically, Ortega and Sinicrope demonstrated that specific L2 complexity measures, like POS, are more reliable and valid for shorter language samples (typically collected from beginner learners) than global measures, like sentence or clause length.

Ortega and Sinicrope (2008) remains a rare study that applied developmental profiling to an L2 corpus collected from novice learners of languages other than English. Whereas Ortega and Sinicrope focused on speaking, the first author of this study continued this line of research by collecting and analyzing beginner German writing corpora. This research has set the stage for the present study and is therefore reviewed below.

2.4. Previous studies with the target learner population

Three longitudinal studies were conducted with a different cohort of the same learner population as in this study: university-level *ab initio* L2 German learners whose writing was explored over four consecutive semesters. Vyatkina (2012) employed general L2 complexity measures and found a linear increase in sentence length, lexical diversity, and subordination, whereas the amount of coordination decreased, and clause length did not change over time. Vyatkina (2013a) focused on selected morpho-syntactic verb forms considered sophisticated for beginning L2 learners (separable prefix verbs, past participles, reflexive verbs, and infinitive constructions) and compared learners' writing with the pedagogical input they received from a workbook. The results showed that the general developmental trend for the learner cohort was toward a greater variety of verb forms, which also generally emulated the instructional progression. However, although this trend toward increasing variety was predominantly linear in the pedagogic corpus, it was never that "clean" in the learner corpus. High inter-individual variation in learner writing was observed throughout the time line (three semesters of study) and was especially salient when inspected for two individual learner cases. These two learner cases were explored by Vyatkina (2013b) with an expanded repertoire of specific complexity measures such as complex nominal structures, coordinate structures, and non-finite verb phrases (cf. Lu, 2011). A general trend for increasing frequency and range of specific syntactic complexity features was found, with learners diverging more from one another in the second half of the observation period. One learner made his sentences progressively longer by using more coordinate and simplex clauses, while the other used increasingly more elaborate phrasal structures per clause. Thus, Vyatkina (2013b) highlighted the importance of combining general complexity measures with fine-grained measures in developmental research and provided illustrations of individual variation at incipient levels of proficiency. The present study contributes to this line of research by an exploration of the developing syntactic modification system in its entirety including sentential, clausal, and phrasal elements at early stages of L2 German writing.

3. Study purpose and research questions

This study aims to fill some important research gaps: The vast majority of L2 writing complexity studies have focused on intermediate to advanced proficiency levels, whereas complexity at incipient levels of L2 proficiency has remained underexplored. Ortega and Byrnes (2008), commenting on the state-of-the-art of SLA research in general, point out that due to this gap, "the full longitudinal course against which advancedness needs to be understood is left rather unexplored in current SLA thinking" (pp. 5-6). Furthermore, as shown in the literature review above, most L2 writing complexity research has operated with general length-based syntactic complexity measures to paint learner developmental profiles with broad strokes or, in contrast, has focused on frequencies of (sometimes randomly) selected syntactic forms in learner

language. Furthermore, as Polio (2012) points out, the few available longitudinal L2 writing studies are either single case studies or single linguistic feature studies. This study aims to fill this gap by exploring the development of the whole system of syntactic modification in L2 writing of a group of learners while taking into account specific components at the sentential, clausal, and phrasal levels. By "specific" components we understand syntactic modifiers expressed by specific clause types and phrases headed by specific POS. Our focus on POS is justified because they have been shown to serve as appropriate developmental complexity metrics at beginning levels of L2 proficiency (Ortega & Sinicrope, 2008). Although POS are multifaceted linguistic categories at the intersection between the syntax and the lexicon, they are still typified abstractions from specific words or word combinations. Whereas Ortega and Sinicrope focused on lexical aspects of POS, this study uses them as proxies for syntactic features. This is the methodological contribution of our study given "the extreme importance of finding ways to measure 'the complex' in novice performance" (Ortega, 2012a, p. 150).

Due to its research focus, this study complements other L2 writing complexity studies collected in this special issue. Importantly, this study focuses on beginning L2 learners unlike other studies in this volume, although it has similarities with some of them in other design aspects. This study is similar to Ryshina-Pankova (this volume) in targeting German as the L2, whereas the participants in all other studies are L2 English learners. Furthermore, it employs specific complexity measures like Mazgutova and Kormos (this volume) but takes this specificity to the next level by differentiating among modifiers expressed or headed by different POS (for instance, among verb modifiers expressed by adjectives, adverbs, and prepositional phrases). Additionally, whereas Mazgutova and Kormos (this volume) select certain modifiers as complexity measures based on the assumption that these modifiers are expected to appear in academic writing at higher proficiency levels, this study is agnostic of genre conventions (in part due to the absence of relevant descriptions of beginner writing) and uses the grammatical function of modification as sole criterion for the selection of measures. For this reason, complement clauses, although they have frequently been used in complexity research, are not considered in this study because they express the syntactic function of an object and thus belong to the main sentence frame.

The specific research questions that guide the study are:

1. (How) Does the size and range of the syntactic modification system in the writing of beginning learners of German change over two years of instructed language study?
2. What are the group developmental trends in the frequency of different modifier classes and (how) are they modulated by individual differences and observation points?

4. Study design and methods

4.1. Modifier categories

Modifiers have been categorized in previous research in numerous ways: according to their meaning, syntactic function, textual function, form, complexity, position in the sentence, and other criteria (e.g., Cresswell, 1985; Demonte, 2011; Ernst, 2002; Maienborn & Schäfer, 2011). In this study, we define modifiers as optional elements that attach to the heads of noun phrases, verb phrases, or sentences. In what follows, we list modifier categories that are operationally defined as POS (adjectives, cardinal numbers, adjectives, and adverbs), phrases

headed by specific POS (prepositional phrases), and clauses which attach to sentences (adverbial clauses) or nouns (relative clauses).

4.1.1. Prenominal (attributive) adjectives (ADJA). These modifiers are prototypical nominal modifiers since they exclusively attach to nominal heads. They occur in the noun phrase between the determiner and the noun (*a **wonderful** idea*). In German, they need to be inflected according to case, number, and gender categories, and, depending on the syntactic context, they have to be inflected strong or weak. Ordinal numbers are also treated under ADJA (*my **second** sister*).

4.1.2. Cardinal numbers (CARD). Cardinal numbers occur in the same syntactic context as prenominal adjectives (*I have **two** sisters*). In German, they are, however, not inflected.

4.1.3. Predicative and adverbial adjectives (ADJP). We subsume under this category predicative adjectives (*the idea is **wonderful***) and adverbial adjectives (*the idea was **wonderfully** expressed*). In German, predicative and adverbial adjectives are uninflected and are expressed by a bare adjectival stem with no adverbial suffix added (unlike English).[2]

4.1.4. Adverbs (ADV). Adverbs function as modifiers of verbs or sentences (*She **always*** [sentence modifier] *travels **alone*** [verb modifier].). These words do not occur in either adjectival positions or in noun-specifying functions. Semantically, adverbs are a very diverse class. Different meanings of adverbs correspond with either verbal or sentence modification: modal, directional, and instrumental adverbs, for instance, are verb specific, whereas epistemic or evaluative adverbs are sentence specific.

4.1.5. Prepositional phrases (PREP). Prepositional phrases are used in various functions and are, syntactically, the most heterogeneous modifier class among the candidates we are examining. They can occur as verbal complements (*believe **in** justice*), verbal modifiers (*swim **in** the sea*), sentence modifiers (***In** contrast, …*), and adnominal modifiers (*the man **in** me*), among other functions. We will examine only verbal and sentential prepositional phrases, which have a semantic-syntactic function equal to adverbs, because adnominal ones were found to be rare in the data during preliminary screening.[3]

4.1.6. Adverbial subordinate clauses (SUB). Adverbial clauses are in most cases fronted by a subordinating conjunction with a specific meaning, which can alternatively be expressed by a prepositional phrase (***while** they were meeting – during their meeting*).

4.1.7. Relative clauses (REL). Subordinate clauses that modify nouns are in most cases relative clauses. Generally speaking, relative clauses are interchangeable with prenominal adjectives (*the embarrassing mistake – the mistake **which** was embarrassing*). In this study we do not distinguish between restrictive and non-restrictive relative clauses.

In the aggregate, these modifier types can be regarded as a basic inventory of grammatical modification. This taxonomy does not cover some more peripheral modifiers, which have been shown to be extremely rare in our data during a preliminary screening (e.g., adverbial infinitives such as *in order to…* and parenthetical insertions). However, by targeting all listed modifiers, we aim to explore the core system of modification in beginner learners' writing.

---

[2] For a contrasting view on English and German adjectives, see König and Gast (2007, pp. 65-58).

[3] This extremely low frequency of adnominal prepositional phrases in beginner-level writing is not surprising, given that they have been shown to be distinctive features of advanced writing in expository written registers (e.g., Biber, 2006).

4.2. The corpus

The study was conducted on a subset of the KanDeL corpus[4] which belongs to the Falko L2 German corpus family (Lüdeling, Walter, Kroymann, & Adolphs, 2005).[5] Whereas other Falko corpora comprise cross-sectional high intermediate to advanced L2 German data as well as comparison L1 German data, KanDeL is developmental and comprises beginning to intermediate L2 data. The subset analyzed in this study includes longitudinal data for 12 participants totaling 185 learner texts and 29,635 tokens[6] (see Appendix 1). Not all participants submitted all essays, hence the difference in sample size between data collection points. This fact does not affect the analysis, however, because the selected method (multilevel modeling) allows for missing data points (Barkaoui, 2014).

