Interaction between grammatical accuracy and syntactic complexity at different proficiency levels: Insights from a learner corpus

Jennifer Thewissen[1,2] & Alena Anishchanka[1]

[1]University of Antwerp, & [2]Université Saint-Louis

This learner corpus study tracks the development and interactional dynamics between grammatical accuracy (GrA) and syntactic complexity (SC) at four Common European Framework proficiency levels (intermediate and advanced levels B1/B2/C1/C2). Data from the *International Corpus of Learner English* (Granger et al. 2020) were error-tagged and submitted to the automatic *L2 Syntactic Complexity Analyzer* (Lu 2010). Interaction between B1 and B2 is characterised by a competitive relationship between GrA and SC, with the most marked improvement displayed in accuracy. Development between B2 and C1 displays subtle shifts, with GrA improving slightly and SC showing mild signs of reorganisation. Between C1 and C2, a more supportive relationship appears: SC shows the most marked signs of development, with GrA continuing to subtly improve.

Keywords: grammatical accuracy, syntactic complexity, interactional dynamics, proficiency levels

1. Introduction and study rationale

The aim of this chapter is to trace the development of grammatical accuracy and syntactic complexity as a function of proficiency level in L2 English writing and to better understand the interactional relationship between the two constructs as learners' writing proficiency develops. Complexity and accuracy regularly appear in Second Language Acquisition (SLA) research as two major components of the CAF (complexity-accuracy-fluency) triad which serves towards measuring and profiling L2 proficiency, development and performance (Housen & Kuiken 2009; Housen, Kuiken & Vedder 2012). This chapter focusses on two of these constructs, namely accuracy and complexity, leaving aside the concept of fluency, i.e. "the capacity to produce speech at normal rate and without interruption" (Skehan 2009: 510), which, as the definition indicates, has traditionally been associated with the study of speech (Götz 2013). Although the notion of 'fluency in writing' has arguably garnered substantial research attention of late (Van Waes & Leijten 2015), its precise definition and operationalisation remain rather elusive (Abdel Latif 2013). One frequently used measure which is believed to reflect fluency in writing is the number of words produced in a given amount of time. The corpus data studied in this chapter were deemed unsuitable for the study of fluency as (1) the number of words per text was predetermined in the text selection procedure (only texts of between 500 and 900 words), and (2) the L2 essays were written in untimed settings, thereby hindering words per timing calculations. The remainder of this chapter thus tightly zooms in on accuracy and complexity in L2 writing.

One issue that regularly surfaces in academic research is the need for accuracy and complexity to be studied together rather than in isolation. Referring to the CAF constructs, Larsen-Freeman (2009: 582) rightly pointed out that "if we examine the dimensions one by one we miss their interaction and the fact that the way that they interact changes with time as well". In the same vein, Plakans, Gebril and Bilki (2019: 165) stressed the need for "organic modelling to include the ecology of their relationships with each other in learning contexts". In response to these challenges, a growing body of SLA research has recently adopted this organic approach, with the awareness that by focusing on changes in one area only, the related changes happening in another related area may be overlooked (Larsen-Freeman 2006; Norris & Ortega 2009; Verspoor, Schmid & Xu 2012; Plakans et al. 2019). Dynamic Systems Theory (DST; de Bot & Larsen-Freeman 2011; Lowie et al. 2020; Yu & Lowie 2020) is particularly relevant for this chapter as it has provided theoretical and empirical evidence that complexity and accuracy are multidimensional, dynamic and interconnected constructs that are impacted by a complex interplay of multiple factors, including L1 background, L2 proficiency level, instructional style, motivation levels, personality traits, to name but a few. Learner corpus research (LCR) to-date has done formidable work in dissecting complexity (e.g. Paquot 2019), accuracy (e.g. Thewissen 2015) and fluency (e.g. Götz 2013) but, compared to SLA research, has very much tended to do so by zooming in on each construct separately.[1] This observation provides the impetus for the present learner corpus chapter which studies the developmental dynamics at play between accuracy and complexity at the intermediate and

[1] Although some recent LCR studies have arguably investigated certain aspects of complexity, accuracy and/or fluency together (e.g. Alexopoulou et al. 2017; Brand & Götz 2011), such endeavours remain the exception rather than the rule.

advanced levels of the *Common European Framework of Reference* (CEFR; Council of Europe 2001) levels, namely B1, B2, C1 and C2. The corpus data were taken from the *International Corpus of Learner English* (ICLE; Granger et al. 2020) which is representative of argumentative student writing. Focus lies specifically on L2 grammatical and syntactic development, with the corpus data having been annotated with a total of 56 indices to capture development in these areas. The overarching research question at the heart of this chapter is the following:

RQ: How do grammatical accuracy and syntactic complexity develop across the intermediate and advanced CEFR proficiency levels and how do the two constructs interact?

In the sections that follow, we first define the constructs of grammatical accuracy and syntactic complexity and draw from DST to discuss the different types of interactions one may expect between the two. The question of whether more syntactic complexity necessarily means better text quality is also discussed as it constitutes an important aspect of this chapter. Subsequently, the corpus data and methodology employed in this study are outlined, followed by the quantitative analysis of grammatical accuracy and syntactic complexity which outlines their dynamic development across the levels as well as the nature of their interaction.

2. Construct definitions: Grammatical accuracy and syntactic complexity

Compared to complexity and fluency, accuracy is regularly presented in the literature as the most transparent and easiest construct to operationalize (Housen, Kuiken & Vedder 2012; Pallotti 2009). The consensus is that accuracy is conceptualized as "freedom from errors" (Foster & Skehan 1996: 303–304) or "form-as-conservatism" (Skehan & Foster 2012: 200) and is thus viewed as the degree of *control* which learners exert over the L2 system. While it is true that there is generally a clearer understanding of what accuracy is compared to complexity and fluency, we nevertheless consider it misleading to characterise it as the 'easiest' CAF construct. The main reason for this is that accuracy is operationalised via Error Analysis (EA) which relies on the systematic analysis of learner errors that are defined as "a linguistic form or combination of forms which, in the same context and under similar conditions of production would, in all likelihood, not be produced by the speakers' native speaker counterparts" (Lennon 1991: 182). Far from being easy, identifying 'forms which would not be produced by native speakers' is in fact one of the core challenges of error analysis, especially in higher-proficiency L2 productions, where errors can occur in subtle areas and are often a question of infelicitous rather than strictly erroneous language. The error analysis literature (Díez-Bedmar 2021; Thewissen 2021) has highlighted a number of core issues faced by error analysis researchers which prove that the study of accuracy is not transparent and easy. These include the usage norms (speech vs. writing; native vs. non-native) against which the presence of errors should be ascertained, which accuracy measures to apply (Polio & Shea 2014), and inter-rater reliability concerns in the error identification and tagging processes (Díez-Bedmar 2021).

The current study defines grammatical accuracy as the level of correctness with which learners adhere to descriptive English grammatical rules in the context of argumentative writing. Grammatical accuracy in this chapter is operationalised based on the classification of grammatical errors described in the *Louvain Error Tagging Manual* (Dagneaux et al. 2008) (see Table 2 in the methods section): errors in grammar are objectively identified based on the part-of-speech affected by the error (e.g. article, verb, noun, adjective, etc.) as well as the nature of the error (e.g. article confusion, pluralisation, wrong tense, etc), e.g. **the life is beautiful* vs. *life is beautiful*. Errors pertaining to syntax (missing and redundant words, erroneous word order) are also considered, e.g. *think about *how would be your house without the last century's inventions* vs. *think about how your house would be.*

L2 complexity has recently received considerable research attention, both in SLA and learner corpus studies. (Lu 2010, 2011, 2017; Bulté & Housen 2012, 2014; Yang, Lu & Weigle 2015; Michel et al. 2019; Vyatkina & Housen 2021). The individual foci of the chapters in the present volume are testimony to the surge of activity in L2 complexity research, with five chapters devoted to that construct, three pertaining to accuracy, two to fluency and the current chapter that studies the dynamic interactions between accuracy and complexity. One of the reasons (though by no means the only one) for the hive of activity in L2 complexity research may be the existence of a number of Natural Language Processing tools that enable the automatic analysis of syntactic complexity measurements. The *L2 Syntactic Complexity Analyzer* (L2SCA; Lu 2010), the *Tool for the Automatic Analysis of Syntactic Sophistication and Complexity* (TAASSC; Kyle 2016; Kyle & Crossley 2018), *CohMetrix* (McNamara et al. 2014) and the *Common Text Analysis Platform* (Chen & Meurers 2016) are some of the best-known available syntactic complexity analysis programmes. One advantage of capturing L2 complexity automatically is that it allows for the analysis of much larger data samples than when studying accuracy which is dependent on detailed manual annotation methods and thus tends to be limited to comparatively smaller amounts of data.

L2 complexity is often described in the literature as the most 'complex' of the CAF dimensions (Housen & Kuiken 2009). It is known as "form-as-ambition" (Skehan & Foster 2012: 200) which corresponds to linguistic risk-taking and is understood to encompass several subconstructs, including syntactic complexity and lexical complexity (Bulté & Housen 2012) as well as the newly identified dimensions of phraseological complexity (Paquot 2019) and morphological complexity (De Clercq & Housen 2019). This paper zooms in on syntactic complexity which, following Pallotti (2015: 119), we momentarily define neutrally “as a purely descriptive category”. Specifically, the current study operationalises syntactic complexity as a set of 14 objective measures described in Lu’s (2010) L2SCA which targets (1) length of production, (2) amount of coordination, (3) amount of subordination, and (4) degree of phrasal sophistication (see Table 3 in the methods section). A variety of other syntactic complexity measures arguably exist, as shown in Bulté and Housen (2014) who list a total of more than 40 different complexity measures taken from a sample of 40 empirical L2 studies. However, given the exploratory nature of this study, we focused on a limited set of variables, specifically those included in the L2SCA and identified as reliable syntactic complexity indices to trace L2 development (Norris & Ortega 2009; Lu 2011; Polio & Yoon 2018).

