

COMPLEXITY AND SECOND LANGUAGE WRITING QUALITY

Andrea Nava

UNIVERSITÀ DEGLI STUDI DI MILANO

Abstract

È un dato di fatto che l'abilità di scrittura in lingua inglese svolge un ruolo fondamentale nella formazione accademica e professionale nel secondo decennio del XXI secolo. Non sorprende, quindi, che negli ultimi anni la ricerca in ambito linguistico-acquisizionale e glottodidattico abbia preso in esame i processi con cui si sviluppa l'abilità di scrittura in lingua straniera e abbia cercato di individuare i criteri per poter misurare la competenza degli apprendenti e poter determinare la 'qualità' della produzione scritta. In particolare, si è proposto di analizzare l'interlingua degli apprendenti, ivi inclusi i processi di acquisizione dell'abilità della scrittura in ambito accademico e professionale, da tre prospettive interdipendenti – la correttezza, la scorrevolezza e la complessità. Di questi tre elementi costitutivi la competenza linguistica, la complessità è l'ambito che risulta a tutt'oggi meno esplorato. Questo contributo si prefigge lo scopo di presentare i più recenti sviluppi della ricerca nell'ambito della scrittura in lingua straniera dalla prospettiva della complessità linguistica. Dal momento che la maggioranza degli studi ha preso in esame apprendenti di lingua inglese in contesti accademici anglofoni, si presentano anche i risultati di uno studio di carattere esplorativo che ha coinvolto un campione ristretto di apprendenti italo-foni di lingua inglese nell'università italiana.

Given the key role played by writing skills in higher education and in professional contexts, in both one's native language and, in a globalized world, increasingly in English, research purporting to determine the best ways

to assess writing proficiency/quality and to track writing development has grown exponentially in recent years. In the Second Language Acquisition (SLA) field, an influential approach analyses the quality of learners' output in a second language in terms of the degree of complexity, accuracy and fluency it exhibits. While the notions of fluency and accuracy have a long history, complexity is little explored. This article reviews the most recent developments in second language studies that have sought to relate linguistic complexity indices to criteria of writing proficiency/quality. The article further reports on the findings of an exploratory study into complexity and writing quality which has analysed a small corpus of L2 English writing by Italian undergraduates, an as yet under-researched population.

I. INTRODUCTION

Given the key role played by writing skills in higher education and in professional contexts, in both one's native language and, in a globalized world, increasingly in English, research purporting to determine the best ways to assess writing proficiency/quality and to track writing development has grown exponentially in recent years (cf. the 2014 issue of the *Journal of Second Language Writing* devoted to «Comparing perspectives on L2 writing: Multiple analyses of a common corpus»). In the Second Language Acquisition (SLA) field, an influential approach to research into the quality of learners' output in a second language has come to be known by the acronym CAF (Complexity – Accuracy – Fluency). In this approach learners' language performance (as well as language proficiency and development) is analysed in terms of the degree of complexity, accuracy and fluency it exhibits. The more general aim of studies employing measures of CAF is, in Norris and Ortega's (2009: 557) words, «to account for how and why language competencies develop for specific learners and target languages, in response to particular tasks, teaching, and other stimuli, and mapped against the details of developmental rate, route, and ultimate outcomes».

While the notions of fluency and accuracy have a long history, particularly as methodological options associated with the communicative language teaching approach (Howatt 2004), complexity was singled out as a separate element of L2 proficiency more recently (Housen and Kuiken 2009). Research undertaken in the last two decades seems to point to the fact that each CAF element is multidimensional and cannot be fully understood in isolation, unless, that is, it is viewed as part of a «dynamic system» whose elements «interact in often unpredictable ways» (Norris and Ortega: 556), so much so that in language use and development «several sub-dimensions of CAF may compete» (Bulté and Housen 2012: 33).

In this article, the concept of linguistic complexity is examined as part of the CAF triad for the insights it can provide to researchers and practitioners on the elusive notion of second language writing quality. The article is organized as follows. I will first touch on some key issues in the definition and measurement of complexity in second language research. I will then review in more detail a number of studies that have sought to relate linguistic complexity indices to criteria of writing proficiency/quality. Finally, I will report on the findings of an exploratory study into complexity and writing quality which has analysed a small corpus of L2 English writing by Italian undergraduates, an as yet under-researched population.

2. DEFINING AND MEASURING COMPLEXITY IN SLA RESEARCH

What qualifies as ‘complex’ vis-à-vis ‘accurate’ and ‘fluent’ language production? While definitions of accuracy and fluency can be provided rather uncontroversially (Housen et al. 2012), complexity has been a more elusive concept (Housen and Kuiken 2009; Norris and Ortega 2009; Housen et al. 2012; Bulté and Housen 2012; Pallotti 2014). Reviewers of «empirical CAF research» have lamented the fact that it «has taken a rather narrow, reductionist, perhaps even simplistic view on and approach to what constitutes L2 complexity» (Bulté and Housen 2012: 34).