4.3. Participants

The participants were five male and seven female students who enrolled in a basic German language program over four consecutive 16-week-long semesters at a large public U.S. university. Only 12 participants were selected because they were the only ones out of their student cohort who progressed through the whole four-semester-long sequence due to attrition. All of them had American English as their L1. Eight participants were *ab initio* learners, and four were so-called "false beginners" who had had taken High School German courses but chose to have a fresh start with the language in college. All of these four students rated their own German proficiency as "beginner" in a pre-course survey. Eleven students ranged in age from 18 to 22 years (mean 19.5) at the beginning of the study, and one student was in her 30s (see Appendix 2 for learner metadata). All students were taking the course to fulfill a language requirement for a major other than German.

4.4. Instructional setting and tasks

The writing samples were rough drafts of essays written by the students in response to curricular tasks every three to five weeks during each semester rather than external experimental tasks. In this sense, this study follows the "instruction-embedded total-sampling approach" (Byrnes et al., 2010, p. 165, see also Vyatkina 2012, 2013b). Writing genres were level appropriate (Byrnes et al., 2010; Dykstra-Pruim & Redmann, 2012) and progressed from informal writing (personal narratives and descriptions) in the first year to more formal writing (essays with explanatory and interpretive elements, formal letters) in the second year. All participants progressed through the first three semesters together as a cohort in a multi-section language program, taught by different graduate student instructors, who nevertheless followed the same syllabus under the supervision of the first author. In the fourth semester, nine participants followed the regular program track, while three participants enrolled in the beginning business German language course, which entailed changes in their syllabus and, respectively, writing tasks. There was one additional variation in writing conditions. Essays at times 1-10, 13, and 17 were written under controlled conditions during 50-minute class periods

[4] https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/research/kandel

[5] https://korpling.german.hu-berlin.de/falko-suche

[6] A token is defined as a word or a punctuation mark here.

without access to supplementary materials, while essays at times 11-12 and 14-16 were written at home under uncontrolled conditions (see Appendix 3 for task metadata).

4.5. Annotations

The learner corpus used in this study, KanDeL, is annotated on multiple tiers, as are all corpora in the Falko corpus family (Lüdeling et al., 2005), including automatic POS annotation with the Tree Tagger (Schmid, 1994) which uses the POS tagset STTS (Schiller, Teufel, Stöckert, & Thielen, 1999), an established standard for German. However, although the Tree Tagger is up to 97% accurate on native German data, its accuracy drops dramatically while tagging learner data. Table 1 illustrates such a tagging error in the learner sentence from our dataset "This photo will be very impressive for the filmmakers." The *ctok* tier contains the tokenized learner sentence and shows that the learner made two mistakes in the noun *Filmmacher* by not capitalizing it and by adding the unnecessary ending *-en*. That led the POS-tagger to incorrectly assign it the attributive adjective tag, ADJA (*pos* tier).

Table 1. Example learner sentence and annotation tiers in KanDeL.

ctok	Dieses	Foto	wird	sehr	beeindruckend	zu	die	**filmmacheren**	sein
pos	PDAT	NN	VAFIN	ADV	ADJD	APPR	ART	**ADJA**	VAINF
ZH1 tok	Dieses	Foto	wird	sehr	beeindruckend	für	die	**Filmmacher**	sein
ZH1 pos	PDAT	NN	VAFIN	ADV	ADJD	APPR	ART	**NN**	VAINF

Therefore, the search for modifiers in this study was performed not on raw learner data but on the manually corrected data. More specifically, we worked with the KanDeL *ZH1* tiers (the acronym for the German equivalent of "target hypothesis 1," see Lüdeling et al., 2005). In order to create the *ZH1 tok* tier, annotators correct morpho-syntactic and orthographic learner errors to arrive at a grammatical German sentence but they keep the sentence as close to the original surface structure as possible. The ZH1 annotation of KanDeL was performed prior to and independent of this study by several annotators and in several annotation rounds following strict guidelines that ensure high inter-annotator agreement (Reznicek et al., 2012). For this study, we also performed the automatic POS annotation of the ZH1 tokens. Next, one of the authors manually checked a sample from the *ZH1 pos* tier (ca. 6% of our data) and found that the Tree Tagger assigned correct POS tags to ca. 98% of the modifiers. This level of tagger accuracy is considered very high and we accepted it as satisfactory for this study. Table 1 illustrates that the word *Filmmacher* was correctly tagged as a noun (NN) in the *ZH1 pos* tier and, therefore, was not considered among modifiers in this study.

For the purpose of the statistical analysis, the frequencies of different modifiers were automatically extracted from the corpus with the help of ANNIS,[7] a search engine for complex annotated corpora, using the following POS proxies.[8] *Prenominal (attributive) adjectives (ADJA), cardinal numbers (CARD), predicative and adverbial adjectives (ADJP),* and *adverbs (ADV)* consisted of one token each and were directly searched for by respective POS tags.

[7] http://www.sfb632.uni-potsdam.de/annis/

[8] Some original STTS tags are replaced here with more transparent English acronyms and abbreviations.

Prepositional phrases (PREP) were identified by the lexical category of the head of the phrase – the prepositions. Next, one of the authors manually filtered out prepositional phrases modifying nouns and kept only phrases modifying verbs and sentences for analysis. Finally, relative pronouns served as proxies for *relative clauses (REL)* and subordinating conjunctions (except *dass - that* and *ob - whether)*, see section *Modifier categories* above) for *adverbial subordinate clauses (SUB)*. Figure 1 shows a sample screenshot of the search tool with two search hits in one corpus sentence.

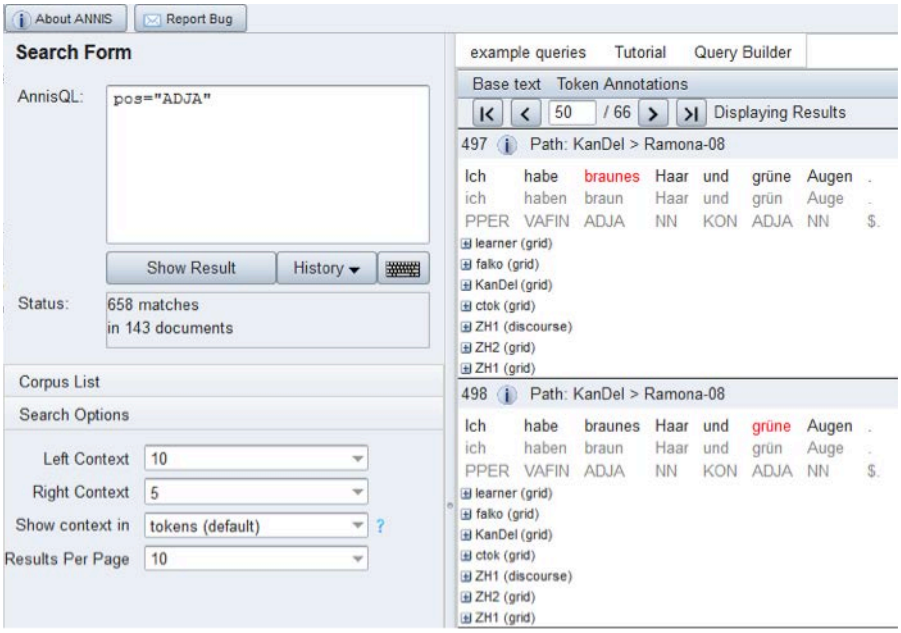

Figure 1. Screenshot of the ANNIS interface. The search window (upper left) shows a simple query for all prenominal adjectives in the corpus (ADJA). On the right side two hits are displayed (*Ich habe braunes Haar und grüne Augen – I have brown hair and green eyes*).

4.6. Statistical methods

To analyze our longitudinal data, we use a generalized linear mixed model, a specific form of mixed effects modeling (see Cunnings, 2012; Gilquin & Gries, 2009). In the following, we explain the basic features of this and the reasons for our choosing it. The dependent variable is the frequency of each of the seven target modifiers (see section *Modifier categories*). For each modifier count we consider the following independent variables: time of the text production, task, and writer. We define all independent variables except for time as random variables. These are categorical variables from which only a small sample of the target population (learners of German as a foreign language) exists in the data, which are irreproducible (any new learner would be different from any learner in the data), and whose particular characteristics are less interesting than the characteristics of the whole class for the given research question. We consider time a prototypical fixed effect because it is continuous and reproducible as well as central to our research focus, longitudinal L2 development. We chose a mixed model because it can account for random and fixed variables simultaneously.

We display the data in plots depicting normalized frequencies of different modifiers at

each point in time. The data points can be between 0 (no use of the respective modifier category) and 1 (use of the respective modifier in all possible cases, which naturally is never the case). Our data can be modeled by the binomial distribution, which in many cases can be approximated by a normal distribution. Two disadvantages which occur when using a normal distribution are, however, that the approximation breaks down when the counts are close to 0, and that the variance depends on the relative frequency of the respective modifier, a situation not foreseen by the basic methods for normally distributed data. Due to this fact, we avoid the normal approximation and directly model the binomial nature of the data by using generalized linear models with the binomial family and the logit link function.

Since the temporal development of modifiers is not always linear, but in some cases changes its direction, we do not model it with a straight line, but with (orthogonal) polynomials up to the fifth order. The possibility that individual learners follow a time development fluctuating around the average development is incorporated into the model as so-called random slopes for the time polynomials. The necessary order of the polynomials for the time dependency in both fixed and random effects is selected by choosing the model with the lowest AIC (Akaike Information Criterion; Akaike, 1974). This measure reconciles the opposing requirements for a model to describe the data as closely as possible but also with as few parameters as possible. If the AIC gives no clear decision for one single model we choose the simplest of the competing models. After selecting a model we report the significance of the fixed effect parameters by citing the z-score-based p-values as given by the summary function *lme4* (Bates et al., 2014) using the program R (R Development Core Team, 2014).