3. The dynamic relationship between L2 accuracy and complexity

Much SLA scholarly effort has been invested into studying accuracy and complexity as a set of interconnected components of L2 development (e.g. Ellis & Yuan 2004; Skehan 2009; Spoelman & Verspoor 2010; Verspoor et al. 2012; Lambert & Kormos 2014; Alexopoulou et al. 2017). Dynamic Systems Theory (e.g. de Bot & Larsen-Freeman 2011; Yu & Lowie 2020) constitutes the theoretical backbone of many of these studies and has revealed itself to be a particularly insightful framework against which to study L2 development. DST sees language learning as a dynamic process characterised by progress, backslide, stagnation and sudden jumps (Rosmawati 2014). As a theory taken up mainly by SLA researchers, it presents a number of key methodological differences with learner corpus studies: (1) DST adopts a process-oriented approach which involves studying numerous observations of the dynamic development within individuals over time (Yu & Lowie 2020), as opposed to the product-oriented approach more typical of learner corpus studies that focus on accounting for differences between groups of learners. The aim of the process-oriented approach is to capture the intra- and inter-individual variability that characterises L2 development. (2) DST studies frequently rely on a small number of learners (sometimes as few as one or two) compared to learner corpus work that values big learner groups for generalisation purposes.

In spite of such fundamental differences, it is argued that DST and its already extensive body of research have much to offer LCR as present-day learner corpus studies are also starting to rise to the challenge of looking at L2 language learning developmentally rather than statically at one point in time (Hasko 2013). Specifically for this chapter, DST has yielded insightful terminology which describes the types of changing relationships that the language subsystems (including accuracy and complexity) show over time. Different subsystems may be in a precursor, competitive or supportive relationship (Chan, Verspoor & Vahtrick 2015). In a precursor relationship, one system needs to be in place before another can develop: e.g. learners will first start building short subject-verb-object sentences before they attempt subordinate constructions. In a competitive relationship, one system develops at the expense of the other. This refers to Skehan's (2009) well-documented Trade-off Hypothesis that predicts a 'natural' tension between accuracy and complexity: under certain circumstances, raised levels in one performance area (e.g. accuracy) may take attention away from the other area (e.g. complexity). A supportive relationship occurs when the subsystems grow simultaneously as "connected growers" (Lowie & Verspoor 2007: 12).

The relationships between the subsystems may vary from very strong to very weak and may change altogether over time in what have been called moments of self-organisation (Spoelman & Verspoor 2010). When the language resources of the learner change from one state to another, this results in increased variability, non-linearity and chaos. Having gone through this transitional phase, the resources restructure and the newly created system may be different from the earlier system, but a key point is that learner language, even at advanced levels, never stops reorganising and changing (Larsen-Freeman 2018). An additional noteworthy concept is that language is "soft-assembled" (Thelen & Smith 1994), i.e. dynamically moulded to suit the immediate environment. The extent to which learners are able to sculpt their production to suit the communicative context depends on various factors, such as their levels of linguistic proficiency, mother tongue, instructional background, levels of motivation, personality types and cognitive styles, to name but a few (see Kuiken et al. 2019). It follows from DST observations that "it is illusory to think that what we are measuring in CAF is some kind of universal construct that can be applied across all possible learners and contexts" (Norris & Ortega: 2009: 573). The current chapter is keenly aware of this and cautiously limits its findings to the corpus data sample investigated here (see the methods section).

4. Is more syntactically complex better?

Pallotti (2009: 597) raises a key issue when he writes that "many studies seem to implicitly assume that higher levels of CAF are 'better' and that less CAF depends on limitations in language processing capacities." This question is particularly relevant regarding syntactic complexity, with Ortega (2015) explicitly asking whether more syntactically complex writing necessarily means 'better' writing (Ortega 2015). Housen et al. (2019) stress that the term L2 complexity has been equated with notions such as 'more advanced', 'more proficient', 'rarer', 'more mature', or 'better', to name but a few. The authors emphasise that L2 complexity must be clearly distinguished from the above notions which constitute conceptually separate constructs. In other words, although L2 complexity and text quality are interrelated, increased complexification does not de facto mean increased L2 quality. Rather, for increased complexity to make a text 'better', it needs to be *the right type of complexity* for a given communicative context. This is what Ortega (2015: 87) means when she says that "from both developmental and educational perspectives it is important to understand better how and *what kinds of syntactic complexity* contribute to making writing "better"" (our emphasis).

Biber, Gray and Poonpon (2011) carried out extensive corpus-based research about the type of syntactic complexity that characterises different genres. Taking English native-speaker data as their baseline, they found that spoken genres rely more on syntactic embedding (e.g. dependent clauses) than academic writing which favours a more compressed style, as evidenced by the reliance on noun phrases and greater phrasal modification. In the same vein, Crossley and McNamara (2014) longitudinally studied the syntactic complexity development of 57 learners of English as a second language in an L2 writing course. They found that over the course of a semester the learners produced texts that were increasingly aligned with the academic conventions described in Biber et al. (2011) (more nouns and phrasal complexity). Interestingly however, the human raters who were asked to assign a score to the learner texts continued to assess the L2 essays on the basis of structures more typical of spoken discourse (clausal embedding). The study thus found that although the learner productions were getting 'better' when compared to corpus-attested descriptions of syntactic complexity in academic writing, the L2 raters did not necessarily value these features (nouns and phrasal complexity) as signs of 'better' performance. There was thus a mismatch between corpus-attested descriptions of academic syntactic complexity features and raters' perceptions of these.

The above finding points to a wider underlying issue in the study of L2 syntactic complexity, namely the elusiveness of what constitutes L2 complexity for some teachers and assessors, how its features change depending on the communicative context and how to encourage its growth in pedagogical settings. Recent research on the pedagogical treatment of L2 complexity has brought to the fore that this construct is misunderstood in L2 teaching and assessment, with focus shifting to accuracy instead: "particularly in writing tasks, teachers tend to focus on accuracy, as this is relatively easy to assess, explain, and grade" (de Graaff 2019: 253). Kuiken and Vedder (2019) is a noteworthy study in this respect. The researchers investigated two groups of language teachers (eleven teachers of L2-Dutch and sixteen of L2-Italian), who were asked to individually evaluate the syntactic complexity of a sample of argumentative texts written by L2 university students of Dutch and Italian. They subsequently discussed the motivation behind their assigned scores and the qualitative feedback they gave the learner texts. A striking finding was that, although the teachers had been explicitly instructed to zoom in on syntactic complexity when rating the texts, their primary motivation for giving a particular score was levels of accuracy and comprehensibility. When analysing the qualitative feedback, the learners were frequently provided with accuracy-promoting comments such as "*Keep it simple. Short sentences. Main clause plus subordinate clause is OK. NOT: subordinate clause within subordinate clause*" (Kuiken & Vedder 2019: 239) or "*Pay more attention to agreement in gender and number within the NP!*" (ibid: 239). This instructional style is likely to inhibit the development of L2 syntactic complexity by fostering a 'play-it-safe' attitude on part of the learners. Another key finding is that when the teachers did explicitly refer to syntactic complexity, they mainly concentrated on subordinate clauses, less so on coordination and very rarely on phrasal complexification. This points to a limited conceptualization of what linguistic complexity entails in concrete terms and how it should be made to suit a given genre. De Graaff (2019) suggests that, to drive L2 development forward, it is as important to encourage complexity as it is to control for accuracy. He promotes the following message: "*Avoid avoidance. Don't keep it simple. Take a walk on the wild side*." (p. 250).

The above section has defined the constructs (grammatical accuracy and syntactic complex) at the heart of this chapter; it has drawn from DST terminology to describe the types of dynamic interactions likely to be expected between the two constructs, and it has highlighted the complex relationship between syntactic complexity and L2 writing quality. We now proceed to the description of the data, methodology and results yielded by the analysis of the ICLE corpus sample investigated in this chapter.

5. Data and methodology

5.1 Learner corpus data

To capture the development of and interaction between grammatical accuracy and syntactic complexity in English learner writing, this study relies on data from ICLE which includes texts by third- and fourth-year university learners of English from 26 L1 backgrounds. The essays investigated here were initially selected on the basis of the learners' first language background and were randomly taken from the L1 French, German and Spanish ICLE components.[2] In terms of genre, the texts are categorised as argumentative and are written on a variety of different topics, such as *The prison system is outdated. No civilised society should punish its criminals: it should rehabilitate them;* or *Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value.*

Each text is between 500 and 900 words in length. A total of approximately 50,000 tokens was selected per L1 group, amounting to c. 70 learner essays per L1 and a total corpus sample of about 150,000 tokens (223 texts overall). The restricted size of the corpus is due to the fact that each text was to be manually annotated for errors. One limitation of the ICLE corpus is that it does not include information about each essay's individual proficiency level. This feature needed to be added to the data as a separate subsequent step: the 223 essays were rated by two professional raters (+ 1 rater in the case of

[2] These L1 backgrounds were selected because of the first author's knowledge of French and German, which is helpful to understand the reason behind certain errors (e.g. transfer effects). Spanish was selected as an L1 given our collaborative research ties with Spanish colleagues also involved in error analysis work.

disagreement between the first two raters).[3] The CEFR proficiency descriptors were used for the proficiency assessment and included the following criteria: vocabulary control, grammatical accuracy, orthographic control, vocabulary range, coherence and cohesion. The raters assigned a CEFR level to each of the above criteria and additionally provided a holistic score (either B1, B2, C1, C2) for each text.[4] The mean of the holistic scores given by the two (or three) raters was subsequently computed to assign a final CEFR score to each production (see Thewissen 2015 for a more detailed description of the rating procedure).

The breakdown of texts per CEFR score is presented in Table 1. It shows proficiency level differences between the L1 groups, with the texts written by German speakers mainly situated in the advanced C1/C2 range, the French-speaking texts equally spread across the upper-intermediate B2 and early advanced C1 levels, and the Spanish-speaking data firmly representative of the lower-intermediate B1 level.