To rectify this situation, within the SLA field, recent proposals have stressed the need to tell apart complexity as a ‘cognitive/relative’ vs. an ‘absolute’ concept (Kusters 2003; Miestamo 2008). What is ‘cognitively’ complex to learn or to perform is mainly a ‘relative’ (Pallotti 2014) notion as it is often dependent on more subjective aspects of the language learner/user (motivation, language aptitude etc.). On this view, cognitive complexity or ‘difficulty’ refers to «the mental ease or difficulty with which linguistic items are learned, processed or verbalized in the processes of language acquisition and use» (Bulté and Housen 2012: 23). ‘Absolute’ complexity, on the other hand, rests on more objective, quantitatively determined traits, such as those that make up ‘linguistic’ complexity, which may be operationalised as «phonemic, lexical, morphological or syntactic items, structures or rules manifested in a language sample or in the language user’s linguistic repertoire» (Bulté and Housen 2019: 160)¹. Linguistic complexity can be a property of the whole interlanguage system (‘global/system linguistic complexity’) or of given features of the system (‘local/structural linguistic complexity’), across its different levels and domains (e.g. phonology, morphology, syntax, lexis). The latter type of linguistic complexity can be determined on the

1 In addition to the ‘absolute’ and ‘relative’ concepts, Pallotti (2014: 118) identifies in current research a third way of conceiving complexity, encapsulated in the term ‘developmental complexity’, i.e. «the order in which linguistic structures emerge and are mastered in second (and, possibly, first) language acquisition».

basis of ‘formal’ and/or ‘functional’ criteria (Kusters 2003). For example, the English present perfect may be thought of as both functionally (e.g. it involves a three-way relation between speech time, reference time and situation time; there is no one-to-one relation between form and meaning) and formally (e.g. it is made up of several parts, each bearing semantic content, and is built up through different – both analytic and synthetic – strategies) complex (Davydova 2011).

System and structure complexity have also been associated with the notions of ‘breadth’ (or ‘range’) and ‘depth’ (or ‘sophistication’) (Lu 2011; Lan et al. 2019). From a syntactic point of view, for example, a learner’s output may be deemed complex if it features a wide range of grammatical constructions and/or it shows evidence of mastery of more sophisticated grammatical constructions. As Lan et al. (2019: 2) claim, this notion of complexity as variation and sophistication has now become «accepted widely» in the SLA field. However, pinning down what exactly is a ‘sophisticated’ linguistic unit is not necessarily straightforward (*ibidem*). In the realm of syntax, for instance, ‘sophistication’ has often been taken as a synonym of elaboration (e.g. the use of longer, more elaborate phrases) (Kuiken et al. 2019). It has also been pointed out that different linguistic registers may exhibit complexity in different ways: for example, higher syntactic complexity may be achieved through clausal subordination in spoken discourse, but it tends to be signalled through nominal modification in written academic registers (Biber et al. 2011; Biber and Gray 2016). Other factors that have been singled out as potentially impacting the way complexity is manifested are the topic (Yang et al. 2015), genre (e.g. narrative vs. argumentative writing, Lu 2011) and the L1 (Lu and Ai 2015).

The L2 research studies carried out to date featuring complexity as either a dependent (e.g. the effect of various instructional options on the degree of complexity of learners’ output is investigated) or an independent or primary variable have relied on a range of measures of linguistic complexity (Wolfe-Quintero et al. 1998; Ortega 2003; Bulté and Housen 2012), targeting in particular syntactic complexity, which is acknowledged to be the «most frequently and intensively measured component of linguistic complexity in SLA research» (Kuiken et al. 2019: 162).

The sheer number of complexity measures devised by researchers may be viewed as a natural consequence of the fact that complexity is a ‘complex’ notion (*sic*) and no single measure can fit all L2 production contexts. For example, in studies of L2 writing, complexity has been operationalized according to four parameters: length (e. g. mean length of sentence, clause or T-unit²), ratio (e.g. clauses per T-unit), index (e.g. syntactic variety) and frequency (e.g. number of noun phrases) (Lan et al. 2019), with «mean length

2 A T-unit (terminal unit) is defined a structure with a main clause and the subordinate clauses attached to it (Hunt 1965).

of T-unit» being «the single most employed complexity measure» (Norris and Ortega 2009: 566). By contrast, in research on task-based L2 production, measures that equal complexity with subordination (e.g. dependent clauses per T-unit) are most commonly used.