For most categories we can calculate one best model, which we present by graphs and tables of the statistical results of the calculation (Appendix 4). In these cases we are able to show statistically significant developments which are true for the whole learner population. This does not mean that individual developments must necessarily follow these trends. So we also discuss aspects of language variation within the population, supporting this discussion by appropriate visualizations using the smoothers calculated with the LOESS regression method (van Geert, 2008). The primary data input for our development modeling is measurements made at 17 time points spread over four semesters of study. However, since the learner texts vary in size from one point to another the data has to be normalized to be comparable. This normalization is not trivial and cannot be done homogenously because the seven modifier categories do not refer to the same linguistic elements. For instance, prenominal adjectives and (regular) relative clauses always refer to and only occur in the presence of nouns, whereas adverbial clauses are bound to clauses. On the other hand, adverbial adjectives, adverbs, and prepositions are not bound to nouns, verbs, or clauses, so the best reference point to normalize these POS is tokens. Therefore, when we compare frequencies of all modifiers, we normalize them uniformly vs. token count (*Results, RQ1*) but when we analyze the development of each modifier, we use different normalization bases (*Results, RQ2*): ADJA, CARD, and REL vs. noun count, SUB vs. finite verb count (=clause count), and ADJP, ADV, and PREP vs. token count.

5. Results

> **RQ 1.** (How) does the size and range of the syntactic modification system in the writing of beginning learners of German change over two years of instructed language study?

To answer the question about the size of the modification system, summary frequency

counts for all modifiers as used by each participant[9] were plotted on the timeline spanning all 17 data collection points over four subsequent academic semesters along with mean frequencies and individual trendlines (Figure 2).

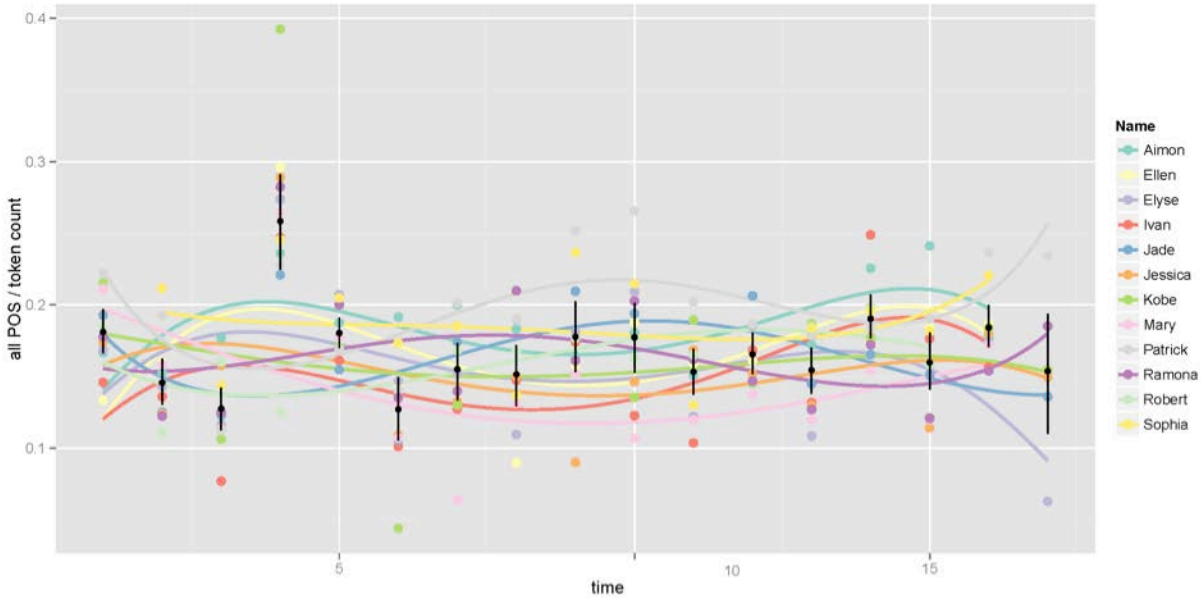

Figure 2. Individual learner trajectories for the sum of all modifiers for each participant. Smoothened lines symbolize the individual developments. Black dots represent the mean frequencies per time of measurement, black vertical lines are bootstrapped 95% confidence intervals. Frequencies are normalized by token count.

What we see from this is no clear development either for the group (black dots) or for individual learners (colored dots and trendlines), as the mean frequencies for the sum of all modifiers are neither increasing nor decreasing but rather oscillating around 12-18 modifiers per 100 written words. Then to address the question about the range of the modification system, the mean frequency counts for each modifier category normalized by text token counts and the variation of frequencies were plotted vs. time (Figure 3).

[9] All names are pseudonyms.

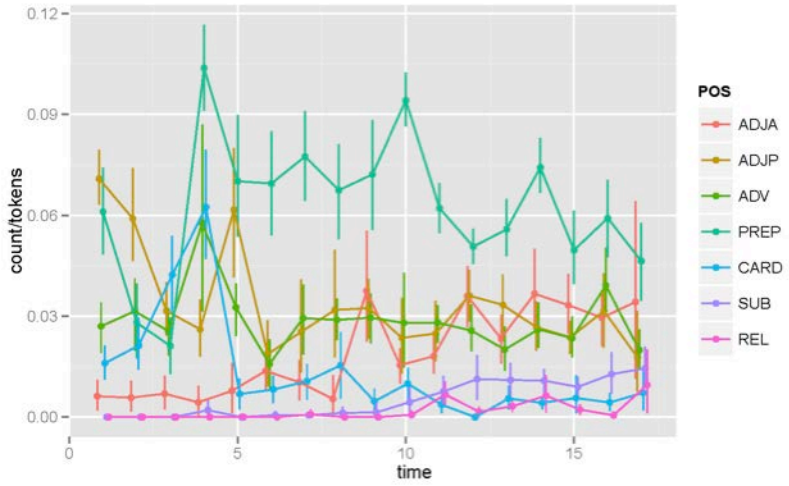

Figure 3. Development in mean frequency counts of all modifier categories normed by token count with bootstrapped 95% confidence intervals. Inter-person correlation not presented.

What these two initial plots demonstrate is that the syntactic modification system in L2 writing is complex throughout the observed period. Learners in this study did not begin writing in their second language with bare subject-verb-(object) sentence frames but modified this frame from the outset. In fact, most modifier categories are present at every point including the earlier ones except for subordinate clauses (SUB and REL), which emerge later. However, each of the modifier types follows a different development path which can be divided into three basic groups: curves showing an increasing trend (ADJA, SUB, REL), curves showing a decreasing trend (ADJP, CARD), and curves showing no clear development trend (ADV, PREP). On a very basic level this suggests that there is no uniform developmental pattern for modification as a global syntactic category.

To summarize, the answer to RQ1 is that the size and range of the modification system as a whole does not change considerably over the observed period. However, different modifiers contribute to the overall modification range to a different extent at each observation point. The development of each specific modifier is then explored below in response to RQ2.

> **RQ 2.** What are the group developmental trends in the frequency of different modifier classes and (how) are they modulated by individual differences and observation points?

To answer this question, normalized frequencies for each specific modifier were plotted along the timeline and the average developmental trend along with confidence intervals were calculated. The results are presented in graphs below for each modifier in turn. Black dots represent each learner's normalized usage frequency at a given time. The black line is the average developmental trend calculated with the mixed effects method, taking individual and text-specific variation into account. 95% confidence intervals are displayed by the grey spectra around the black line.

5.1. Attributive adjectives (ADJA)

Figure 4 shows that there is an increase in the mean ADJA frequencies normalized by

noun frequencies. Furthermore, it is obvious that there are large differences between learner texts which were written at the same time under similar conditions. This variation clearly increases with time.

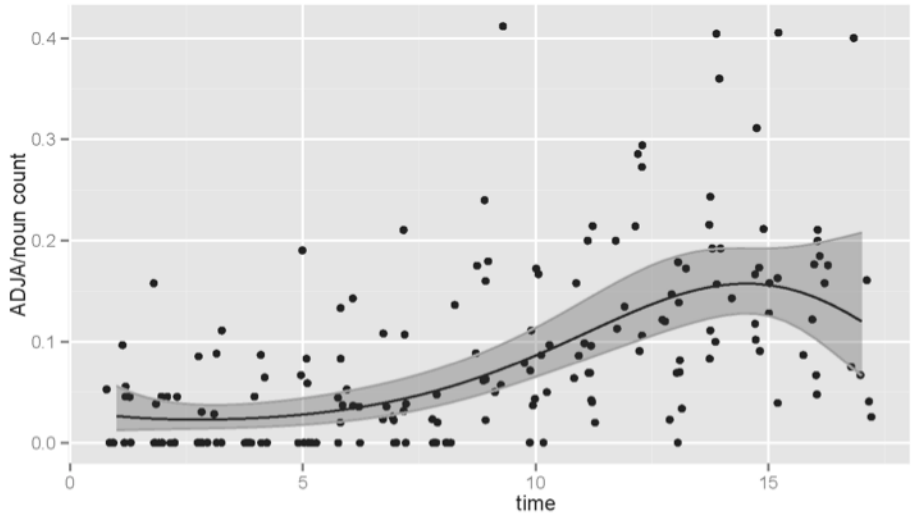

Figure 4. ADJA frequencies per observation point normalized by noun count, including a plot of a mixed effects model with 4th/4th order in time (fixed/random). See Appendix 4 (Table 2) for fixed effects values.

While the average increase of ADJA frequencies is significant (Table 2, Appendix 4), it does not tell us whether this increase holds true for each individual learner. However, by plotting the individual trajectories (and mean frequencies for each time of measurement), it is clear that individual learner development is consistent with the average development (Figure 5).