Table 1. Breakdown of ICLE texts per CEFR score

	French	**German**	**Spanish**	**TOTAL**
B1	5	8	53	66
B2	33	11	18	62
C1	30	30	7	67
C2	6	22	0	28
	74	71	78	223

[3] The inter-rater reliability agreement for the whole corpus was substantial, with an overall *r* value of 0.70 for the scores attributed by raters 1 and 2.
[4] Because the *ICLE* texts were written by learners who were already in their third or fourth year of university (modern language programme), the data are representative of the intermediate and advanced proficiency levels. No beginner-level texts are included in the corpus.

-

Because the proficiency level information needed to be fed into the data after the initial L1-based text selection, it was not possible to ensure the presence of an equal number of texts for each L1 per proficiency level. The result is therefore a rather unbalanced corpus sample with, for example, five French texts at level B1 but 53 Spanish essays at the same level. This has an impact on the analysis, making it challenging to yield reliable results on the possible influence of the L1 on developmental profiles across proficiency levels. From Table 1, it is nevertheless clear that the results at proficiency level B1 are mainly based on the Spanish data; those for B2 are slightly more mixed, with a majority of L1 French texts and a more or less similar number of L1 German and Spanish scripts. The data for C1 are equally represented by L1 French and German texts. C2 level is mostly populated by the L1 German scripts. No L1 Spanish texts were rated as C2.

5.2 Grammatical accuracy and syntactic complexity indices

We used the computer-aided error analysis method (Dagneaux, Denness & Granger 1998) to manually annotate the grammatical inaccuracies detected in each text. Relying on the *Louvain Error Tagging Manual* (see Section 2), grammatical accuracy is operationalised as a multidimensional construct, with each individual error type explained and exemplified in Table 2.[5] Errors were counted using the potential occasion analysis (POA) method (Thewissen 2015). As its name indicates, POA involves counting the errors of a particular type out of the number of times it could *potentially* have been made rather than out of the total tokens in the data. For instance, errors which involve dropping the third person singular *–s* are best counted as a proportion of all verbs rather than out of the total tokens in the data as it is only verbs which represent potential occasions for this error. POA thus involves carefully considering the best *denominator* out of which to count the different error types instead of de facto relying on the total tokens as the standard denominator. For the grammatical errors in the data, tailor-made part-of-speech (POS)[6] denominators were created to fit each grammatical error type as well as possible (see Thewissen 2015 for an exhaustive account of how each denominator was created and to what extent they are suitable for the individual error types). For example, the 'nounALL' POS denominator was created to count noun number agreement errors (e.g. *three *chair*) which can only occur on nouns (common or proper); similarly, tense errors were counted out of the total lexical verbs per text. Although sometimes imperfect, POA was nevertheless deemed a more realistic environment for potential error occasions than the total tokens in the data. For the more syntactic error types (e.g. word missing, word order, etc.), the total

[5] The following error coding system was used: the tags were manually inserted in the data, with the error code placed in front of the error itself and the correction included between dollar signs following the error.
[6] The data were POS tagged with the CLAWS7 tagger (Garside & Smith 1997).

number of sentences was considered the most suitable denominator as these errors very much occur at the sentential level. Grammatical errors due to word class confusion (e.g. use of an adjective instead of an adverb) were nevertheless counted out of the total tokens as most word classes can be the object of word class confusion. Table 2 specifies out of which denominator each grammatical error type was counted.

Table 2. Grammatical error operationalization (examples taken from the *Louvain Error Manual*; Dagneaux et al. 2008)

Error tags and tag name (counting denominator)	**Tag description**	**Example**
GA: Grammar, article (all nouns)	Article errors involving the definite, indefinite or zero article (sentence grammar error)	*(GA) The 0 life is beautiful.*
GADJCS: Grammar, adjective, comparative/superlative (adjectives)	Errors on the comparative or superlative use of an adjective	*The role that women should play in a (GADJCS) more fair $fairer$ present-day society*
GADJN: Grammar, adjective, number (adjectives)	Adjective erroneously used in the plural	*The last sentences have been (GADJN) favourables $favourable$ to women.*
GADJO: Grammar, adjective, order (adjectives)	Erroneously ordered adjectives	*A (GADJO) leather black small $small black leather$ handbag*
GADVO: Grammar, adverb, order (adverbs)	Misplaced adverb	*They (GADVO) see only $only see$ other criminals.*
GNC: Grammar, noun, case (all nouns)	Errors involving the use of the Saxon genitive	*Behind the (GNC) Berlin's wall $Berlin wall$*
GNN: Grammar, noun, number (all nouns)	Addition or omission of the plural morpheme on nouns	*Bearing in mind that sex equality is one of the great (GNN) reason $reasons$ for fights in most places around the world (…)*
GPD: Grammar, pronoun, demonstrative (demonstratives)	Errors on demonstrative pronouns	*What is harassment? The dictionary says (GPD) that it is a behaviour which is intended to trouble or annoy someone.*

Error tags and tag name (counting denominator)	Tag description	Example
GDD: Grammar, determiner, demonstrative (demonstratives)	Errors on demonstrative determiners	*(GDD) This $These$ elements cannot be separated.*
GPP: Grammar, pronoun, personal (personal pronouns)	Errors on personal pronouns	*The big majority of children have a computer or a video-game, with which (GPP) 0 $they$ spend (waste, in my opinion) a great number of hours.*
GPO: Grammar, pronoun, possessive (possessives)	Errors on possessive pronouns	*My computer did not cost as much as my sister's. (GPO) His $Hers$ was more sophisticated.*
GDO: Grammar, determiner, possessive (possessives)	Errors on possessive determiners	*People accept jobs according to how much they get paid but not according to (GDO) his $their$ preferences.*
GPI: Grammar, pronoun, indefinite (indefinites)	Errors on indefinite pronouns	*A lower-class man is not on an equal footing with his middle- or upper-class (GPI) one $counterpart$.*
GDI: Grammar, determiner, indefinite (indefinites)	Errors on indefinite determiners	*He does not have (GDI) some any expectations.*
GPF: Grammar, pronoun, reflexive/reciprocal (reflexives & reciprocals)	Errors on reflexive or reciprocal pronouns	*They didn't need to communicate (GPF) themselves 0 outside their homes.*
GPR: Grammar, pronoun, relative/ interrogative (relative pronouns)	Errors on relative orinterrogative pronouns	*The government took several measures to stop the strikes, (GPR) that $which$ was not effective.*
GPU: Grammar, pronoun, unclear reference (personal pronouns)	Use of a pronoun whose reference is unclear	*But there are also imprisoned people waiting for their execution who are innocent. They never had a fair trial and a real chance to get out of (GPU) it $jail$. These people often do not have enough money to get their own attorney.*
GVAUX: Grammar, verb, auxiliary (modal auxiliaries)	Misuse of modal auxiliaries (*can, should, may*,etc.), primary auxiliaries (*be, do, have*) or semi-auxiliaries (*ought to, used to, dare, need*)	*So if there is an army it (GVAUX) might $should$ be professional, and formed by people who believe in that and want to dedicate their lives to it.*
GVM: Grammar, verb, morphology	Erroneous use of existing verb forms (e.g.	*It is generally (GVM) agree $agreed$ today that we live*

Error tags and tag name (counting denominator)	Tag description	Example
(lexical verbs)	a simple past form instead of a past participle form, an infinitive instead of a past participle)	*in a world where television plays an important part.*
GVN: Grammar, verb, number (all nouns)	Subject-verb agreement errors	*How do you think that a man (GVN) react $reacts$ when he hears that a woman is shocked when she receives a letter beginning with "Dear Sirs"?*
GVNF: Grammar, verb, non-finite/finite (lexical verbs)	Errors involving non-finite/finite verb forms	*(GVNF) To travel $Travelling$ by public transport is recommended.*
GVT: Grammar, verb, tense (lexical verbs)	Erroneously chosen tense or aspect	*He learned a profession in the prisson, and now he (GVT) wrote $writes$ poetry and (GVT) took $takes$ part in the publication of a prisson's journal.*
GVV: Grammar, verb, voice (lexical verbs)	Use of the active instead of passive voice or passive instead of active voice	*This seems impossible to (GVV) be achieved $achieve$.*
GWC: Grammar, word class (total tokens)	Inappropriate use of a word class: adjective used instead of a noun, adverb instead of an adjective, etc.	*We are going to review the following subjects: Labour discrimination, the right to vote, the fight against male (GWC) chauvinist $chauvinistic$ behaviours.*
WM: Word missing (sentences)	Omission of words, except pronouns, dependent prepositions, articles, connectors, auxiliaries	*The future soldiers make an strict physical training and (WM) 0 sit some exams.*
WO: Word order (sentences)	Problems of word order that do not fall into the categories of Adverb Order (GADVO) or Adjective Order (GADJO)	*Think about (WO) how would be your house $how your house would be$ without the last century's inventions.*
WRS: Word redundant singular (sentences)	Unnecessary use of a single word, except articles (GA), connectors (LC*), pronouns (GP*), dependent prepositions (X*PR) and auxiliaries (GVAUX)	*Actual life is very complicate and (WRS) extremely 0 full of worries.*

Error tags and tag name (counting denominator)	Tag description	Example
WRM: Word redundant multiple (sentences)	Unnecessary use of multiple words, except articles (GA), connectors (LC*), pronouns (GP*), dependent prepositions (X*PR) and auxiliaries (GVAUX)	*Others comment that (WRM) the fact is that 0 once you are inside, if you like it, you can even re-enlist.*
XADJPR: Lexico-grammar, adjective, preposition (adjectives)	Adjective used with an erroneous dependent preposition	*How many public places are easily (XADJPR) accessible for $accessible to$ wheelchairs?*
XNCO: Lexico-grammar, noun, complementation (common nouns)	Erroneous complementation of nouns	*Students have the (XNCO) possibility to leave $possibility of leaving$.*
XNPR: Lexico-grammar, noun, preposition (common nouns)	Nouns used with an erroneous dependent preposition	*He has a (XNPR) thirst of $thirst for$ knowledge.*
XNUC: Lexico-grammar, nouns, uncountable/ countable (common nouns)	Errors involving the countable use of uncountable nouns	*The tremendous (XNUC) progresses $progress$ realized by science have disrupted our habits and our way of living.*
XVCO: Lexico-grammar, verb, complementation (lexical verbs)	Erroneous complementation of verbs	*What about the people who cannot (XVCO) afford going $afford to go$ to these kind of centres?*
XVPR: Lexico-grammar, verb, preposition (lexical verbs)	Verb used with an erroneous dependent preposition	*The classroom must often have (XVPR) resembled to $resembled$ a "Chamber of Horrors".*