The wide range of measures proposed and the preferences expressed by researchers for one or another measure to be used in particular contexts, however, do not seem to have necessarily resulted from an appreciation of what each measure is supposed to specifically tap. It has been pointed out, for example, that some complexity measures are general in scope (e.g. «any length-based measure with a potentially multi-clausal denominator», e.g. mean length of T-unit, Norris and Ortega 2009: 566) while others have more specific remits, targeting e.g. sentential, clausal or phrasal complexity. For instance, subordination measures (e.g. dependent clauses per T-unit) are meant to gauge sentential complexity and mean length of clause is thought to be a measure of phrasal complexity (although the latter association has been disputed, cf. Bulté and Housen 2012). It has also been suggested that emergence of syntactic complexity is to be signalled in different ways at different levels of a learner's proficiency, following three main stages (Norris and Ortega 2009). At lower levels, syntactic growth is signalled by the learners' use of coordination, at the intermediate/upper intermediate level, by an increase in sentential subordination while at more advanced levels it is indexed by phrasal elaboration – learners' use of more and more sophisticated phrases³. Insisting on using a single complexity index (e.g. a subordination measure) may lead researchers to misinterpret, for example, «a decrease in subordination at the highest levels of proficiency», thus failing to account for «an increase in the overall complexity of the language performance» (Norris and Ortega: 566). Moreover, as mentioned above, register should also be taken into account when selecting syntactic complexity measures, as the pervasively used «T-unit measures are much more strongly associated with conversational complexities than the complexities of writing, while a new set of grammatical measures is required to account for the actual complexities of formal written discourse» (Biber et al. 2011: 17). In particular, researchers are highlighting the need to consider measures targeting levels of syntactic organization that have traditionally been given short shrift, viz. the clausal and phrasal levels (De Clercq and Housen 2017; Bulté and Housen 2018; Kuiken et al. 2019).

Another issue is related to the 'hybrid' nature of some measures of complexity. The reason why subordination ratios are such popular ways of operationalizing complexity may also lie in the fact that subordination can be associated with both the linguistic/absolute and the cognitive notions

³ The straightforward association between development of complexification according to levels of syntactic organization and proficiency seems to have been called into question by more recent studies (Bulté and Housen 2014; Ortega 2015).

of complexity (Bulté and Housen 2012). Subordinate clauses are arguably held to be complex not only because they are made up of more parts than independent clauses but also because they are more (cognitively) difficult to process and produce (Bygate 1999).

There is no gainsaying that the measurement of complexity (and of CAF more generally) is still a partially unresolved issue in SLA. Critical surveys of the field have identified the «analytic challenges» presented by different metrics as well as their «reliability, validity, and sensitivity» (Housen and Kuiken 2009: 464), stressed that measures need to be chosen for the specific facets of complexity they tap (Norris and Ortega 2009) and include «more developmentally sensitive complexity measures targeting different aspects of complexity» (Kuiken and Vedder 2019: 193), and assessed the potential of the increasingly widely available automated tools for computing complexity (Polio and Yoon 2018).

3. COMPLEXITY AND SECOND LANGUAGE WRITING QUALITY

As was mentioned above, several studies of L2 production have featured complexity as a dependent or independent variable. Since the 1990s task-based language learning researchers have attempted to identify which task design and implementation options lead to more complex output (e.g. Skehan 1998; Ellis 2003; Skehan and Foster 2012). Underlying this research is a psycholinguistic approach (Ellis and Barkhuizen 2005) that posits that in language performance learners may prioritize one or more of the three CAF components depending on specific conditions of L2 learning and use (Housen et al. 2012), leading, according to some researchers, to trade-offs (Skehan 2009).

Another well-developed strand of complexity research has focused on L2 writing and sought to investigate learners' writing proficiency/quality and/or development over time (e.g. Larsen-Freeman 2006), or under the influence of given instructional options (Kuiken and Vedder 2011), as mirrored in interlanguage complexity. The underlying premise of this research is that higher L2 written proficiency is signalled by higher degrees of linguistic complexity (in other words, complexity increases linearly as proficiency develops) – with complexity being routinely operationalised as the use of a wide range of and/or more sophisticated linguistic features.

Two recent studies (Bulté and Housen 2014 and Crossley and Macnamara 2014) have addressed the issue of which complexity measures can best encapsulate the L2 development which intermediate/upper intermediate level ESL students receiving writing instruction at an American university achieved over a semester, and whether complexity indices may be correlated with human ratings of writing quality, expressed both holistically and

according to five analytical categories (including assessment of vocabulary and language use). Complexity was operationalized in slightly different ways in the two studies. Bulté and Housen (2014) relied on syntactic complexity measures mainly targeting the sentence level (e.g. mean length of sentence) and also focused on lexical complexity. Crossley and Macnamara (2014) operationalized syntactic complexity through a wider range of measures targeting the phrasal and clausal (as well as the sentential) levels, which they computed using an automatic tool (Coh Metrix⁴).