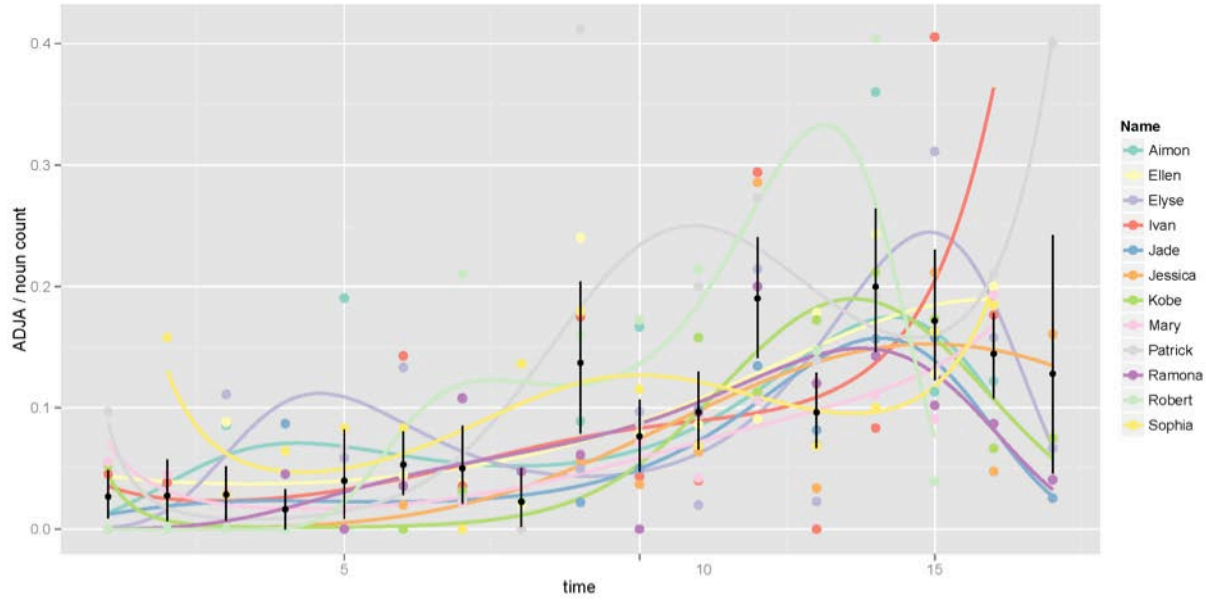

Figure 5. Individual trajectories of ADJA for all participants. Smoothened lines symbolize the individual developments. Black dots represent the mean frequencies per time of measurement, black vertical lines are bootstrapped 95% confidence intervals. Frequencies are normalized by noun frequencies.

Each learner's writing shows an increasing ADJA frequency throughout most of the observation period. However, some learners tend to produce fewer prenominal adjectives at later data collection points, while others use continuously more ADJA over time. These heterogeneous tendencies explain the expansion of the confidence interval in Figure 4 and the fact that the last section of the calculated trajectory is not a significant development (see Table 2, Appendix 4).

5.2. Cardinal numbers (CARD)

Cardinals are syntactically similar to ADJA because they also function as nominal attributes. However, in German they are not inflected (while prenominal adjectives are). Figure 6 illustrates how CARD frequencies decrease over time, the inverse of ADJA. In other words, while one type of prenominal modifier grows, the other decreases in learner writing.

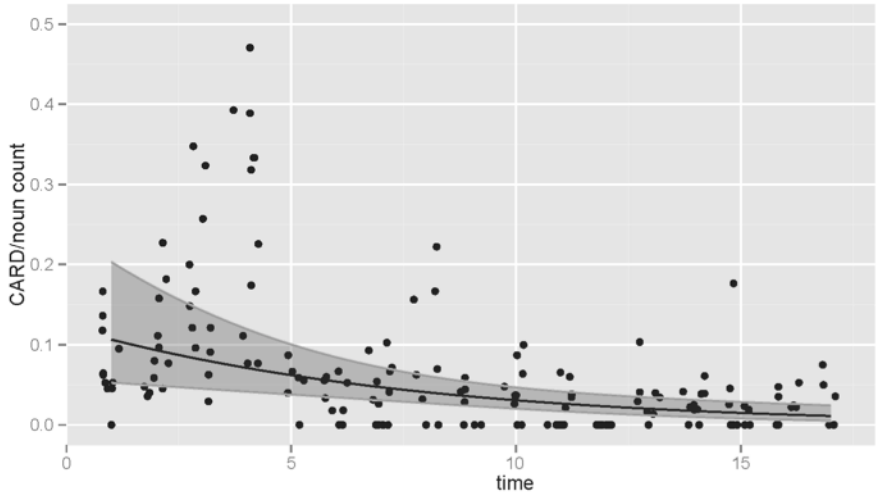

Figure 6. CARD frequencies per time point normalized by noun count, including a plot of a mixed effects model with 1st/1st order in time (fixed/random). See Appendix 4 (Table 3) for fixed effects values. See Appendix 5 (Figure 12) for individual trajectories.

The frequency variation of cardinals is lower than the variation of prenominal adjectives, apart from the third and fourth time of measurement, which can be explained by a task factor (see Section 6).

5.3. Predicative and adverbial adjectives (ADJP)

Frequencies of ADJP, like CARD, decrease over time, most drastically during the first quarter of the observation period (Figure 7). Inter-individual variation is lower for both CARD and ADJP

than for ADJA (see Appendix 5).[10]

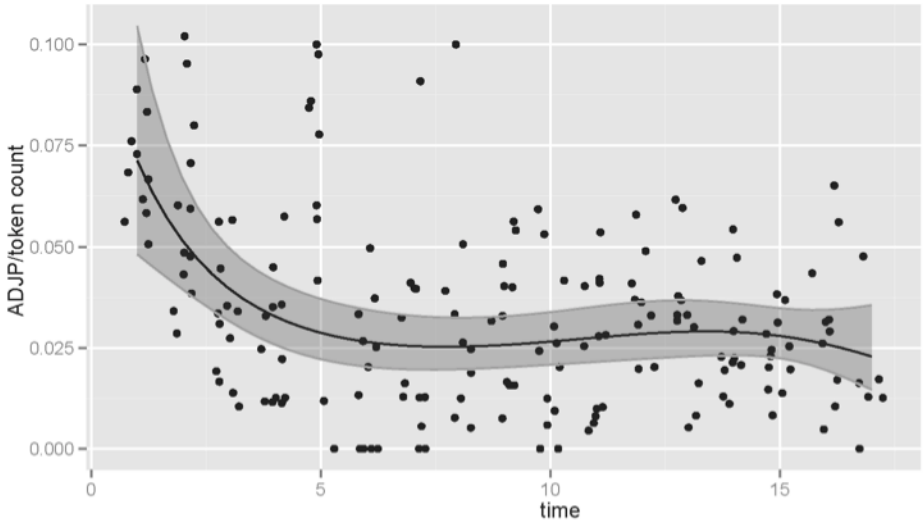

Figure 7. ADJP frequencies per time point normalized by token count, including a plot of a mixed effects model with 3rd/2nd order in time (fixed/random). See Appendix 4 (Table 4) for fixed effects values. See Appendix 5 (Figure 13) for individual trajectories.

5.4. Adverbs (ADV)

While we have seen a significant increase in prenominal adjectives and a significant decrease of cardinals and predicative/adverbial adjectives, neither of these tendencies is true for adverbs, where mixed effects modeling shows no significant correlation between the use of adverbs and time. Hence we only show the mean frequencies per time and the individual trajectories for ADV (Figure 8). The individual trajectories shown here support the fact that no significant development over time can be found for the category ADV, as opposed to, for example, ADJA, where the development of individual learners (Figure 6) goes hand in hand with the general developmental trend (Figure 5).

[10] For space considerations, only individual curves for ADJA and ADV are included in the main text, whereas individual plots for other categories are shown in Appendix 5.

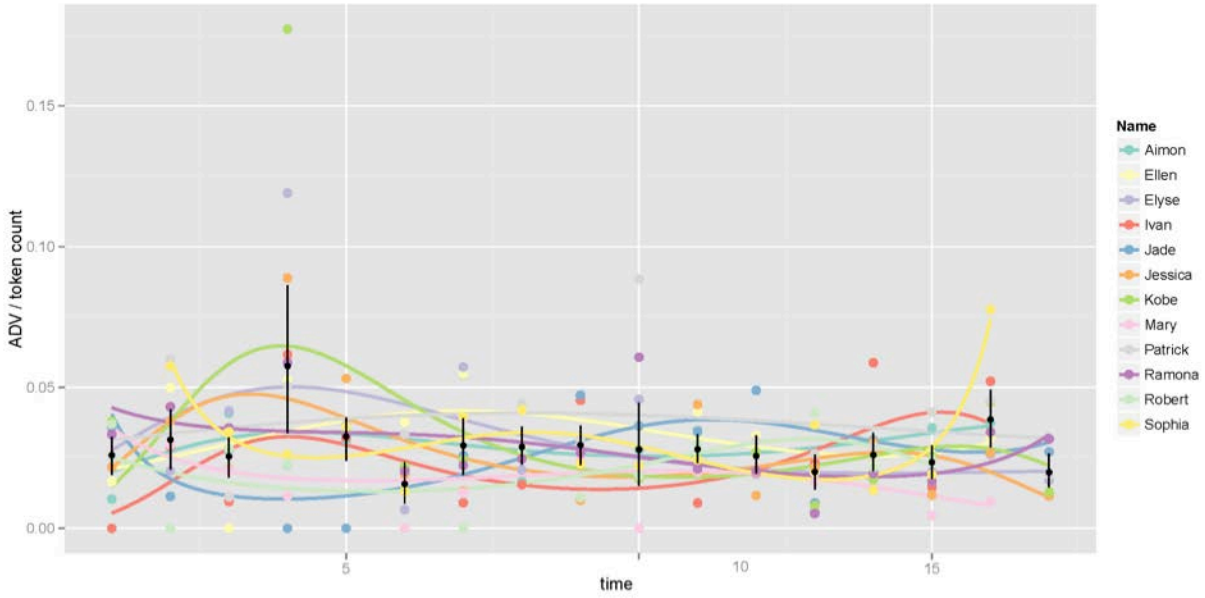

Figure 8. Individual trajectories of ADV for all participants. Smoothened lines symbolize the individual developments. Black dots represent the mean frequencies per time of measurement, black vertical lines are bootstrapped 95% confidence intervals. Frequencies are normalized by token frequencies.