In addition to being annotated for grammatical errors, each text was analysed using the L2SCA (see Section 2). For each script, the tool calculates a number of frequencies (e.g. words, verb phrase, clauses, etc.) as well as 14 syntactic complexity ratios which are described in Table 3.[7]

[7] For more refined indices of syntactic complexity, the data can be submitted to TAASC (*Tool for the Automatic Analysis of Syntactic Sophistication and Complexity;* Kyle 2016), which goes beyond the more classic measures in the L2SCA to encompass as many as 372 indices in five

Table 3. Syntactic complexity measures in L2SCA (based partly on Kyle & Crossley 2018 and Polio & Yoon 2018)

Abbreviation	**Syntactic structures**	**Description**	**Example**
W	Word	Series of letters separated by a white space	*I* *Think*
VP	Verb phrase	Finite or non-finite verb phrase	*am tired* *to go there*
C	Clause	Subject + finite verb structure	*I am tired* *because I worked all day*
DC	Dependent clause	Finite adverbial, nominal or adjective clause	I am tired *because I worked all day.*
T	T-Unit	A main clause and all its dependent clauses	*I am tired as I have been doing all the housework.*
CP	Coordinate phrase	Adjective, adverb, noun and verb phrases linked by a conjunction of coordination	She *looks tired but happy.*
CN	Complex nominals	Noun with modifiers, gerunds and infinitives as subjects	*Beautiful smile* *Hiking* is healthy.
S	Sentence	A group of words with sentence-ending punctuation	*Do you want to leave now?* *She looks tired but happy.*
Syntactic constructs	**Measures**	**Formula**	
Length of production unit			
MLC	Mean length of clause	No. words/no. clauses	
MLS	Mean length of sentence	No. words/no. sentences	
MLT	Mean length of T-unit	No. words/n. T-units	
Sentence complexity			
C/S	Sentence complexity ratio	No. clauses/no. sentences	
Subordination			
C/T	T-unit complexity ratio	No. clauses/no. T-units	

categories: clause complexity (32 indices), phrase complexity (132 indices), syntactic sophistication (190 indices), syntactic component scores (9 indices) and classic syntactic complexity indices (14 indices).

CT/T	Complex T-unit ratio	No. complex T-units/ no. T-units
DC/C	Dependent clause ratio	No. dependent clauses/no. clauses
DC/T	Dependent clauses per T-unit	No. dependent clauses/no. T-units
Coordination		
CP/C	Coordinate phrases per clause	No. coordinate phrases/no. clauses
CP/T	Coordinate phrases per T-unit	No. coordinate phrases/no. T-units
T/S	Sentence coordination ratio	No. T-units/no. sentences
Phrasal sophistication		
CN/C	Complex nominals per clause	No. complex nominals/no. clauses
CN/T	Complex nominals per T-unit	No. complex nominals/no. T-units
VP/T	Verb phrases per T-unit	No. verb phrases/no. T-units

5.3 Study design

A combination of quantitative and qualitative methods was used to explore the interactions between the constructs of syntactic complexity and grammatical accuracy at two levels of granularity. First, we applied multidimensional scaling (MDS) to study the correlational structure between the measures of syntactic complexity and grammatical accuracy at the most general level. This macro-analysis provides a bird's eye view of the distribution of the variables and interaction between the two constructs. As a second step, the available measures were tested using one-way analysis of variance (ANOVA) and post-hoc Tukey HSD tests to identify significant differences in syntactic complexity and grammatical accuracy between texts at the four CEFR levels and to infer information about the nature of their dynamic interactions.

6. Results

6.1 Mapping the grammatical accuracy and syntactic complexity constructs

MDS was applied as an exploratory visualization method to identify the distribution of the 34 measures of grammatical accuracy Table 2) and 22 syntactic complexity (Table 3) and their potential interaction at the most general level. An ordinal MDS model was fitted using the MDS algorithm based on stress minimization by means of majorization (smacof).[8] Proximities for each pair of variables included in the analysis were derived from the matrix of pairwise correlations for all variables calculated as the absolute values of the Kendall rank correlation coefficient *tau*. This correlation coefficient was selected as the most adequate measure of similarity between the variables, given their different range and scale (absolute frequencies, relative frequencies and ratios).

[8] The analysis was carried out with the R-CRAN package *smacof* (de Leeuw & Mair 2009; Mair et al. 2021).

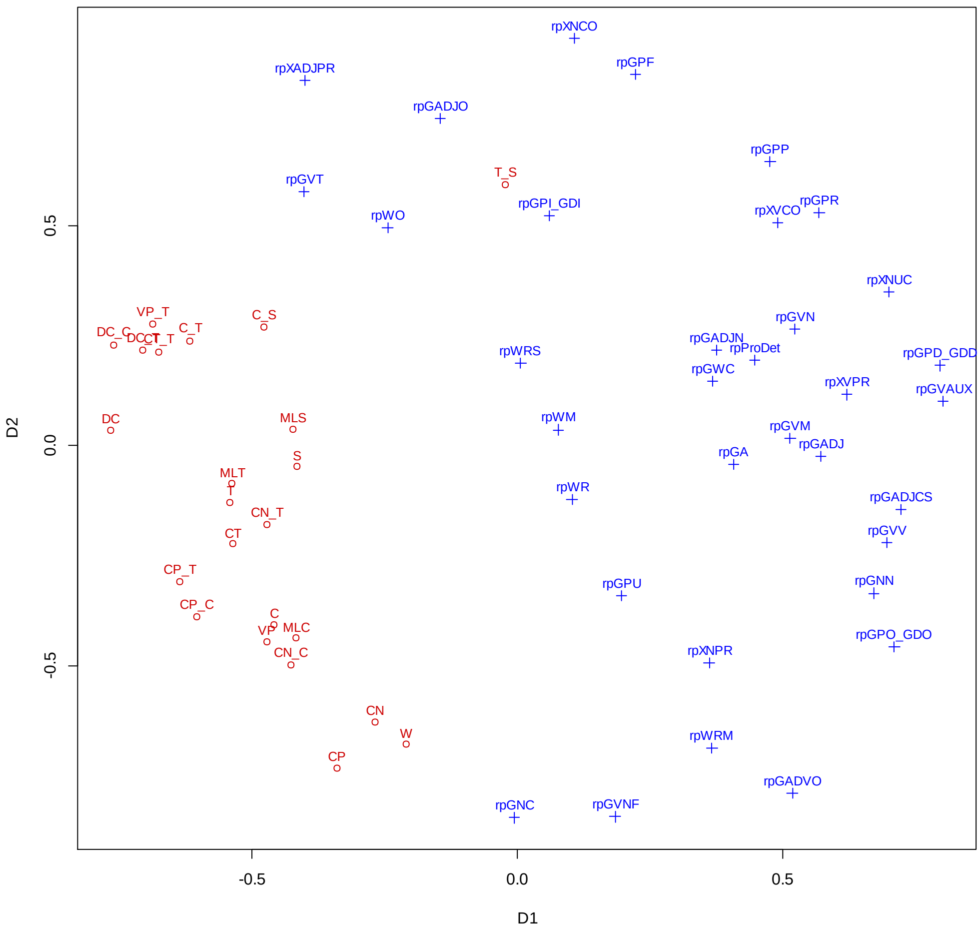

Figure 1. Exploratory MDS solution for 56 measures of grammatical accuracy (+) and syntactic complexity (○) (*stress-1 = 0.3*)

Figure 1 represents the two-dimensional MDS solution for 56 measures of syntactic complexity and grammatical accuracy. The MDS solution shows a clear configuration with two distinct clusters: measures of syntactic complexity (○) and grammatical accuracy (+) form two distinct groups along dimension 1 with little overlap. One exception is the ratio of T-units per sentence (T/S). The group of syntactic complexity measures on the left is denser and more compact, suggesting higher within-group correlations between the variables. The grammatical accuracy indices on the right are more dispersed in the MDS space, which can be attributed to the more heterogeneous nature of the included measures corresponding to the different error types.

The observed pattern in the distribution of variables calls for two comments. On the one hand, the compact clustering of the syntactic complexity measures suggests that they are similar in nature and measure closely related aspects of written texts. As has been pointed out by Norris and Ortega (2009: 560), there is a certain degree of redundancy in the syntactic complexity measures calculated by the L2SCA and clustered on the left side of the graph which needs to be taken into account when measuring this construct. On the other hand, the clear separation between the two clusters of metrics shows that grammatical accuracy and syntactic complexity are very much two distinct constructs that each require their own set of measurements.

6.2 Capturing dynamic developmental patterns

The data sample was submitted to a one-way analysis of variance (ANOVA) to determine whether significant differences exist in grammatical accuracy and syntactic complexity between the B1, B2, C1 and C2 proficiency levels. Complexity measures based on frequency counts (W, VP, C, DC, T, CP, CT, CN, S) were excluded from this analysis. Since these measures are sensitive to text length, even statistically significant results can be biased. Appendix 1 lists the descriptive results for all the indices used. The ANOVA test is particularly suitable for comparing mean differences of continuous variables across different proficiency levels (Jin, Lu & Ni 2020). The CEFR level constituted the independent variable and the accuracy and complexity indices the dependent variables. In addition, the Tukey HSD (Honestly Significant Difference) post-hoc test was used to identify statistically significant differences between the means for the four CEFR levels. The left part of Table 4 shows the detailed ANOVA results for the 48 measures included in the analysis, providing the F score, *p*-value and effect size (*eta*-squared). Figures highlighted in bold are statistically significant at the level $p<0.05$. The right part of Table 4 lists the results of the Tukey test including the differences between the mean values for each pair of adjacent levels: e.g. 4.37 for MLS in the C1-C2 column indicates that there are on average 4.37 more words per sentence in C2 texts. Values for the differences between proficiency levels should be interpreted as follows: negative scores mean that the level of syntactic complexity decreases, while positive values mean that it increases. Negative accuracy scores indicate fewer errors and thus an increase in accuracy, while positive values point to a higher error mean and thus a decrease in accuracy.