The findings from these two studies provide important insights into the construct of linguistic complexity as an index of L2 writing proficiency and development. First, it emerged that not all aspects of complexity had developed in the same way over the one-semester course. While results did not point to statistically significant changes in measures of lexical complexity, syntactic complexity seemed to have undergone development. This suggests that at least *some* measures of complexity are able to account for L2 writing development. Second, results lent support to Norris and Ortega's (2009) cautionary attitude to the 'one-size-fits-all' approach that has made subordination ratios among the most commonly used measures of complexity. Indeed, the development which the intermediate/upper intermediate level students underwent over the one-semester course was not indexed by significant changes in their use of subordination but by growth in phrasal complexity (e.g. longer noun phrases and more words before the main verb). However, rather unexpectedly, the students' progress was also significantly correlated (in Bulté and Housen's study) with changes in their use of coordination, a phenomenon usually associated with lower level language users. The authors interpreted these results pointing to the nonlinear development of complexity through the lens of Complex Systems theory (e.g. Larsen-Freeman and Cameron 2008), whereby language development is thought to be «characterized by periods of growth and progress alternating with periods of stabilization or even temporary backsliding before progress picks up again (if at all)» (Bulté and Housen 2014: 54).

With regard to the correlation between complexity indices and human ratings of writing quality, both Bulté and Housen (2014) and Crossley and Macnamara (2014) found that most of the measures that had accounted for syntactic development (i.e. those targeting the phrasal level) were not predictive of human assessments. Other components of linguistic complexity – associated with clausal subordination (e.g. subclause ratio) and clause and sentence length – appeared to be correlated with the assessors' judgements. In other words, while learners were increasingly complexifying their output towards features more typical of written academic discourse (e.g.

4 Coh Metrix is a computational tool aimed at measuring text difficulty which includes a range of measures of text cohesion, some of which are also related to linguistic complexity (Polio and Yoon 2018).

nominalization, phrasal elaboration), assessors seemed to lay greater store by the use of syntactic features which have been shown to mark complexity in spoken discourse. This may have been a result of the fact that the writing prompts the students were given were meant to lead them to produce descriptive essays, a genre which the assessors tended to associate with less formal, spoken discourse as opposed to argumentative writing. As Crossley and Macnamara (2014: 75) point out in the discussion of their study's findings, what clearly emerged from the analysis was that «the syntactic features that develop in L2 learners are not the same syntactic features that will assist them in receiving higher evaluations for essay quality».

Mazgutova and Kormos's (2015) study provides further insights into some of the issues raised by the two earlier studies described above. In particular, this study, which analysed a corpus of essays written by university students at two different ESL proficiency levels (upper intermediate and advanced) over a short but intensive academic writing course, found that while the trends towards complexification highlighted in previous studies were confirmed, it was the lower proficiency group that made the largest gains. With regard to noun phrase complexity indices, no significant increase was recorded for the higher proficiency group over the period – as the use of phrasal units and the degree of phrasal elaboration were already high for this group at the beginning of the study, development seemed to entail «the use of syntactically less complex but conceptually more abstract lexical units to express their views and opinions» (Mazgutova and Kormos 2015: 12). In other words, a learner's proficiency was shown to affect not only which aspects of syntactic complexification (e.g. phrasal vs. sentential) undergo development but also whether complexification favours one linguistic domain (e.g. lexicon over syntax) over another.

Another important factor impacting complexity in its multidimensionality and its relation to writing quality is explored in Kuiken and Vedder's (2019) study – the degree of typological distance between L1 and L2 and between different L2s. The researchers' use of finer-grained complexity measures (e.g. coordination within T-units, between T-units and between constituents) enabled them to detect «patterns and variation in the process of gradual complexification of L2 production across proficiency levels, across languages, and between L2 and L1» (Kuiken and Vedder 2019: 195). The study involved university students of L2 Dutch, Italian and Spanish in a Dutch university whose proficiency ranged from CEFR A2 to B1 levels as well as a sample of L1 Dutch, Italian and Spanish speakers. Each informant produced two argumentative writing samples. A range of complexity measures were used to analyse the essays, targeting both overall and more specific (subordination, coordination, phrasal elaboration) aspects of syntactic complexity. The results yielded by the analysis provided some evidence (though statistical significance was obtained by only one of the language groups investigated) in support of the hypothesis

that syntactic complexification is achieved in different ways at different proficiency levels. More interesting, it emerged that the differences in complexification across levels and between L1 and L2 were signalled by the more specific measures (e.g. types of subordinate clauses) used in the study – for example, it was found that higher level Italian L2 learners used «more coordination within T-units, relative clauses, and longer postmodifying phrases than their less proficient peers» (Kuiken and Vedder 2019: 206) and that L1 and L2 Spanish essays differed in terms of «clauses per T-unit, coordination within T-units, and relative clauses» (201).

The study also pointed to the fact that complexification patterns may vary across languages. For example, the (albeit moderate) correlation found between proficiency and complexity for L2 Italian was not matched by similar patterns for L2 Spanish and Dutch. This suggests that language-specific ways of attaining complexification may exist and points to the possible influence of the learners' L1s on L2 complexification patterns (in the case of L2 Italian and Spanish, all informants were native speakers of Dutch whereas the L2 Dutch informants had a wide range of different L1s). While the issue of L1 complexification was not among the research questions targeted by the study, the analysis of the L1 essays found that L1 Italian writers displayed higher degrees of overall complexity (higher number of clauses per T unit) and more elaborate postmodification compared to the L1 Dutch and Spanish writers.