5.5. Prepositional phrases (PREP)

Prepositions (like adverbs) show no clear development in terms of the fact that there is no uniform development over time (Figure 9). There is, however, a significant upward development in the first half and a significant downward development in the second half of the time plot (Table 5, Appendix 4). A unique characteristic of prepositions is that they are used in almost all learner texts (only at T4 are there two learners who do not use any prepositions at all).

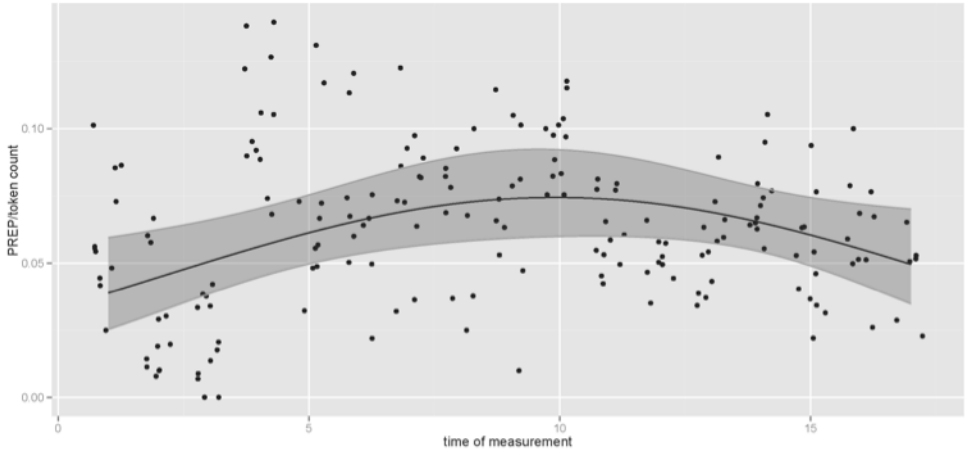

Figure 9. PREP frequencies per time point normalized by token count, including a plot of a

mixed effects model with 2nd/1st order in time (fixed/random). See Appendix 4 (Table 5) for fixed effects values. See Appendix 5 (Figure 14) for individual trajectories.

5.6. Adverbial subordinate clauses (SUB) and relative clauses (REL).

The two last categories are clausal modifiers. Figures 10 and 11 show similar developmental trajectories.

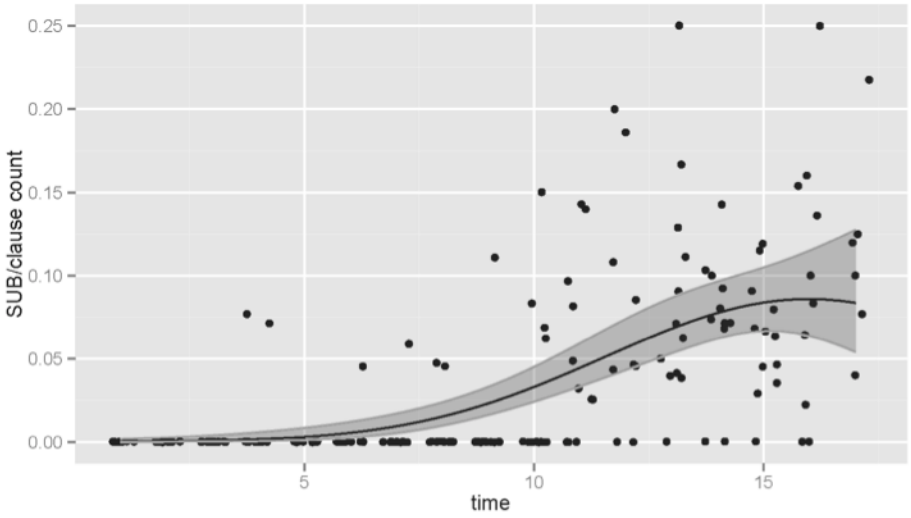

Figure 10. SUB frequencies per time point normalized by clause count, including a plot of a mixed effects model with 2nd/0 order in time (fixed/random). See Appendix 4 (Table 6) for fixed effects values. See Appendix 5 (Figure 15) for individual trajectories.

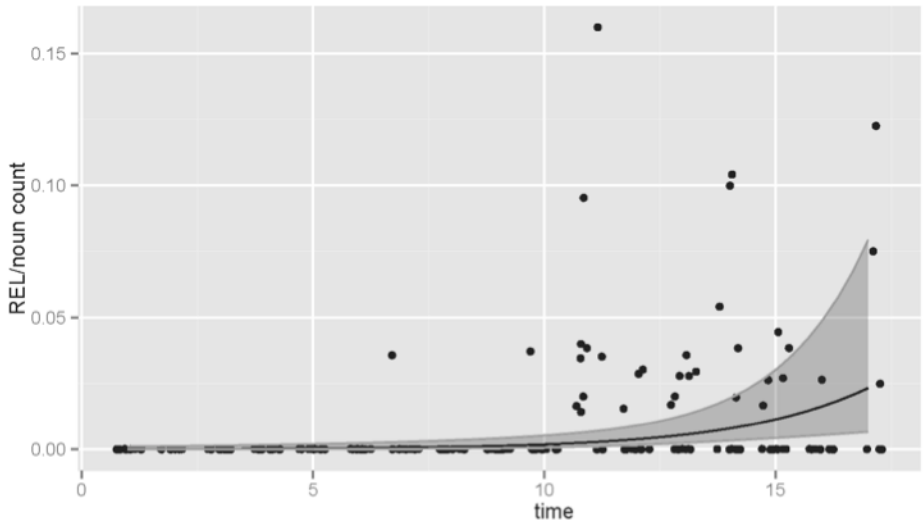

Figure 11. REL frequencies per time point normalized by clause count, including a plot of a mixed effects model with 1st/0 order in time (fixed/random). See Appendix 4 (Table 7) for fixed effects values. See Appendix 5 (Figure 16) for individual trajectories.

Frequencies for both SUB and REL – like for ADJA – increase, with the SUB curve being steeper and rising earlier. Neither modifier category appears at earlier stages, but both tend to increase after the first occurrence. The variation for each non-zero point in time is relatively high (see also Appendix 5), and for each time point there are learners who do not use these modifiers at all.

To summarize, the answer to RQ2 is that different modifiers have different developmental trends in the writing of our participants. ADJA, SUB, and REL show increasing trends, CARD and ADJP decreasing trends, ADV no discernible trend, and PREP an increasing and then decreasing trend. This finding fine-tunes the results in response to RQ1 above, illustrating the specific contribution of each modifier to the complex weave of the whole modification system at each observation point. As frequencies of some modifiers increase and those of the other ones decrease over time, the overall size and range of the modification system remains almost constant (cf. Figures 2 and 3). Furthermore, our analysis shows that individual differences modulate group trends to a larger or smaller extent depending on the modifier type and observation point, although it is overall considerable as illustrated by the wide dispersion of dots in the plots above.

6. Discussion

Our analysis shows that syntactic modification as a global construct is a constant characteristic of L2 writing from its inception, whose size and range remain relatively stable over four semesters of instructed language study. At the same time, this system exhibits considerable inter-individual and intra-individual variation in the contribution of its specific components. We agree with Larsen-Freeman (2012, p. 80) that this "interplay of stability and variability offers potentially useful information" for explaining the L2 development from the DUB perspective. We explain the relative stability of the overall system by the fact that the learners in this study were adult mature individuals who needed to express complex thoughts even with limited L2 resources, which led them to use syntactic modifiers beginning with their very first writing assignment. The variation, on the other hand, was manifested in many forms and merits a detailed discussion. In what follows, we discuss the found variation patterns against the backdrop of previous L2 writing complexity research and explain them from the DUB perspective in terms of different complexity aspects, task effects, and inter-individual differences.

6.1. Upward and downward trends

Five variables in this study (CARD, ADJP, ADJA, SUB, REL) showed clear developmental trends. The learners primarily relied on predicative adjectives and prenominal cardinals at earlier stages. As they progressed, they used fewer of these modifiers and more attributive adjectives. This finding confirms results from previous research that associated high frequencies of predicative adjectives with lower L2 proficiency (Hinkel, 2002) and those of attributive adjectives with higher L2 proficiency (Granfeldt & Nugues, 2007; Grant & Ginther, 2000). Next, we found an increase in clausal modifiers in the second half of our observation period, which confirms findings from previous research showing that frequencies of adverbial clauses increase up to intermediate levels and those of relative clauses up to advanced levels (Byrnes & Sinicrope, 2008; Verspoor et al., 2012). Moreover, our study extends these results by showing that ADJA, ADJP, CARD, SUB, and REL frequencies can serve as benchmarks that

capture development at lower levels of L2 writing proficiency and between relatively dense data collection waves.

In terms of the reasons for these developmental trends, we need to tap into L2 complexity dimensions beyond the focus of our study on linguistic complexity (operationally defined here as frequency of syntactic modifiers), namely cognitive difficulty. Whereas CARD and ADJP are uninflected POS, the use of ADJA requires learners to utilize the German inflection system, which has been shown to cause difficulties to L2 German learners even at advanced proficiency levels (Baten, 2011; Diehl, 1994). Subordination measures (SUB and REL in our study) have also been shown to be "hybrid measures" (Bulté and Housen, 2012, p. 36) that capture distinct yet interrelated complexity aspects: diversity, depth, compositionality, and cognitive difficulty. Our results show that features associated not only with structural diversity (new modifiers) but also cognitive difficulty emerge and begin growing in learner writing at later time points, whereas less cognitively difficult features decline. Therefore, our results are in line with the Dynamic Usage-Based approach (DUB) and its definition of development "as the growth or increase of more developmentally advanced […] variables and the decline or decrease of less developmentally advanced variables" (Verspoor et al., 2012, p. 242; see also van Geert, 2008).