	ANOVA			Tukey HSD		
Measure	**F-score**	**p-value**	**Effect size η^2**	**B1-B2 difference**	**B2-C1 difference**	**C1-C2 difference**
Grammatical accuracy						
GA	39.215	**0.000**	0.351	**-3.07**	-0.82	-0.87
GADJCS	6.205	**0.000**	0.079	**-0.42**	-0.12	-0.03
GADJN	8.144	**0.000**	0.101	**-0.80**	0.03	0.02
GADJO	3.355	**0.020**	0.044	-0.04	**0.32**	-0.01
GADVO	5.300	**0.002**	0.068	-0.12	**-1.22**	-0.29
GNC	0.903	0.440	0.012	-0.01	-0.03	-0.12
GNN	6.358	**0.000**	0.080	-0.22	-0.29	-0.38
GPD_GDD[9]*	2.462	0.063	0.033	-0.43	-0.47	-2.03
GPP	2.412	0.068	0.032	-0.07	-1.57	-0.49
GPO_GDO*	4.134	**0.007**	0.054	-3.14	-1.11	-1.16
GPI_GDI*	5.280	**0.002**	0.068	**-3.18**	0.61	-0.34
GPF	1.795	0.149	0.024	0.96	-4.46	-1.37
GPR	3.381	**0.019**	0.044	-1.57	-2.76	-2.88
GPU	10.364	**0.000**	0.125	**-2.08**	-1.38	-0.83
GVAUX	2.517	0.059	0.033	-4.19	-0.75	-2.43
GVM	8.387	**0.000**	0.103	**-0.39**	-0.04	0.07
GVN	14.647	**0.000**	0.168	**-1.86**	-0.31	0.07
GVNF	0.806	0.492	0.011	0.04	0.03	-0.26
GVT	2.319	0.076	0.031	-0.41	-0.75	-0.07
GVV	4.729	**0.003**	0.061	-0.11	-0.14	0.02
GWC	21.077	**0.000**	0.225	**-0.15**	-0.05	-0.01
WM	17.096	**0.000**	0.190	**-3.77**	-0.62	-0.54
WO	4.058	**0.008**	0.053	-0.16	-1.08	-0.40
WRS	5.369	**0.001**	0.069	**-1.52**	-0.08	-0.44
WRM	3.839	**0.010**	0.050	-0.64	-0.43	-0.04
XADJPR	0.490	0.690	0.007	-0.13	0.02	-0.03
XNCO	1.407	0.242	0.019	-0.07	-0.05	-0.05
XNPR	6.961	**0.000**	0.087	**-0.33**	0.06	-0.13
XNUC	9.986	**0.000**	0.121	**-0.58**	-0.31	0.02
XVCO	1.491	0.218	0.020	-0.15	-0.09	-0.10
XVPR	12.407	**0.000**	0.146	**-1.48**	-0.29	-0.40

[9] A number of aggregated variables (indicated with *) were tested to identify more general patterns for several minor error types and less common errors. Variables GPD_GDD, GPO_GDO, GPI_GDI each combine two similar error types. Variable ProDet combines all errors in pronouns and determiners (GDD, GDO, GDI, GPD, GPP, GPI, GPF, GPR, GPU). Variable GADJ combines errors in the use of adjectives (GADJCS, GADJN, GADJO). Variable WR combines errors in word order (WRS, WRM).

ProDet*	32.125	**0.000**	0.307	**-2.57**	-1.21	-1.22
GADJ*	8.945	**0.000**	0.110	**-1.27**	0.23	-0.03
WR*	9.187	**0.000**	0.112	**-2.16**	-0.50	-0.48
Syntactic complexity						
MLS	8.941	**0.000**	0.110	**-4.29**	-0.04	**4.37**
MLT	7.813	**0.000**	0.097	-1.84	-0.01	**4.46**
MLC	2.672	**0.048**	0.035	-0.46	0.63	0.56
C_S	9.535	**0.000**	0.116	**-0.33**	-0.14	0.28
VP_T	6.192	**0.000**	0.079	-0.12	-0.09	**0.45**
C_T	7.969	**0.000**	0.099	-0.11	-0.11	**0.31**
DC_C	5.013	**0.002**	0.065	-0.01	-0.02	**0.08**
DC_T	6.853	**0.000**	0.086	-0.07	-0.09	**0.29**
T_S	9.087	**0.000**	0.111	**-0.09**	-0.01	-0.05
CT_T	4.807	**0.003**	0.062	-0.03	-0.04	**0.10**
CP_T	2.907	**0.036**	0.038	-0.06	0.02	0.12
CP_C	1.082	0.358	0.015	-0.02	0.03	0.02
CN_T	5.167	**0.002**	0.066	-0.23	0.01	**0.51**
CN_C	1.320	0.269	0.018	-0.07	0.07	0.07

Table 4. Results of the one-way ANOVA and Tukey HSD tests for accuracy and complexity measures

Regarding the development happening between B1 and B2 (see B1-B2 difference column), a first noteworthy finding is that half of the grammatical accuracy measures (16 out of 34) show significant between-level differences. The specific areas of grammatical accuracy significantly improving from B1 to B2 are the following:

- article choice (GA), e.g. *there are families in which* (GA) *the 0 communication does not exist anymore;*

- use of adjectives: comparative and superlative adjectival forms (GADJCS); erroneous pluralisation of adjectives (GADJN), e.g. *in fact, when these* (GADJN) *excelents $excellent$ students finish their degrees and go…*;
- mastery of pronouns and determiners (ProDet), especially the more accurate use of indefinite determiners and pronouns (GPI_GDI) and more precise pronominal reference (GPU), e.g. *they lie to their partner and* (GPI) *all $everything$ finish in divorce*;
- verb morphology (GVM) and verb number agreement (GVN) errors also considerably decrease, e.g. *the feminist movement* (GVN) *ask $asks$ for brave princesses to rescue helpless princes*;
- use of uncountable nouns (XNUC): learners at B2 are better able to identify uncountable nouns and use them grammatically accurately, e.g. *Since the industrial revolution, the tremendous* (XNUC) *progresses $progress$ realized by science have disrupted our habits*;
- the use of dependent prepositions with verbs (XVPPR) and with nouns (XNPR), e.g. *but the problem is that a lot of parents do not* (XVPR) *look for $look after$ their children the way they should*;
- errors in syntax, especially the use of redundant words (WR), are also significantly fewer between these levels, e.g. *it* (WR) *rather 0 ruins your health instead of making you fit*.

It is also important to note that, although they do not display *statistically significant development* between the levels, nearly all the remaining error categories decrease slightly from B1 to B2, as testified by the minus signs in front of the mean differences.[10] Although the results are not statistically significant, the general decreasing tendency seems to point towards an overall improvement trend in accuracy. In line with DST research, we tentatively argue here that non-statistically significant results remain a meaningful source of information on the process of language development as they point to the subtle changes that the L2 is going through. Looking at statistically significant results only means missing out on the dynamicity of written language development. Fogel (2011: 267) hints at this when he explains that "development (…) emerges from sometimes subtle momentary shifts of the system in context. Often, those shifts arise in non-obvious ways". Non-significant developmental results might be considered indicative of such subtle shifts in the L2 system.

Turning to the changes in syntactic complexity measures between B1 and B2, the analyses reveal the following trends: (1) three measures were found to significantly decrease, namely mean length of sentence (MLS) (length of production unit), clauses per sentence (C_S) (sentence complexity) and T-units per sentence (T_S) (sentence coordination); (2) interestingly, all the other mean differences, although non-statistically significant, also point towards a very

[10] There are two minor exceptions to this, namely reflexive pronoun errors (GPF) and the use of finite vs. non-finite verb forms (GVNF).

slightly decreasing trend in syntactic complexity, as shown by the minus signs in front of them. In other words, each of the 14 syntactic complexity measures considered here tends towards slightly lower syntactic complexity results at B2 than at B1. Rather than a sign of "coincidental fluctuations" (Yu & Lowie 2020: 868), this regularity might signify that B2 learners are indeed restructuring their interlanguage in such a way that they are making slightly fewer uses of syntactic complexity devices at B2 than at B1. To try and explain the statistically significant trends, the decreasing reliance on coordination (T_S) may actually be a sign of advancing syntactic complexification as the SLA literature (e.g. Norris & Ortega 2009) has shown that coordination is one of the clause linking devices most relied on by lower proficiency levels who then start using more subordinate clauses towards the intermediate levels. However, in light of the earlier discussion about the fact that syntactically more complex is not necessarily 'better', one needs to be cautious before claiming that lower mean length of sentence (MLS) and lower sentence complexity (C_S) indices are possible signs of 'regression' in these areas. At this stage, our study can merely neutrally claim that lower syntactic complexity indices were found at B2 compared to B1, without venturing any inferences about the impact of this finding on text quality.

The regularities found in the data nevertheless reveal some insightful observations regarding the interaction between grammatical accuracy and syntactic complexity as learners move from levels B1 to B2. Accuracy showed a rather clear trend towards progression: B2 learner texts are considerably more grammatically accurate than their B1 counterparts, which is somewhat of an expected finding. Interestingly, however, syntactic complexity measures were found to be slightly lower at B2 than at B1, indicating that learners were perhaps busier tending to aspects of accuracy at the expense of more marked syntactic complexity development. The results for levels B1-B2 seem to indicate that grammatical accuracy and syntactic complexity are interlinked and competitive in a dynamic system (Chan et al. 2015): As learners move from B1 to B2, grammatical accuracy appears to take precedence over syntactic complexification.