Although still in its infancy, the investigation of the relation between complexity and L2 writing proficiency/quality and development has made great strides since ever more sophisticated computational tools for computing different measures became available. The studies that have been reviewed in this section have on the one hand shown that mismatches often obtain between the results of analyses of learners' writing from the perspective of complexity and human ratings of quality. On the other hand, they have confirmed the often unpredictable effects yielded by factors such as «L1–L2 (L3/L4 etc.) language combinations, developmental sequences and task» (Bernardini and Granfeldt 2019: 226-227). It is thus important to widen the range of learner populations targeted in L2 complexity research, which the exploratory study illustrated in the next section has sought to do by investigating L2 English writing by Italian undergraduates.

4. EXPLORING COMPLEXITY AND WRITING QUALITY IN ITALIAN UNDERGRADUATES' L2 ENGLISH

Written exams have been a distinctive feature of degrees in Foreign Languages and Literatures in Italy ever since they were set up within universities as part of the Faculties of Education (*Magistero*) in the first few

decades of the last century (Nava 2018). This is in contrast to the oral-based assessment system still in place in most other degree courses in the humanities and social sciences. Across the decades, as language teaching methods have evolved along with the set of skills required of modern languages graduates, the focus of written exams has shifted from translation of literary texts from and into the foreign language to more open-ended tasks, such as essay writing.

In order to shed some light on the quality of Italian university students' essay writing from the perspective of complexity research, a small-scale study has been carried out based on a restricted corpus of second-year BA Foreign Languages and Literatures exam papers produced by students at the University of Milan. According to the curriculum of the second-year practical English language course (*esercitazioni*), students are required to reach the CEFR B2+ proficiency level. As part of their assessment, they sit an end-of-year written exam which includes a 250-300 word argumentative essay. The essay prompts are drawn from the range of current affairs issues dealt with during the course and in the textbook adopted by the practical language instructors (an internationally published coursebook for upper intermediate/advanced English students aimed at practising academic reading and writing skills). Students are given 60 minutes to complete their essays and no dictionaries or other reference materials are allowed. As attendance to the practical language classes is not mandatory – as is the norm in the Italian university system – students may take the end-of-year exam without submitting any previous piece of writing to the instructors. However, an increasing number of students take advantage of the option of continuous assessment in lieu of the final exam, which is conditional upon regular attendance. As a result of this, the number of students who have actually sat the final-year written exams over the past five years has decreased exponentially.

The findings that will be illustrated henceforth are yielded by the analysis of a restricted corpus of 16 essays produced by L1 Italian students during three exam sessions between 2019 and 2020. The students were required to write 250-300 word essays on three different topics: in the first exam session, the issue the students were asked to discuss was mandatory vaccination, in the second session, the impact of technology on identity and in the third session the effects of globalization on local cultures and identities⁵. Due to COVID 19 restrictions, the students sat the exams using an online videoconferencing platform. At the start of the exam, they were given the essay prompts, they were reminded of the word and time limits and that the use of dictionaries or other reference materials was not allowed and they

5 As pointed out by an anonymous reviewer, the topic of the May 2020 exam session may have led to a more emotional response by the students (*vis-à-vis* the topics chosen for the following exam sessions) given the health emergency. This may have had an impact on the quality of the output produced.

were instructed to type their essays into Word documents. The essays were marked using a holistic rating (a mark out of 30 was allocated to each essay) by the two (highly experienced) instructors who had taught the practical language course. As per the usual assessment protocol in Italian universities, each essay was marked by one examiner only (no double marking was carried out). The two examiners had, however, been working together as instructors/examiners for over 20 years, during which time they had developed and relied on common assessment criteria. The students were contacted by email by the researcher a few days after the exam results had become available and were asked whether they agreed for their exam papers to be used for research purposes. All the students gave their consent to have their papers used.

The analysis of the corpus aimed at identifying what kind of syntactic complexification was displayed in the essays written by Italian undergraduates and how complexity indices related to examiners' holistic evaluations. In order to facilitate the comparison between complexity measures and examiners' evaluations, the essays were divided into 3 groups: the first group (7 essays) was made up of the essays that had received higher holistic ratings (30-25), the second group (7 essays) included the essays that had received lower ratings but were deemed of sufficient quality to pass the exam (18-24), and the third group (2 essays) featured essays that had received fail marks (less than 18). For each essay group, a roughly equal number of essays had been marked by each of the two examiners.

Complexification was operationalized from four syntactic perspectives, following e.g. Norris and Ortega (2009): global, coordination, subordination and phrasal elaboration. It was also decided to account for complexification as breadth or range of syntactic constructions used. The 11 different complexity indices considered (Table 1) were calculated manually as well as through computerized tools (SCA⁶, Coh Metrix).