Furthermore, it is noteworthy that all increasing and decreasing trends in our data are curvilinear, with smoother curves for some variables and abrupt curvature for others. For instance, attributive adjectives show a pronounced smooth growth that becomes increasingly steeper, whereas cardinals drop drastically by T5 and remain only slightly above zero until the end of the observation period. Next, the increase in adverbial clauses is followed by relative clauses, although at a much more moderate rate. This finding confirms conclusions reached in other DUB complexity studies that development is never linear (e.g., Verspoor, 2012; Zhang & Lu, 2012).

6.2. No growth and mixed pattern

As opposed to the clear trends described above, this study found no change in the frequency of adverbs. One possible explanation is that at the early stages of language acquisition (our focus here), learners mostly rely on a few high frequency adverbs. Although a systematic inquiry into lexical complexity (including counts of lexical types and tokens) was beyond the scope of this study, we performed a qualitative screening of adverbs in our corpus. The most frequent adverbial lexemes turned out to be intensifiers (*sehr* – very, by far the most frequent adverb in the corpus) and prototypical local (*hier* – here), temporal (*jetzt* – now*; dann* – then), or modal (*gern*(*e*) – used in *to like to do smth.* constructions) adverbs, whereas sentence adverbs were hardly used at all. This finding may be an indication of the fact that beginning learners use some modifiers as memorized chunks such as *sehr gut* (very good) (cf. Myles, 2012), whereas considerable growth in adverb frequency and variety may begin at later developmental stages. Moreover, the lack of growth in adverb frequencies in our data is not surprising given that even advanced learners of German were found to significantly underuse adverbs in comparison with native speakers, although there was no difference in other modifiers (Hirschmann, in press).

For prepositional modifiers, the picture was mixed with no clear developmental trend. Our screening of a sample of prepositional modifiers showed that they are, like adverbs, mainly locational (*Die Musik ist meine Lieblingssache* ***in Lawrence*** – music is my favorite thing **in Lawrence**) and temporal (*und es ist heiß und schwül* ***im Sommer*** – and it is hot and muggy **in the summer**). As previous research consistently associated higher frequencies of prepositions

with higher proficiency, we expect that at later acquisition stages the use of prepositional sentence and nominal modifiers could lead to a further increase of prepositional modifiers in general.

6.3. Variation due to task effects

As in all longitudinal studies with dense data collection intervals, writing prompts could not have been kept constant in our study. Moreover, because the study was administered in a tutored setting, instruction was also expected to influence learner writing development. Although the design of this study did not single out task effects, the metadata for all topics, prompts, and instructional foci have been collected and can be used to explain oscillations in modifier frequencies at certain observation points. Topic effects can be seen, for instance, at T4 and T10, when the topics "your daily routine" and "compare two trips" triggered high frequencies of prepositional phrases (*Ich esse Abendessen* ***um*** *sieben Uhr* – I eat dinner **at** seven o'clock; *sie fuhren* ***mit*** *dem Zug* – they traveled **by** train). The drastic increase of cardinals at T3 and T4 can be explained in the same vein: The students were asked to write about their families and friends (T3) and about their everyday life (T4). The most reoccurring statements in the respective texts are specifications about the number of family members and friends, pets, rooms, and time of day, all of which contain cardinals. Importantly, however, these oscillations due to topic effects did not change the developmental trends (e.g., there was no upward fluctuation in the CARD trendline at T3 - T4, see Figure 6).

On the other hand, instructional foci did have an effect on the models. Such effects can be observed at T8, when subordinate clauses begin to grow after focused instruction, and at T9, when the growth curve for attributive adjectives becomes steeper after this grammar feature was the instructional target. These patterns of variation have important methodological ramifications. Although the conflation of the topic variable is hard to completely avoid in longitudinal research, the random topic effects did not significantly influence the developmental trends in this study. Also, instructional foci did not cause the onset of development (e.g., for prenominal adjectives) or change its direction but rather modulated growth curves. This finding provides evidence that, whereas other designs (e.g., counterbalancing topic and time variables among subjects) may be better for singling out task effects, mixed effect models can still capture task-independent developmental trends in studies situated in ecologically valid real-life instructional settings, where writing topics change over time in conjunction with changing instructional foci (Byrnes et al., 2010).

6.4. Inter-individual variation

Finally, a large inter-individual variation was found in the data, as illustrated by a large dispersion of data points and individual trendlines in the data plots. The inter-individual variation is especially pronounced for modifiers with growing trendlines (attributive adjectives, adverbial and relative clauses). As a general tendency, the variation becomes larger toward the end of the observation period, which is illustrated by widening confidence intervals. For example, Figure 5 demonstrates that five learners use increasingly more attributive adjectives at the end of the observation period, whereas trendlines for seven other learners start declining (which causes the group trendline to also slightly decline as shown in Figure 4). Furthermore, some categories are used especially frequently by individual learners: ADJA are used with high frequency by Patrick

(Figure 5), ADJP by Aimon and Sophia, ADV by Sophia and Ivan, SUB by Ramona, Robert, and Jessica, and PREP by Mary (Appendix 5, Figures 12-16). On the other hand, there are also clear dislikes such as ADJP for Elyse and SUB for Ellen and Elyse (Appendix 5, Figures 13 and 15). Additionally, at all time points there are modifiers that are not used by some or by any learners (only prepositions are used in all but two texts). The finding that each learner uses some selected modifier types much more than others may point to trade-off effects within learners' developing interlanguage systems (Adams, Newton, & Nik, this volume) as well as to the fact that learners "exercise agency in the constructions they choose to use" (Larsen-Freeman, 2012, p. 77). On the one hand, this finding confirms the importance of the study of inter-individual variation in L2 writing complexity development (see also Gunnarsson, 2012; Larsen-Freeman, 2006; Vyatkina, 2013a, 2013b; Zhang & Lu, 2013). On the other hand, our study shows that even though inter-individual variation is considerable, there are still clear general developmental trends for several modifiers that hold for all learners. In particular, we found that frequencies of subordinate and relative clauses clearly increase as a group trend for our beginning learners. These results show the potential of dynamic methods to systematically study both uniform group trends and inter-individual variation that contribute to the overall picture of long-term L2 writing complexity development.

7. Conclusion and Implications

This study makes a contribution to the study of L2 writing development and supplements other studies in this Special Issue by presenting a dynamic profile of syntactic modification in the writing of *ab initio* learners of German. First, this study has proved that modification as a grammatically optional category is present in learner writing from the onset of tutored language study, and its size (the number of modifier tokens) and range (the number of modifier types) remains relatively stable over four semesters of instructed language study. Furthermore, this study confirmed with longitudinal data some developmental assumptions that have been made in previous cross-sectional research. In particular, it showed that not only ubiquitous global measures of syntactic complexity but also more specific measures, namely syntactic modifiers, can serve as developmental indices. At the same time, we presented a complex picture of development as a non-linear phenomenon that combines decline in cognitively easier (e.g., uninflected) structures and increase in more difficult (e.g., inflected and clausal) structures. Furthermore, our results illustrate the importance of a differentiated approach to developmental data: only by presenting average trends and individual variation together do we get a truly multifaceted picture of longitudinal development in a learner population. This analysis was afforded by the mixed effects modeling method and the theoretical DUB approach.

This study has important implications for L2 writing instruction and research. Our results show that learners use syntactic modification in their writing from the very beginning of their language study. We argue that this fact merits attention of L2 teachers. Although we did not specifically focus on instructional effects, our results also indicated that focused grammar instruction supported and accelerated development in more advanced modification categories. These findings can encourage L2 writing instructors to explicitly address writing complexity in their teaching, attract learners' attention to it, and give feedback on complexity aspects of learner writing (see also Connor-Linton & Polio, 2014). We also hope that our results will prompt L2 writing researchers to explore specific instructional effects with mixed effect modeling approaches, which can help account for a variety of learner-related and task-related factors

(Byrnes et al., 2010; Norris & Ortega, 2012).

Finally, we would like to acknowledge the limitations of this study, which may indicate promising avenues for future research. First, the number of participants was low, so the results cannot be generalized beyond our target learner population, although significant trends that we found may be indicative of beginning stages of L2 German (L1 English) acquisition in instructed settings. This assumption can be tested in replication studies with other learner groups. Next, we only targeted the development of L2 writing complexity as syntactic modification. This focus needs to be expanded in future studies by considering lexical complexity aspects as well as associated developmental dimensions such as accuracy. Furthermore, as already mentioned above, instructional effects and task effects should feature more prominently in future studies of longitudinal language development in tutored acquisition settings. Also, explorations of dynamic relationships between different modifiers in learner writing may supplement this study's results. Finally, extending the observation period and increasing the density of observation points would be invaluable in order to arrive at a fuller L2 development picture, although conducting such studies is admittedly a very significant undertaking.

Acknowledgements

This study was supported in part by the Language Learning Small Grants Research Program, the University of Kansas General Research Fund allocation No. 2301446, and the Fulbright U.S. scholar grant, U.S.-Germany Fulbright Commission. Many thanks go to Emily Hackmann, Michael Gruenbaum, and Melanie Piltingsrud for their help with data annotation and Marc Reznicek for his collaboration on an earlier draft of this article.