Concerning the difference between levels B2 and C1 (see B2-C1 difference column), there is at first glance little to report on in terms of statistically significant findings in grammar/syntax. Except for two marginal grammatical error types, namely adjective order errors (e.g. *a* (GADJO) *leather black small $small black leather$ handbag)* which are found to significantly increase at C1 and adverb order errors (e.g. *they* (GADVO) *see only $only see$ other criminals*) which decrease, no further significant changes in grammatical accuracy are noted. Concerning subtle developmental shifts (i.e. non-significant results), the mean differences for the majority of the other grammatical error types (26 out of 34) are once more preceded by minus signs. This indicates further fine-tuning towards slightly more grammatically accurate texts at C1 as most of the heavy work has been done by now and learners are more generally at ease accessing L2 grammatical knowledge. To use DST terminology (Chan et al. 2015), it may be that by the time learners reach level B2, the grammatical accuracy system has become more robust/well-established and, although it continues to develop, it does so less markedly than before.

Regarding the changes in syntactic complexity measures, none developed in a statistically significant way. The non-statistically significant results point to a rather variable developmental profile. Nine out of the 14 syntactic complexity indices were found to be slightly lower at C1 than at B2. These include for example mean length of production units (MLS and MLT), as well as three subordination measures (DC_C, CT_T, DC_T) and one coordination measure (T_S). Mean difference scores which were very slightly higher at C1 than at B2 include two phrasal sophistication indices, namely complex nominals per T-unit (CN_T) and complex nominals per clause (CN_C). If one considers the L2 as being in a state of constant dynamic reconfiguration, these very slight changes might indicate a subtle shift in the syntactic complexity system whereby learners are busy reorganising their interlanguage system. Fogel (2011: 267) speaks of "microscopic" change and rightly argues that "microlevel change provides the seeds for developmental (macroscopic) change".

The interaction between grammatical accuracy and syntactic complexity as learners develop from B2 to C1 is rather complex to qualify given the almost total absence of statistically capturable changes. From the microscopic developmental ‘seeds’ available in Table 2, it might be tentatively suggested that the relationship between grammatical accuracy and syntactic complexity is at a “point of transition” (Larsen-Freeman 2018: 89). The constructs are not in competition but are not exactly showing clear signs of growing in tandem either: grammatical accuracy is displaying subtle improvement, while syntactic complexity is also showing small signs of reorganisation that are quite variable at this stage. An explanation for the lack of obvious development in grammar between these levels may be that learners’ attention is geared towards lexical complexity development. Some preliminary indication that this is the case was reported by Thewissen and Anishchanka (2019) who submitted the current data to the Lexical Complexity Analyzer (Lu 2012) and found that more than half of the 25 measures of lexical complexity used in the tool developed significantly from B2 to C1. Lexical sophistication and type/token ratio measures were particularly revealing: C1 learners were found to use more sophisticated and a wider range of vocabulary in their texts than B2 learners. Although beyond the scope of this paper, carrying out an in-depth analysis of the lexical complexity development in the current dataset constitutes an interesting avenue for further research.

Concerning the development between the two most advanced levels (see C1-C2 difference column), grammatical accuracy shows a rather similar profile to the one identified between B2 and C1: It continues to subtly improve, with 30 out of the 34 accuracy indices slightly decreasing as learners reach the C2 level. Much noteworthier is the developmental activity in syntactic complexity, with as many as eight out of the 14 indices showing a significant change between C1 and C2. At C2, texts have become significantly longer (MLS and MLT), use significantly more subordination devices as all four subordination measures (CT_T, C_T, DC_C, DC_T) were found to be statistically significantly higher, and also rely statistically significantly more on phrasal complexification (VP_T, CN_T). Are these changes for the better? As explained in this chapter, in order to align with the conventions of academic writing, texts typically need to move away from a reliance on subordinate clauses in favour of denser phrasal components. The evidence yielded here partly corresponds to this trend, with a statistically significantly higher use of complex phrases. However, the C2 texts also rely significantly more on complexification via subordination, which goes against the expected development towards a more academic style. This may be because of the nature of the writing topics. The learners were asked to write argumentative essays which they may not deem to be a prototypical academic genre (Crossley & McNamara 2014). The essay prompts included a wide variety of topics which involved writing about marriage, the death penalty, the value of university education, money, drugs, children's education, etc. This can lead to writing that is of a more interpersonal register (Biber & Conrad 2009). This register implies that learners are often concerned about conveying their own feelings and attitudes, which is frequently done by using features that are more typical of spoken registers (e.g. subordination).

Concerning the interaction between grammatical accuracy and syntactic complexity as learners move towards the highest C2 proficiency level, the data indicate that it has shifted to take on a more supportive character as both develop in tandem: syntactic complexity shows significant changes between these levels and grammatical accuracy continues to slightly improve. One may also suggest that the constructs display signs of a precursor relationship (Chan et al. 2015) as it is only once the grammatical accuracy system is more firmly in place that syntactic complexification shows such active development. In other words, because C2 learners do not need to give as much cognitive attention to grammatical accuracy, this frees up space for syntactic complexification development.

The above section has described the development of grammatical accuracy and syntactic complexity across four CEFR proficiency levels and has studied the interaction between the two concepts. The broad emerging picture is that the major change in grammatical accuracy was found between the two intermediate levels (B1 and B2). The most marked development in syntactic complexity was found between the two advanced proficiency levels (C1 and C2). Surprisingly, the shift from upper-intermediate B2 to advanced C1 status was characterised by a slowed-down developmental trend, with no significant changes in either of these areas (for grammar at least). One implication of this finding might be that the way the CEFR levels have been established with two intermediate (B1 and B2) and two advanced (C1 and C2) levels may not necessarily correspond to the complexities of L2 grammatical development, especially given the fact that the existence of a universally-valid developmental trajectory is questionable (e.g. Larsen-Freeman 2018). Thewissen (2015) had already formulated this concern with respect to the construct of accuracy. The current study opens up the debate even further as the same question can be asked regarding syntactic complexity development. Although we are not in a position to provide a straight answer to the above question, the findings have pointed out the important underlying issue of the "naturalness" of the CEFR proficiency level yardstick which we argue should be borne in mind in the subsequent learner corpus developmental work.

7. Conclusion and avenues for further research

This chapter has adopted an integrated approach to the study of grammatical accuracy and syntactic complexity, a methodological decision which we claim needs to feature more prominently in learner corpus research. The integrative approach could be further expanded to also encompass other key constructs such as lexical and morphological complexity, additional areas of accuracy (e.g. formal, lexical, punctuation) as well as fluency. Although it zoomed in exclusively on aspects of grammar, this study has convincingly shown the dynamicity of L2 development and the fact that it is not stable over proficiency levels (Yu & Lowie 2020). Various interrelationships emerged from the data analysis: grammatical accuracy and syntactic complexity appeared to be in a rather competitive rapport between the B1 and B2 levels, with the main efforts going towards improving levels of grammatical accuracy and resulting in slightly lower syntactic complexity indices at B2 than at C1. A developmental lull was identified between B2 and C1, with very mild changes in both constructs. Transition from C1 to C2 was marked by significant cognitive investment in syntactic complexification, with grammatical accuracy also continuing to slightly improve but in a non-significant way. Importantly, although learner corpus work gives prominence to results which are found to be statistically significant and therefore ‘not due to chance’, we purposely also considered the trends indicated by the non-statistically significant findings which may seem random but, in their regularity, may also provide information on how the L2 system is reorganising itself. Rather than be written off as random results which should not be reported on, we believe L2 developmental work should in fact take the time to analyse such microscopic changes which may reveal themselves to be precursors of more macroscopic development.

A key issue considered in this chapter was whether higher complexity and accuracy indices necessarily indicated better L2 text quality. Few would object that more accurate is indeed better, especially in argumentative writing in an academic context. The trend towards more accurate grammar as learners progress along the proficiency continuum was confirmed in this chapter. However, the question was thornier regarding syntactic complexity. Rather than the amount of syntactic complexity per se, it is the types of syntactic devices used and whether these are suitable to a given genre which impacts L2 text quality. The data analysis results revealed that the learners in the analysed corpus sample were partly on track towards acquiring a more academic style, with a significant increase in complex noun phrases, but that they continued to rely on complexity features more typical of an interpersonal style (embedded clauses). This was hypothesised to be due, among other things, to the nature of the writing topics which favoured personal involvement. Also highlighted in this chapter was the related issue of the teachability of L2 complexity which, compared to accuracy, is still largely misunderstood, ill-assessed (Crossley & McNamara 2014) and under-taught. This line of research is deserving of further exploration to make concrete suggestions about the implementation of efficient teaching methods which encourage L2 complexification.

One caveat of the current study is the unbalanced nature of the corpus sample: the L2 texts were initially selected on the basis of the L1 background of the learners, without the possibility of simultaneously controlling for proficiency level. The assessment of each individual text considered in this chapter was thus carried out as a separate step. This implies that the combined impact of the L1 and the proficiency level on the grammatical accuracy and syntactic complexity profiles could not be reliably tested in a multivariate analysis. One way forward for developmental learner corpus work is to either (1) rely on truly longitudinal corpus data (e.g. the *Longitudinal Database of Learner English* (LONGDALE; cf. Meunier 2016) to adopt a process-oriented approach to L2 accuracy and complexity development or (2) to rely on a learner corpus which enables the simultaneous selection of proficiency level, mother tongue background and other additional metavariables that come into play to shape L2 developmental trajectories.

References

Abdel Latif, Muhammad M. Mahmoud. 2013. What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics* 34(1): 99-105. https://doi.org/10.1093/applin/ams073

Alexopoulou, Theodora, Michel, Marije, Murakami, Akira & Meurers, Detmar. 2017. Task effects on linguistic complexity and accuracy. A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning* 67(1): 180–208. https://doi.org/10.1111/lang.12232

Bestgen, Yves, Granger, Sylviane & Thewissen, Jennifer. 2012. Error patterns and automatic L1 identification. In *Approaching Transfer through Text Classification: Explorations in the Detection-based Approach*, Scott Jarvis & Scott A. Crossley (eds), 127–153. Bristol: Multilingual Matters.