6 The Syntactic Complexity Analyzer (SCA) is a computerised tool developed by X. Lu (Lu 2010). It computes 14 measures of syntactic complexity divided into five categories (length of production units, sentence complexity, subordination, coordination, and particular structures).

TYPE OF COMPLEXIFICATION	MEASURES	MEASUREMENT TOOL
Global	Mean Length of Sentence (MLS)	SCA
	Mean Length of T-Unit (MLTU)	SCA
	Mean Length of Clause (MLC) ⁷	SCA
Coordination	Coordinate Clause Ratio: Number of Coordinate Clauses/Number of Sentences (CCR)	Manual
Subordination	Dependent Clause Ratio (DC/C)	SCA
Phrasal elaboration	Incidence of NPs/VPs/PPs ⁸	Coh Metrix
	Number of Words before Main Verb	Coh Metrix
	Number of Modifiers per NP	Coh Metrix
Breadth	Syntactic similarity between adjacent Sentences	Coh Metrix

Table 1. Complexity measures used in the analysis.

In addition to the 11 complexity measures, number of words, T-units, sentences (full stops determine sentence boundaries), clauses and dependent clauses for each essay were calculated by means of the SCA. The mean values of all these quantitative indices for each of the three essay groups are shown in Table 2. For Groups 1 and 2, each consisting of 7 essays, extremes were disregarded for each value so as not to skew the means.

⁷ I follow Bulté and Housen (2012) in considering MLC a measure of global complexity.

⁸ Incidence values are normalized – they show the number of NPs/VPs/PPs for a span of 1000 words.

	GROUP 1 (30-25)	GROUP 2 (24-18)	GROUP 3 (<18)
Words	311.2	273	241.5
MLS	21.318	25.895	32.447
MLTU	19.281	22.219	25.556
T-units	15.8	13.4	9.5
MLC	10.258	11.45	11.02
CCR	0.118	0.324	0.607
Coordinate clauses	1.8	3.4	4.5
Sentences	15	11.2	7.5
DC/C	0.456	0.428	0.563
Clauses	32	26	22
Dependent clauses	14.4	10.8	12.5
NPs	356.591	352.405	336.561
VPs	244.616	226.527	180.238
PPs	108.014	119.521	122.332
Number of words before main verb	3.991	4.266	5.116
Number of words per NP	0.762	0.783	0.887
Syntactic similarity	0.077	0.075	0.044

Table 2. Quantitative findings.

Although the study is meant to be exploratory and, given the limited size of the corpus, no statistical analysis has been carried out, some tentative insights can be drawn from the quantitative findings concerning complexity indices for the essays grouped according to the examiners' holistic evaluations.

Compared to the other essays in the corpus, the more highly rated essays (Group 1) appear to be longer and feature a higher number of both clauses and sentences. On the other hand, they seem to display a lower level of global complexity – their sentences, T-units and clauses are indeed shorter than those featured in the other two groups of essays. Subordination is preferred to coordination – nearly every other clause is a subordinate clause. As regards phrasal elaboration, Group 1 essays feature the highest number of both NPs and VPs while fewer PPs than both of the other groups. The number of words used before the main verb and the number of modifiers per NP are, however, surprisingly low. A rather limited range of syntactic constructions are employed, as shown by a relatively high mean value of the syntactic similarity metric.

It seems that the less favourably an essay is evaluated, the shorter it is, the lower the number of production units (sentences, T-units, clauses) used and, a result, the longer each production unit is, the more use is made of coordination, the lower the number of NPs and VPs (and the higher the number of PPs), the more elaborate NPs are and the wider the range of different constructions used. The only indices that run counter to these trends are those associated with the use of dependent clauses in the two essays that failed the exam. These display not only a high use of coordination, but also of subordination, as shown by the DC/C metric (which is higher than the one for the Group 1 essays) and the value for the number of Dependent Clauses used.

Taken as a whole, the findings seem to hint at the fact that second-year Italian BA students in Foreign Languages and Literatures produce argumentative essays in English which display features of syntactic complexification that have been deemed typical of an intermediate/upper intermediate level of proficiency. As was illustrated above, according to Norris and Ortega (2009), a learner's interlanguage follows a developmental pattern in the process of complexification, with lower level learners tending to complexify their interlanguage using coordination, intermediate/upper intermediate students relying on subordination, and advanced level students resorting to phrasal elaboration. The essays in the corpus also seem to point to the fact that students are starting to become aware of the distinctive features of written English. Research carried out by Biber and colleagues, which I have touched on above, has indeed shown that proficient writers tend to produce a higher number of NPs than VPs, which is of course evidence of the fact that more formal writing adopts a nominal – or synoptic – style as opposed to a verb-based – or dynamic – style typical of informal spoken discourse.