Appendix A. Corpus composition

Semester	Time (T)	Number of texts	Tokens (total)	Tokens (mean)	SD
first	1	11	1033	94	13.4
	2	12	1251	104	15.1
	3	12	1452	121	27.2
	4	12	1114	93	20.8
	5	10	883	88	5.1
second	6	11	1554	141	26.0
	7	12	1638	137	33.8
	8	10	1482	148	56.7
	9	11	1565	142	21.7
	10	12	1669	139	30.1
third	11	12	2928	244	65.2
	12	9	2170	241	70.0
	13	12	1990	166	36.8
fourth	14	11	2874	261	82.1
	15	12	2886	241	38.3
	16	10	2111	211	61.9
	17	6	1035	173	13.6
Total		185	29635	2744	69.6

Appendix B. Participant metadata (all names are pseudonyms)

Name	Gender	Age, onset	High School German	4th semester track
Aimon	m	19	yes	business
Ellen	f	18	no	business
Sophia	f	>25	yes	business
Elyse	f	19	no	regular
Ivan	m	18	no	regular
Jade	f	19	yes	regular
Jessica	f	19	no	regular
Kobe	m	22	no	regular
Mary	f	19	no	regular
Patrick	m	20	no	regular
Ramona	f	21	yes	regular
Robert	m	20	no	regular

Appendix C. Task metadata

Semester	Time point	Track	Topic	Genre	Condition
First	1	regular	Who are you?	personal narrative / description	50 minutes
	2	regular	Your apartment, your friends	personal narrative / description	50 minutes
	3	regular	Your family	personal narrative / description	50 minutes
	4	regular	Your daily routine	personal narrative / description	50 minutes
	5	regular	Description: Your favorite clothes OR Your lucky charm OR A shopping day	personal narrative / description	50 minutes
Second	6	regular	Party plan	personal narrative / description	50 minutes
	7	regular	Your last weekend	personal narrative / description	50 minutes
	8	regular	Describe yourself or another person	personal narrative / description	50 minutes
	9	regular	Your town	personal narrative / description	50 minutes
	10	regular	Comparison of two trips	personal narrative / description	50 minutes
Third	11	regular	Tips for an extraterrestrial	personal letter	untimed
	12	regular	Journalist for a day	expository essay	untimed
	13	regular	Book review	summary and review	50 minutes
Fourth	14	regular	Interpret a short story	interpretation	untimed
	14	business	Letter of self-introduction	formal letter	untimed
	15	regular	Interpret a short story	interpretation	untimed

15	business	Letter of complaint	formal letter	untimed
16	regular	Interpret a short story	interpretation	untimed
16	business	Letter to a friend	informal letter	untimed
17	regular	Film review	summary and review	50 minutes

Appendix D. Data tables for mixed effects models (fixed effects)

Table 2
Fixed effects values for category ADJA (4th/4th order).

	Estimate	Std. Error	z value	Pr(|z|)
(Intercept)	-2.76	0.13	-21.34	<0,001
poly(time,4)1	11.13	1.58	7.05	<0,001
poly(time,4)2	-1.03	1.63	-0.63	0.528
poly(time,4)3	-2.73	1.48	-1.84	0.066
poly(time,4)4	-0.39	1.70	-0.23	0.819

Table 3
Fixed effects values for category CARD (1st/1st order).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.26	0.22	-14.93	<0,001
poly(time, 1)	-9.56	2.72	-3.52	<0,001

Table 4
Fixed effects values for category ADJP (3rd/2nd order).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.44	0.08	-44.12	<0,001
poly(time,3)1	-2.71	0.92	-2.94	0.003
poly(time,3)2	2.32	0.99	2.35	0.019
poly(time,3)3	-1.86	0.87	-2.14	0.033

Table 5
Fixed effects values for category PREP (2nd/1st order).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.73	0.08	-33.68	<0,001
poly(time,2)1	1.30	1.04	1.25	0.212
poly(time,2)2	-2.39	1.04	-2.31	0.021

Table 6
Fixed effects values for category SUB (2nd/0 order).

Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-4.52	0.33	-13.83	<0,001
poly(time,2)1	26.71	5.02	5.32	<0,001
poly(time,2)2	-8.02	2.82	-2.84	0.004

Table 7
Fixed effects values for category REL (1st/0 order).

Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-6.80	0.64	-10.56	<0,001
poly(time,1)	23.75	7.48	3.17	0.002

Appendix E. Individual trajectories not shown in the main text

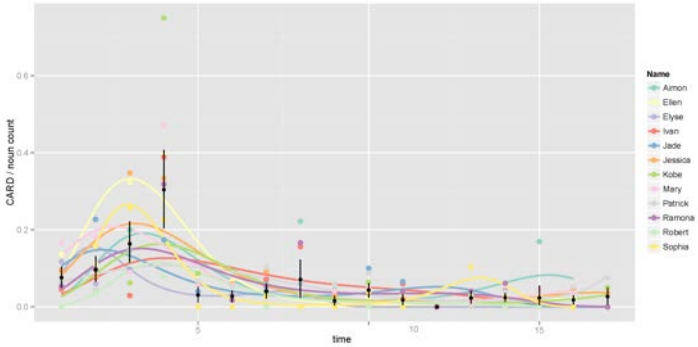

Figure 12. Individual trajectories of CARD for all persons in the corpus. Smoothened lines symbolize the individual developments. Black dots represent the mean frequencies per time of measurement, black vertical lines are bootstrapped 95% confidence intervals. Frequencies are normalized by noun frequencies.

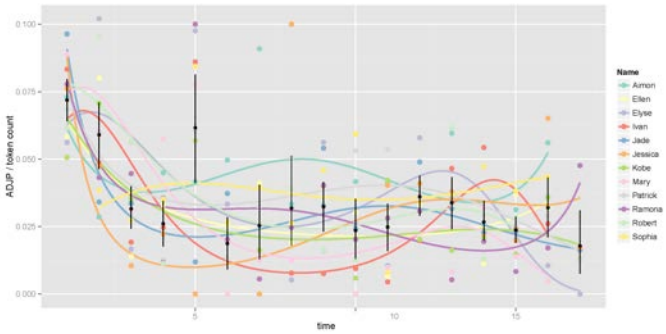

Figure 13. Individual trajectories of ADJP for all persons in the corpus. Smoothened lines symbolize the individual developments. Black dots represent the mean frequencies per time of measurement, black vertical lines are bootstrapped 95% confidence intervals. Frequencies are normalized by token frequencies.

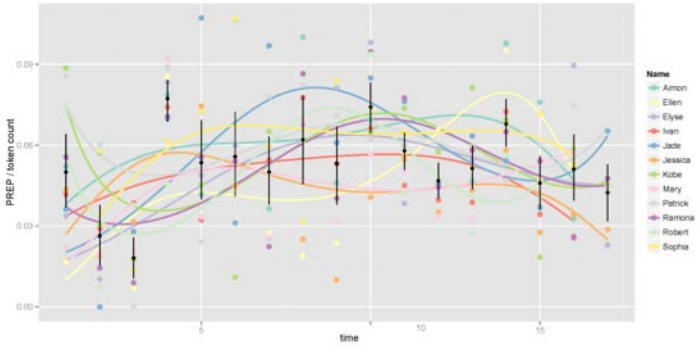

Figure 14. Individual trajectories of PREP for all persons in the corpus. Smoothened lines symbolize the individual developments. Black dots represent the mean frequencies per time of measurement, black vertical lines are bootstrapped 95% confidence intervals. Frequencies are normalized by token frequencies.

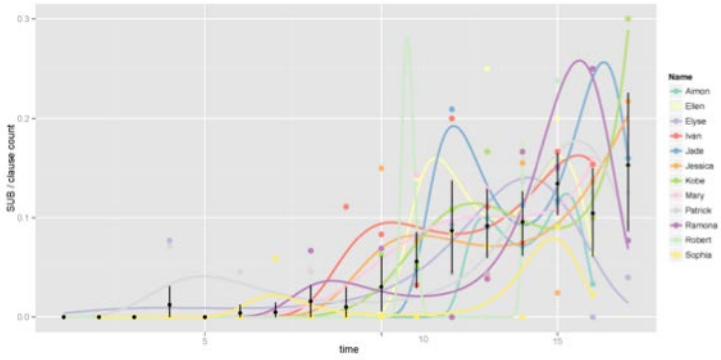

Figure 15. Individual trajectories of SUB for all persons in the corpus. Smoothened lines symbolize the individual developments. Black dots represent the mean frequencies per time of measurement, black vertical lines are bootstrapped 95% confidence intervals. Frequencies are normalized by clause frequencies.

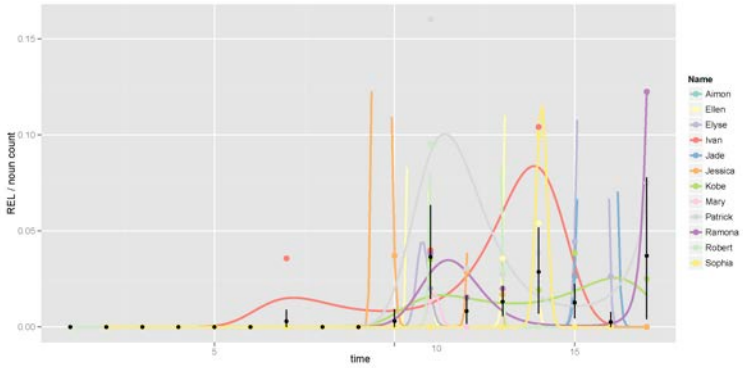

Figure 16. Individual trajectories of REL for all persons in the corpus. Smoothened lines symbolize the individual developments. Black dots represent the mean frequencies per time of measurement, black vertical lines are bootstrapped 95% confidence intervals. Frequencies are normalized by noun frequencies.

References

Aarts, J., & Granger, S. (1998). Tag sequences in learner corpora: A key to interlanguage grammar and discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 132–141). New York: Longman.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

American Council on the Teaching of Foreign Languages (2012). *ACTFL proficiency guidelines 2012*. Alexandria, VA: ACTFL.