Biber, Douglas & Conrad, Lauren. 2009. *Register, Genre, and Style*. Cambridge: CUP.

Biber, Douglas, Gray, Bethany & Poonpon, Kornwipa. 2011. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly* 45(1): 5–35. https://doi.org/10.5054/tq.2011.244483

Brand, Christiane & Götz, Sandra. 2011. Fluency versus accuracy in advanced spoken learner language: A multi-method approach. *International Journal of Corpus Linguistics* 16(2): 255–275. https://doi.org/10.1075/ijcl.16.2.05bra

Bulté, Bram & Housen, Alex. 2014. Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing* 26(4): 42–65. https://doi.org/10.1016/j.jslw.2014.09.005

Bulté, Bram & Housen, Alex. 2012. Defining and operationalizing L2 complexity. In *Dimensions of L2 Performance and Proficiency* [Language Learning and Teaching 32], Alex Housen, Folkert Kuiken & Ineke Vedder (eds), 21–46. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.32.02bul

Chan, HuiPing, Verspoor, Marjolijn & Vahtrick, Louisa. 2015. Dynamic development in speaking vs. writing in identical twins. *Language Learning* 65(2): 298–325. https://doi.org/10.1111/lang.12107

Chen, Xiaobin & Meurers, Detmar. 2016. CTAP: A web-based tool supporting automatic complexity analysis. In *Proceedings of The Workshop on*

Computational Linguistics for Linguistic Complexity, Dominique Brunato, Felice Dell'Orletta, Giulia Venturi, Thomas François & Philippe Blache (eds), 113-119. Osaka: The COLING 2016 Organizing Committee.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Council of Europe.

Crossley, Scott & McNamara, Danielle. 2014. Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing* 26: 66–79. https://doi.org/10.1016/j.jslw.2014.09.006

Dagneaux, Estelle, Denness, Sharon & Granger Sylviane. 1998. Computer-aided error analysis. *System: An International Journal of Educational Technology and Applied Linguistics* 26(2): 163–174. https://doi.org/10.1016/S0346-251X(98)00001-3

Dagneaux, Estelle, Denness, Sharon, Granger, Sylviane, Meunier, Fanny, Neff, JoAnne, & Thewissen, Jennifer. 2008. Error Tagging Manual Version 1.3. Centre for English Corpus Linguistics. Université catholique de Louvain, Louvain-la-Neuve. Unpublished manual.

de Bot, Kees & Larsen-Freeman, Diane. 2011. Researching second language development from a dynamic systems theory perspective. In *A Dynamic*

Approach to Second Language Development: Methods and Techniques [Language Learning and Language Teaching 29], Marjolijn Verspoor, Kees de Bot & Wander Lowie (eds), 5–24. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.29.01deb

De Clercq, Bastien & Housen, Alex. 2019. The development of morphological complexity: A cross-linguistic study of L2 French and English. *Second Language Research* 35(1): 71–98. https://doi.org/10.1177/0267658316674506

de Graaff, Rick. 2019. You've got to crack a few eggs to make an omelette. Linguistic complexity and instruction in SLA: What's in it for SLA? *Instructed Second Language Acquisition* 3(2): 249–258. https://doi.org/10.1558/isla.39398

de Leeuw, Jan & Mair, Patrick. 2009. Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software* 31(3): 1–30. https://doi.org/10.18637/jss.v031.i03

Díez-Bedmar, María Belén. 2021. Error analysis. In *The Routledge Handbook of Second Language Acquisition and Corpora*, Nicole Tracy-Ventura & Magali Paquot (eds), 90–104. New York NY: Routledge.

Ellis, Rod & Yuan, Fangyuan. 2004. The effects of planning on fluency, complexity and accuracy in second language narrative writing. *Studies* *in Second Language Acquisition* 26(1): 59–84. https://doi.org/10.1017/S0272263104026130

Fogel, Alan. 2011. Theoretical and applied dynamic systems research in developmental science. *Child Developmental Perspectives* 5: 267–272. https://doi.org/10.1111/j.1750-8606.2011.00174.x

Foster, Pauline & Skehan, Peter. 1996. The influence of planning and task type on second language performance. *Studies in Second Language Acquisition* 18(3): 299–324. https://doi.org/10.1017/S0272263100015047

Garside, Roger & Smith, Nicholas. 1997. A hybrid grammatical tagger: CLAWS4. In *Corpus Annotation: Linguistic Information from Computer Text Corpora*, Roger Garside, Geoffrey Leech & Anthony McEnery (eds), 102–121. London: Longman.

Götz, Sandra. 2013. *Fluency in Native and Nonnative English Speech* [Studies in Corpus Linguistics 53]. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.53

Granger, Sylviane, Dupont, Maïté, Meunier, Fanny, Naets, Hubert & Paquot, Magali. 2020. *The International Corpus of Learner English*, Version 3. Louvain-la-Neuve: Presses universitaires de Louvain.

Hasko, Victoria. 2013. Capturing the dynamics of second language development via learner corpus research: A very long engagement. *The*

Modern Language Journal 97(1): 1–10. https://doi.org/10.2307/23361908

Housen, Alex, De Clercq, Bastien, Kuiken, Folkert & Vedder, Ineke. 2019. Multiple approaches to complexity in second language research. *Second Language Research* 35(1): 3–21. https://doi.org/10.1177/0267658318809765

Housen, Alex, Kuiken Folkert & Vedder, Ineke. 2012. Complexity, accuracy, fluency. *International Journal of Applied Linguistics* 23(2): 1–20. https://doi.org/10.1111/ijal.12031

Housen, Alex & Kuiken, Folkert. 2009. Complexity, accuracy, and fluency in Second Language Acquisition. *Applied Linguistics* 30(4): 461–473. https://doi.org/10.1093/applin/amp048

Jin, Tan, Lu, Xiaofei & Ni, Jing. 2020. Syntactic complexity in adapted teaching materials: Differences among grade levels and implications for benchmarking. *The Modern Language Journal* 104(1): 192–208. https://doi.org/10.1111/modl.12622

Kuiken, Folkert, Housen, Alex, Vedder, Ineke & De Clercq, Bastien. 2019. Variation in syntactic complexity: Introduction. *International Journal of Applied Linguistics* 29(2): 161–170. https://doi.org/10.1111/ijal.12255

Kuiken, Folkert & Vedder, Ineke. 2019. Investigating teachers' perceptions of syntactic complexity in L2 academic writing. *Instructed Second*

Language Acquisition 3(2): 228–248. https://doi.org/10.1558/isla.37977

Kyle, Kristopher. 2016. Measuring Syntactic Development in L2 Writing: Fine Grained Indices of Syntactic Complexity and Usage-based Indices of Syntactic Sophistication. PhD Dissertation, Georgia State University.

Kyle, Kristopher & Crossley, Scott. 2018. Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal* 102(2): 333–349. https://doi.org/10.1111/modl.12468

Lambert, Craig & Kormos, Judit. 2014. Complexity, accuracy and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics* 35(5): 607–614. https://doi.org/10.1093/applin/amu047

Larsen-Freeman, Diane. 2018. Second Language Acquisition, WE, and language as a complex adaptive system (CAS). *World Englishes* 37(1): 80–92. https://doi.org/10.1111/weng.12304

Larsen-Freeman, Diane. 2009. Adjusting expectations: The study of complexity, accuracy and fluency in Second Language Acquisition. *Applied Linguistics* 30(4): 579–589. https://doi.org/10.1093/applin/amp043

Larsen-Freeman, Diane. 2006. The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of

English. *Applied Linguistics* 27(4): 590–619. https://doi.org/10.1093/applin/aml029

Lennon, Paul. 1991. Error: Some problems of definition, identification, and distinction. *Applied Linguistics* 12(2): 180–196. https://doi.org/10.1093/applin/12.2.180

Loewen, Shawn & Hui, Bronson. 2021. Small samples in instructed second language acquisition research. *The Modern Language Journal* 105(1): 187–193. https://doi.org/10.1111/modl.12700

Lowie, Wander, Michel, Marije, Keijzer, Merel & Steinkrauss, Rasmus. 2020. *Usage-Based Dynamics in Second Language Development*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781788925259

Lowie, Wander & Verspoor, Marjolijn. 2007. A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition* 10(1): 7–21. https://doi.org/10.1017/S1366728906002732

Lu, Xiaofei. 2017. Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing* 34(4): 493–511. https://doi.org/10.1177/0265532217710675

Lu, Xiaofei. 2012. The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal* 96(2): 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

Lu, Xiaofei. 2011. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly* 45(1): 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, Xiaofei. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15(4): 36–62. https://doi.org/10.1075/ijcl.15.4.02lu

Mair, Patrick, Groenen, Patrick J.F. & de Leeuw, Jan. 2021. More on multidimensional scaling in R: Smacof version 2. *Journal of Statistical Software*. <https://cran.r-project.org/web/packages/smacof/vignettes/smacof.pdf> (12 February 2022)

McNamara, Danielle, Graesser, Arthur, McCarthy, Philip & Cai, Zhiqiang. 2014. *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511894664

Meunier, Fanny. 2016. Introduction to the LONGDALE project. In *Studies in Learner Corpus Linguistics: Research and Applications for Foreign Language Teaching and Assessment*, Erik Castello, Katherine Ackerley & Francesca Coccetta (eds), 123–126. Berlin: Peter Lang.