Is a higher degree of syntactic complexity in Italian undergraduates' L2 English writing rewarded by examiners? If we look at the indices of global complexity, which are length-based measures, it would actually appear that the less 'globally' complex an essay is the better evaluation it is allocated. In other words, examiners seem to rate more highly those essays which are made up of shorter sentences, T-units or clauses. While the Group 1 essays are the longest of all the three groups, they also feature the highest number of sentences, T-units and clauses, which thus tend to be rather compact. The quantitative findings also seem to suggest that examiners do not reward students' use of more elaborate phrases. The essays that display the highest index of phrasal elaboration and number of words before main verb are indeed those that failed the exam. While syntactic breadth – or wider range of syntactic constructions used – is usually taken as evidence of more complex interlanguage, examiners appeared to allocate a higher mark to the essays that displayed a high index of syntactic similarity, which obviously

entails that the same restricted range of constructions were repeated across the essays.

It could be hypothesised that these findings, which appear to run counter to the assumption that ‘more complex equals better’ (as far as language proficiency/quality and development are concerned), may be accounted for by the students’ L1 – Italian – and the typological, stylistic, and rhetorical differences between Italian and English. It is often assumed that Italian writers tend to resort to a more ‘flowery’ style which entails the use of longer sentences rich in subordination and coordination, unlike the more ‘compact’ English style. It is, moreover, a fact that Romance languages prefer post-modification and right-embededness while Germanic languages rely more frequently on premodification and left-embededness (Gyllstadt et al. 2014). The use of longer sentences and more elaborated noun phrases may thus have been viewed by the examiners as evidence that the students were still unable to depart from their L1 linguistic influence. Producing longer production units may also have led students to be less accurate (given the time constraints they had to work with), which likely resulted in lower ratings by the examiners. By the same token, students who were more ‘adventurous’ and attempted to use a wider range of constructions (earning lower indices of syntactic similarity) may have failed to control the accuracy of their use of those constructions that had still not been wholly automatized – again leading to lower marks.

5. CONCLUDING REMARKS

The study of complexity in the second language acquisition field is a relatively recent endeavour and while great strides have been made in the last few decades in the way the concept has been operationalized and measured, little is yet known about how its subdimensions interact with other aspects of language proficiency and with factors affecting language development. This is brought home by the findings of studies on complexity and writing quality, which seem to be corroborated by the exploratory investigation reported on in the latter part of this paper. In particular, when complexity indices are correlated with human ratings of writing quality, factors such as the task genre or the relation of the L1 and the L2 appear to skew examiners’ judgements away from the simple axiom ‘more complex – more breadth or depth – equals higher quality’. However tentative, such findings point to the need to investigate in more depth some of these ‘confounding’ factors (e.g. by having the same informants produce samples of different written genres and/or produce samples of the same genre in both the L1 and the L2) using finer-grained measures of subdimensions of complexity (e.g. types of subordinate clauses: complement, relative etc.). They also raise questions about

the key issue of assessor training. While experienced assessors/examiners are bound to be familiar with the concepts of fluency and accuracy, they are perhaps less aware of the notion of complexity in second language acquisition, its multidimensional nature and the ‘complex’ relations it can have with the other elements of the CAF triad.

References

- Bernardini P.-Granfeldt J., 2019, *On cross-linguistic variation and measures of linguistic complexity in learner texts: Italian, French and English*, «International Journal of Applied Linguistics» 29.2: 211-232.
- Biber D.-Gray B., 2016, *Grammatical Complexity in Academic English. Linguistic change in writing*, Cambridge, Cambridge University Press.
- Biber D.-Gray B.-Poonpon K., 2011, *Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?* «TESOL Quarterly» 45.1: 5-35.
- 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*, Amsterdam, John Benjamins: 21-46.
- , 2014, *Conceptualizing and measuring short-term changes in L2 writing complexity*, «Journal of Second Language Writing» 26: 42-65.
- , 2018, *Syntactic complexity in L2 writing: Individual pathways and emerging group trends*, «International Journal of Applied Linguistics» 28.1: 147-164.
- , 2019, *Beginning L2 complexity development in CLIL and non-CLIL secondary education*, «Instructed Second Language Acquisition» 3.2: 153-180.
- Bygate M., 1999, *Quality of language and purpose of task: Pattern of learners' language on two oral communication tasks*, «Language Teaching Research» 3.3: 185-214.
- Crossley S. A.-McNamara D. S., 2014, *Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners*, «Journal of Second Language Writing» 26: 66-79.
- Davydova J., 2011, *The Present Perfect in Non-Native Englishes. A corpus-based study of variation*, Berlin, Mouton de Gruyter.
- De Clercq B.-Housen A., 2017, *A cross-linguistic perspective on syntactic complexity in L2 development: Syntactic elaboration and diversity*, «The Modern Language Journal» 101.2: 315-334.
- Ellis R., 2003, *Task-based Language Learning and Teaching*, Oxford, Oxford University Press.