Barkaoui, K. (2014). Quantitative approaches for analyzing longitudinal data in second language research. *Annual Review of Applied Linguistics, 34*, 65-101.

Baten, K. (2011). Processability theory and German case acquisition. *Language Learning, 61*, 455-505.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. (R package version 1.1-7) [Computer software]. Vienna: R Foundation

for Statistical Computing. Retrieved from http://cran.r-project.org/web/packages/lme4/index.html
Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
Borin, L., & Prütz, K. (2004). New wine in old skins? A corpus investigation of L1 syntactic transfer in learner language. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 67–87). Amsterdam: John Benjamins.
Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam: John Benjamins.
Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42-65.
Byrnes, H. (2009). Emergent L2 German writing ability in a curricular context: A longitudinal study of grammatical metaphor. *Linguistics and Education, 20*, 50–66.
Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment [Monograph]. *Modern Language Journal*, *94*(s1).
Byrnes, H., & Sinicrope, C. (2008). Advancedness and the development of relativization in L2 German: A curriculum-based longitudinal study. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 109–138). New York: Routledge/Taylor & Francis.
Clahsen, H. (1985). Profiling second language development: A procedure for assessing L2 proficiency. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 283– 331). Clevedon: Multilingual Matters.
Connor, U. (1990). Linguistic/rhetorical measures for international persuasive student writing. *Research in the Teaching of English, 24*, 67-87.
Connor-Linton, J., & Polio, C. (2014). Comparing perspectives on L2 writing: Multiple analyses of a common corpus. *Journal of Second Language Writing, 26*, 1-9.
Cresswell, M. J. (1985). *Adverbial Modification. Interval semantics and its rivals*. Dordrecht: D. Reidel.
Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research, 28*, 369-382.
Demonte, V. (2011). Adjectives. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning. Volume 2* (pp. 1314-1340). Berlin: Mouton de Gruyter.
Diehl, E. (1994). Probieren geht über Studieren: Individuelle Varianten im Fremdsprachenerwerb [Trial is worth more than study: Individual variation in second language acquisition]. *Deutsch als Fremdsprache [German as a Foreign Language], 31*, 138-147.
Dykstra-Pruim, P., & Redmann, J. (2012). *Schreiben lernen [Learning to write]: A writing guide for learners of German*. New Haven, Conn.: Yale University Press.
Ernst, T. (2002). *The syntax of adjuncts*. Cambridge: Cambridge University Press.
Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly, 28*, 414–420.
Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299–323.

Friginal, E., Li, M., & Weigle, S. C. (2014). Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing, 23*, 1–16.

Gilquin, G., & Gries, S. T. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory, 5*, 1–26.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*, 193–202.

Granfeldt, J., & Nugues, P. (2007). Evaluating stages of development in second language French: A machine-learning approach. In J. Nivre, H.-J. Kaalep, K. Muischnek, & M. Koit (Eds.), *NODALIDA 2007 Conference Proceedings* (pp. 73-80). University of Tartu, Estonia.

Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), *Learner English on computer* (pp. 119–131). New York: Longman.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing, 9*, 123–145.

Gunnarsson, (2012). The development of complexity, accuracy and fluency in the written production of L2 French. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 247–276). Amsterdam: John Benjamins.

Hawkins, J. A., & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. *English Profile Journal, 1*, 1-23.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah, NJ: Lawrence Erlbaum.

Hirschmann, H. (2015). *Modifikatoren im Deutschen [Modifiers in German]*. Tübingen, Germany: Stauffenburg.

Hirschmann, H., Lüdeling, A., Rehbein, I., Reznicek, M., & Zeldes, A. (2013). Underuse of syntactic categories in Falko: A case study on modification. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty years of learner corpus research: Looking back, moving ahead. Corpora and Language in Use – Proceedings 1* (pp. 223-234). Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*, 461–473.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy, and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1–20). Amsterdam: John Benjamins.

Jarvis, S., & Crossley, S. A. (2012). (Eds.). *Approaching language transfer through text classification: Explorations in the detection-based approach*. Bristol: Multilingual Matters.

König, E., & Gast, V. (2007). *Understanding English-German contrasts*. Berlin: Erich Schmidt Verlag.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing, 20*, 148-161.

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing, 21*, 390-403.

Langacker, R. W. (2008). Cognitive grammar as a basis for language instruction. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 66–88). New York: Routledge.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*, 590–619.

Larsen-Freeman, D. (2012). Complexity theory. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 73-87). London: Routledge.

Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. London: Routledge.

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*, 36–62.

Lüdeling, A., Walter, M., Kroymann, E., & Adolphs, P. (2005). Multi-level error annotation in learner corpora. *Proceedings of Corpus Linguistics 2005*, Birmingham, UK.

Maienborn, C., & Schäfer, M. (2011). Adverbs and adverbials. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning. Volume 2* (pp. 1390-1420). Berlin: Mouton de Gruyter.

Manchón, R.M. (2012). Multiple perspectives in the study of L2 writing development: An introduction. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 1-16). Berlin: Mouton de Gruyter.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of quality writing. *Written Communication, 27*, 57–86.

Mellow, J. D. (2006). The emergence of second language syntax: A case study of the acquisition of relative clauses. *Applied Linguistics, 27*, 645-670.

Myles, F. (2012). Complexity, accuracy and fluency: The role played by formulaic sequences in early interlanguage development. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 71–94). Amsterdam: John Benjamins.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*, 555– 578.

Norris, J. M., & Ortega, L. (2012). Assessing learner knowledge. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 573-589). London: Routledge.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*, 492–518.

Ortega, L. (2012a). Interlanguage complexity: A construct in search of theoretical renewal. In B. Szmrecsanyi & B. Kortmann (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127– 155). Berlin: Mouton de Gruyter.

Ortega, L. (2012b). Epilogue: Exploring L2 writing – SLA interfaces. *Journal of Second Language Writing, 21*, 404-415.

Ortega, L., & Byrnes, H. (2008). The longitudinal study of advanced L2 capacities: An introduction. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 3–20). New York: Routledge/Taylor & Francis.

Ortega, L., & Sinicrope, C. (2008). *Novice proficiency in a foreign language: A study of task-based performance profiling on the STAMP test*. [Technical report]. University of Oregon, Center for Applied Second Language Studies.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*, 117-134.

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition, 10,* 217–243.

Polio, C. (2012). The acquisition of second language writing. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 319-334). London: Routledge.

R Core Team. (2014). R: A language and environment for statistical computing [Computer software]. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Reid, J. (1992). A computer text analysis of four cohesion devices in English discourse by native and nonnative writers. *Journal of Second Language Writing, 1*, 79–107.

Reznicek, M., Lüdeling, A., Krummes, C., Schwantuschke, F., Walter, M., Schmidt, K., Hirschmann, H., & Andreas, T. (2012). *Das Falko-Handbuch. Korpusaufbau und Annotationen [The Falko manual: Corpus construction and annotation]. Version 2.01*. Retrieved from https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko/Falko-Handbuch_Korpusaufbau%20und%20Annotationen_v2.01

Schiller, A., Teufel, S., Stöckert, C., & Thielen, C. (1999). *Guidelines für das Tagging deutscher Textcorpora mit STTS [Guidelines for tagging German corpora of written language with STTS]*. Technical Report. Stuttgart, Germany: Institut für maschinelle Sprachverarbeitung [Institute for Machine Language Processing].

Schmid, H. (1994, September). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the international conference on new methods in language processing* (pp. 44–49). Manchester, UK.

Skehan, P. (1989). *Individual differences in second language learning*. London: Hodder Arnold.

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics, 31*, 532–553.

Tono, Y. (2000). A computer learner corpus based analysis of the acquisition order of English grammatical morphemes. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 123-132). Frankfurt am Main: Peter Lang.

van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *Modern Language Journal, 92,* 179-199.

Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *Modern Language Journal, 92,* 214–231.

Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing, 21,* 239–263.

Verspoor, M., & Smiskova, H. (2012). Foreign language writing development from a dynamic usage based perspective. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 17-46). Berlin: Mouton de Gruyter.

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *Modern Language Journal, 96*, 576-598.

Vyatkina, N. (2013a). Analyzing part-of-speech variability in a longitudinal learner corpus and a pedagogic corpus. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty years of learner*

corpus research: Looking back, moving ahead. *Corpora and Language in Use - Proceedings 1* (pp. 479-491). Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
Vyatkina, N. (2013b). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *Modern Language Journal, 97*(s1), 11-30.
Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing, 21,* 321–331.
Wolfe–Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Second Language Teaching & Curriculum Center, University of Hawai'i at Manoa.
Zhang, J., & Lu, X. (2013). Variability in Chinese as a foreign language learners' development of the Chinese numeral classifier system. *Modern Language Journal, 97*(s1), 46-60.

Nina Vyatkina is an Associate Professor of German and Applied Linguistics at the University of Kansas. She is the 2009 co-recipient of the ACTFL/MLJ Paul Pimsleur Award for Research in Foreign Language Education. Her research interests include instructed SLA, applied corpus research, and longitudinal learner language development. Her articles have appeared in leading Applied Linguistics journals and she serves on Editorial Boards of Language Learning & Technology and International Journal of Learner Corpus Research.

Hagen Hirschmann is a Postdoctoral Researcher at the Institute of German Language and Linguistics at Humboldt University Berlin. His research in the last years has focused on the grammatical description of modifiers, the modelling of German grammar for corpus linguistic applications, and corpus based language acquisition studies.

Felix Golcher is a Postdoctoral Researcher at the Institute of German Language and Linguistics at Humboldt University Berlin and works as a statistical consultant. His research interests lie in the statistical properties of written language and methodological questions of statistical analysis of linguistic data.