Michel, Marije, Murakami Akira, Alexopoulou, Theodora & Meurers, Detmar. 2019. Effects of task type on morphosyntactic complexity: Evidence from a large learner corpus of A1 to C2 writings. *Instructed*

Second Language Acquisition 3(2): 124–152. https://doi.org/10.1558/isla.38248

Norris, John & Ortega, Lourdes. 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics* 30(4): 555–578. https://doi.org/10.1093/applin/amp044

Ortega, Lourdes. 2015. Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing* 29: 82–94. https://doi.org/10.1016/j.jslw.2015.06.008

Pallotti, Gabriele. 2015. A simple view of linguistic complexity. *Second Language Research* 31(1): 117–134. https://doi.org/10.1177/0267658314536435

Pallotti, Gabriele. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics* 30(4): 590–601. https://doi.org/10.1093/applin/amp045

Paquot, Magali. 2019. The phraseological dimension in interlanguage complexity research. *Second Language Research* 35(1): 121–145. https://doi.org/10.1177/0267658317694221

Plakans, Lia, Gebril, Ata & Bilki, Zeynep. 2019. Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing* 36(2): 161–179. https://doi.org/10.1177/0265532216669537

Polio, Charlene & Shea, Mark. 2014. An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing* 26(1): 10–27. https://doi.org/10.1016/j.jslw.2014.09.003

Polio, Charlene & Yoon, Hyung-Jo. 2018. The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics* 28(1): 165–188. https://doi.org/10.1111/ijal.12200

Rosmawati, Rosmawati. 2014. Dynamic development of complexity and accuracy: A case study in second language academic writing. *Australian Review of Applied Linguistics* 37(2): 75–100. https://doi.org/10.1075/aral.37.2.01ros

Skehan, Peter. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30(4): 510–532.

Skehan, Peter & Foster, Pauline. 2012. Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research In *Dimensions of L2 performance and Proficiency* [Language Learning and Teaching 32], Alex Housen, Folkert Kuiken & Ineke Vedder (eds), 199–220. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.32.09fos

Spoelman, Marianne & Verspoor, Marjolijn. 2010. Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics* 31(4): 532–553. https://doi.org/10.1093/applin/amq001

Thelen, Esther & Smith, Linda, B. 1994. *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge MA: The MIT Press.

Thewissen, Jennifer. 2021. Accuracy. In *The Routledge Handbook of Second Language Acquisition and Corpora*, Nicole Tracy-Ventura & Magali Paquot (eds), 305–317. New York NY: Routledge.

Thewissen, Jennifer. 2015. *Accuracy across Proficiency Levels: A Learner Corpus Approach.* Louvain-la-Neuve: Presses Universitaires de Louvain.

Thewissen, Jennifer & Anishchanka, Alena. 2019. An integrative approach to L2 accuracy and complexity development. Paper presented at the 5th Learner Corpus Research Conference. 12-14 September 2019. University of Warsaw, Poland.

Verspoor, Marjolijn, Schmid, Monika & Xu, Xiaoyan. 2012. A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing* 21: 239–263. https://doi.org/10.1016/j.jslw.2012.03.007

Van Waes, Luuk & Leijten, Mariëlle. 2015. Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2.

Computers and Composition 38: 79–95. https://doi.org/10.1016/j.compcom.2015.09.012

Vyatkina, Nina & Housen, Alex. 2021. Complexity. In *The Routledge Handbook of Second Language Acquisition and Corpora*, Nicole Tracy-Ventura & Magali Paquot (eds), 318–330. New York NY: Routledge.

Yang, Weiwei, Lu, Xiaofei, Cushing Weigle, Sarah. 2015. Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgements of writing quality. *Journal of Second Language Writing* 28: 53–67. https://doi.org/10.1016/j.jslw.2015.02.002

Yu, Hanjing & Lowie, Wander. 2020. Dynamic paths of complexity and accuracy in second language speech: A longitudinal case study of Chinese learners. *Applied Linguistics* 41(6): 855–877. https://doi.org/10.1093/applin/amz040

Appendix 1

Table 1. Descriptive statistics for measures of grammatical accuracy and syntactic complexity

	All texts		B1 (N=65)		B2 (N=62)		C1 (N=67)		C2 (N=28)	
Measure	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
GA	2.99	3.00	5.62	3.81	2.55	1.89	1.72	1.37	0.86	0.93
GADJCS	0.28	0.84	0.64	1.27	0.21	0.67	0.09	0.44	0.06	0.30
GADJN	0.31	1.11	0.85	1.86	0.05	0.31	0.09	0.46	0.11	0.40
GADJO	0.21	0.69	0.10	0.48	0.06	0.28	0.38	0.92	0.37	0.99
GADVO	2.43	2.63	3.07	2.69	2.95	3.11	1.73	2.05	1.44	1.84
GNC	0.27	0.46	0.30	0.42	0.29	0.54	0.26	0.49	0.14	0.25
GNN	0.89	1.03	1.21	1.17	1.00	0.96	0.70	0.99	0.32	0.53
GPD_GDD*	2.17	4.95	2.93	5.13	2.50	5.46	2.03	5.14	0.00	0.00
GPP	2.32	4.93	3.10	4.65	3.03	7.24	1.46	2.90	0.96	1.39
GPO_GDO*	3.77	8.44	6.61	12.17	3.47	6.64	2.36	6.05	1.21	3.39
GPI_GDI*	2.65	5.11	4.68	6.83	1.50	2.82	2.11	4.87	1.76	3.57
GPF	3.47	14.42	4.87	16.73	5.83	19.68	1.37	7.05	0.00	0.00
GPR	6.50	11.50	9.15	12.20	7.58	13.88	4.82	8.68	1.94	7.93
GPU	3.04	4.44	5.20	5.56	3.12	4.10	1.74	3.27	0.91	1.93
GVAUX	12.73	13.66	16.32	16.01	12.13	12.54	11.38	13.36	8.95	8.76
GVM	0.24	0.58	0.52	0.86	0.13	0.35	0.09	0.27	0.16	0.46
GVN	1.00	2.28	2.44	3.62	0.58	0.96	0.27	0.83	0.34	0.77
GVNF	0.36	0.75	0.35	0.65	0.39	0.98	0.42	0.72	0.16	0.41
GVT	2.52	2.90	3.14	3.54	2.73	2.71	1.98	2.34	1.91	2.65
GVV	0.13	0.41	0.27	0.55	0.16	0.46	0.02	0.16	0.04	0.23
GWC	0.10	0.18	0.23	0.26	0.08	0.13	0.03	0.06	0.02	0.06
WM	3.95	4.51	6.95	5.86	3.18	3.47	2.56	2.81	2.02	2.67
WO	1.94	2.82	2.57	3.24	2.41	3.36	1.33	1.85	0.93	1.75
WRS	1.83	2.92	2.99	3.62	1.47	2.82	1.39	2.13	0.95	2.21
WRM	1.08	2.00	1.72	2.74	1.08	1.79	0.66	1.28	0.62	1.47
XADJPR	0.25	0.70	0.33	0.83	0.20	0.66	0.22	0.67	0.19	0.56
XNCO	0.17	0.41	0.24	0.44	0.17	0.38	0.13	0.43	0.08	0.29
XNPR	0.31	0.51	0.53	0.67	0.20	0.35	0.26	0.43	0.14	0.36
XNUC	0.73	1.06	1.27	1.35	0.69	1.03	0.38	0.64	0.40	0.65
XVCO	0.36	0.81	0.51	0.95	0.37	0.74	0.28	0.81	0.18	0.56
XVPR	1.56	2.14	2.78	2.69	1.30	2.00	1.01	1.44	0.61	0.78
ProDet*	3.96	3.19	6.45	3.45	3.88	2.73	2.67	2.04	1.45	1.64
GADJ *	0.79	1.59	1.60	2.35	0.33	0.75	0.56	1.09	0.53	1.04
WR *	2.91	3.58	4.71	4.08	2.55	3.73	2.05	2.55	1.57	2.50
W	683	96	667	101	668	85	686	92	748	93

S	31.95	9.14	28.51	9.18	33.45	8.84	33.93	8.31	31.89	9.72
VP	91.42	17.06	88.26	18.04	93.37	17.58	91.19	15.39	95.00	16.93
C	68.56	13.89	67.28	14.45	70.84	14.88	67.22	12.48	69.68	13.42
T	38.03	9.85	35.62	9.41	39.77	10.56	39.82	9.27	35.46	9.26
DC	28.49	8.66	28.49	9.22	29.21	8.80	26.18	7.32	32.39	8.75
CT	19.64	5.04	19.51	5.21	20.34	5.34	19.00	4.95	19.93	4.05
CP	15.32	6.45	15.00	6.87	14.27	5.99	15.73	5.67	17.43	7.83
CN	77.19	16.47	76.51	17.34	75.42	15.93	76.51	15.52	84.36	16.83
MLS	22.89	6.45	25.38	7.58	21.10	5.08	21.05	4.40	25.42	7.86
MLT	18.97	4.86	19.71	4.81	17.87	4.82	17.86	3.63	22.32	5.93
MLC	10.26	1.96	10.24	2.18	9.78	2.07	10.41	1.63	10.98	1.71
C_S	2.25	0.55	2.51	0.69	2.18	0.41	2.04	0.38	2.31	0.61
VP_T	2.49	0.51	2.56	0.50	2.44	0.48	2.35	0.39	2.80	0.69
C_T	1.86	0.35	1.94	0.36	1.83	0.31	1.72	0.26	2.03	0.44
DC_C	0.42	0.09	0.42	0.09	0.41	0.09	0.39	0.08	0.47	0.09
DC_T	0.80	0.32	0.85	0.34	0.78	0.28	0.69	0.24	0.98	0.42
T_S	1.21	0.15	1.28	0.19	1.19	0.12	1.18	0.14	1.13	0.11
CT_T	0.54	0.14	0.56	0.14	0.53	0.13	0.49	0.13	0.59	0.18
CP_T	0.43	0.22	0.45	0.23	0.39	0.23	0.41	0.16	0.53	0.30
CP_C	0.23	0.12	0.24	0.14	0.21	0.12	0.24	0.09	0.26	0.14
CN_T	2.16	0.69	2.25	0.62	2.02	0.67	2.03	0.65	2.54	0.83
CN_C	1.17	0.32	1.18	0.32	1.11	0.32	1.18	0.32	1.25	0.31

* A number of aggregated variables were tested to identify more general patterns for several minor error types and less common errors. Variables GPD_GDD, GPO_GDO, GPI_GDI each combine two similar error types. Variable ProDet combines all errors in pronouns and determiners (GDD, GDO, GDI, GPD, GPP, GPI, GPF, GPR, GPU). Variable GADJ combines errors in the use of adjectives (GADJCS, GADJN, GADJO). Variable WR combines errors in word order (WRS, WRM).