- Ellis, R.-Barkhuizen G., 2005, *Analysing Learner Language*, Oxford, Oxford University Press.
- Gyllstadt H.-Granfeldt J.-Bernardini P.-Källkvist M., 2014, *Linguistic correlates to communicative proficiency levels of the CEFR: The case of syntactic complexity in written L2 English, L3 French and L4 Italian*, in L. Roberts-I. Vedder-J. H. Hulstijn (eds.), *EUROSLA yearbook 14*, Amsterdam, John Benjamins: 1-30.
- Housen A.-Kuiken F., 2009, *Complexity, accuracy and fluency in second language acquisition*, «Applied Linguistics» 30.4: 461-473.
- Housen A.-Kuiken F.-Vedder I., 2012, *Complexity, Accuracy and Fluency*, in A. Housen-F. Kuiken-I. Vedder (eds.), *Dimensions of L2 Performance and Proficiency. Complexity, accuracy and fluency in SLA*, Amsterdam, John Benjamins: 1-20.
- Howatt A.P.R., 2004, *A History of English Language Teaching*, Oxford, Oxford University Press.
- Hunt K. W., 1965, *Grammatical Structures Written at Three Grade Levels*. NCTE research report no. 3, Champaign, IL, National Council of Teachers of English.
- Kuiken F.-Vedder I., 2011, *Task performance in L2 writing and speaking: The effect of mode*, in P. Robinson (ed.), *Second Language Task Complexity. Researching the Cognition Hypothesis of language learning and performance*, Amsterdam, Benjamins: 91-104.
- , 2019, *Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish*, «International Journal of Applied Linguistics» 29.2: 192-210.
- Kuiken, F.-Vedder I.-Housen A.-De Clercq B., 2019, *Variation in syntactic complexity: Introduction*, «International Journal of Applied Linguistics» 29.2: 161-170.
- Kusters C. W., 2003, *Linguistic Complexity. The influence of social change on verbal inflection*, Utrecht, LOT.
- Lan G.-Liu Q.-Staples S., 2019, *Grammatical complexity: 'What Does It Mean' and 'So What' for L2 writing classrooms*, «Journal of Second Language Writing» 46: 1-7.
- 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» 27.4: 590-619.
- Larsen-Freeman D.-Cameron L., 2008, *Complex Systems and Applied Linguistics*, Oxford, Oxford University Press.
- Lu X., 2010, *Automatic analysis of syntactic complexity in second language writing*, «International Journal of Corpus Linguistics» 15.4: 36-62.
- , 2011, *A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development*, «TESOL Quarterly» 45.1, 36-62.
- Lu X.-Ai H., 2015, *Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds*, «Journal of Second Language Writing» 29: 16-27.
- Nava A., 2018, *English grammaticography for university students in Italy (1999-2011): Pedagogical grammars or pedagogical presentations of linguistic theories?*, «Studi Italiani di Linguistica Teorica e Applicata» 47.2: 249-268.

- Mazgutova D.-Kormos J., 2015, *Syntactic and lexical development in an intensive English for Academic Purposes programme*, «Journal of Second Language Writing» 29: 3-15.
- Miestamo M., 2008, *Grammatical complexity in a cross-linguistic perspective*, in M. Miestamo-K. Sinnemäki-F. Karlsson (eds.), *Language Complexity. Typology, contact, change*, Amsterdam/Philadelphia, John Benjamins: 22-41.
- Norris J.M.-Ortega L., 2009, *Towards an organic approach to investigating CAF in instructed SLA: The case of complexity*, «Applied Linguistics» 30.4: 555-578.
- Ortega L., 2003, *Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing*, «Applied Linguistics» 24.4: 492-451.
- , 2015, *Syntactic complexity in L2 writing: Progress and expansion*, «Journal of Second Language Writing» 29: 82-94.
- Pallotti G., 2014, *A simple view of linguistic complexity*, «Second Language Research» 31.1: 117-134.
- Polio C.-Yoon H. J., 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.
- Skehan P., 1998, *A Cognitive Approach to Language Learning*, Oxford, Oxford University Press.
- 2009, *Modelling second language performance: Integrating complexity, accuracy, fluency and lexis*, «Applied Linguistics» 30.4: 510-532.
- Skehan P.-Foster, P., 2012, *Complexity, accuracy, fluency and lexis in task-based performance*, in A. Housen-F. Kuiken-I. Vedder (eds.), *Dimensions of L2 Performance and Proficiency. Complexity, accuracy and fluency in SLA*, Amsterdam, John Benjamins: 199-220.
- Wolfe-Quintero K.-Inagaki S.-Kim H.-Y., 1998, *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*, University of Hawaii, Second Language Teaching & Curriculum Center.
- Yang W.-Lu, X.-Weigle S. C., 2015, *Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality*, «Journal of Second Language Writing» 28: 53-67.