24

Complexity

Nina Vyatkina & Alex Housen

Abstract

This chapter focuses on the construct of second language complexity: the range of linguistic (e.g., lexical, syntactic, morphological) forms and the degree of elaboration of these forms in learner language. The chapter reviews how measures of second language complexity have been used to characterize learner proficiency and to chart learner development with the help of learner corpus research methods and tools, as well as how complexity is modulated by task effects and language typology features.

Introduction

In the past few decades, the notion of complexity has raised significant interest in a wide range of disciplines, including Second Language Acquisition (SLA) research. Already in the 1970s and 1980s the notions of *complexification* and its antonym, *simplification,* figured prominently in the first theoretical models of SLA (e.g., Andersen, 1983). Skehan (1989) was the first to propose a model of SLA that added complexity as one of three basic dimensions (with accuracy and fluency) in terms of which L2 learners' performance, proficiency, and development could be investigated (see also Chapters 23 and 24, this volume). At that time, L2 complexity was also given its working definition, which is still widely used today, as the range of forms (i.e. items, structures, patterns, rules) available to a learner and the degree of elaboration of these forms (Ortega, 2003).

Complexity has been measured by means of (subjective) ratings and, more typically, (objective) quantitative measures, of which there are many (see reviews in Ortega, 2003; Wolfe-Quintero, Inagaki, & Kim, 1998). Early studies of L2 complexity (1970s – early 2000s) typically handled a limited amount of learner language as most coding of complexity indicators was conducted manually. The measures were limited to a handful of global, broad measures such as average length of syntactic units (sentences, T-units, clauses) in words and amount of embedding (ratio of subordinate clauses to total clauses). Higher values of these measures were found to be associated with higher L2 proficiency. However, the progress in complexity research was slowing down at the turn of the century, due to its narrow focus on a few measures, lack of theoretical motivation for these measures, and other methodological limitations. The seminal article by Norris and Ortega (2009) has triggered a new wave of interest with their call for a more 'organic approach' to the study of L2 complexity. In particular, Norris and Ortega

invited researchers to supplement global, coarse-grained measures (e.g., mean length of sentence, clauses per T-unit) with finer-grained, more specific measures of phrasal and clausal complexity (e.g., infinitival phrases, passives), and to explore the dynamic relationships among different measures at different stages of L2 development. They also called for building connections to corpus linguistics. This call was echoed by Bulté and Housen (2012), who reviewed 40 studies (1995-2008) and concluded that the majority had used 1-2 complexity measures only, "probably due to the lack of adequate computational tools for automatic complexity measurement and the labour-intensiveness of manual computation" (p. 34).

The situation changed drastically in the 2010s with the development of Natural Language Processing (NLP) tools (especially those with open access) that expanded the repertoire of complexity measures as well as enabled automatic retrieval and analysis of large amounts of language production data from learner corpora. Recently, the field has seen an exponential growth of studies that explored SLA-motivated research questions with Learner Corpus Research (LCR) tools and methods. Rich and nuanced analyses have revealed a more intricate picture of L2 complexity and its development over time, pointed to significant effects of various learner, register, and task variables, and have prompted researchers to take a more critical look at L2 complexity as a construct.

It is worth noting that it is not always easy to distinguish between LCR and non-LCR L2 complexity studies. Many studies have been conducted on learner production data not characterized by authors as learner corpora. The difference between such data pools (or convenience corpora) and learner corpora seems to be that the former are smaller in size, limited in terms of registers and text types, and are collected and utilized only for a specific

study (or a few studies), with no declared intent of making them available to a larger research community – which is typical of learner corpora.

Core issues and topics

Defining L2 complexity

Empirical L2 research on complexity has produced many inconsistent and inconclusive results (Bulté & Housen, 2012; Pallotti, 2015). The problematic status of the complexity construct in SLA research is mainly due to its complex (i.e. multidimensional, multi-faceted, and multilayered) nature and its resultant polysemy. Indeed, the term complex(ity) has been equated with a seemingly disparate variety of terms such as (more) advanced, late(r) acquired, (more) developed, (more) proficient, (more) sophisticated, (more) challenging, (more) difficult, (more) problematic, (more) elaborate, (more) embedded, (more) frequent, long(er), rare(r), (more) marked, (more) diverse or varied, rich(er), better, and (more) mature. The question is whether all these terms refer to one and the same underlying construct or to conceptually distinct and analytically separable constructs that may or may not be empirically (cor)related. To overcome this terminological confusion, researchers have recently called for a more parsimonious approach to defining L2 complexity strictly in terms of *absolute complexity*, which refers to quantitative properties of linguistic structures or systems, as reflected by the number and range of phonemic, lexical, syntactic or morphological items, structures, or rules in a language (Bulté & Housen, 2012). The overwhelming majority of

studies have used syntactic or lexical complexity measures, or combinations thereof, whereas interest in other measures (e.g., morphological complexity) started emerging only recently.

L2 complexity measures as indicators of L2 proficiency and development

Syntactic complexity

Syntactic complexity has been the most popular construct, defined as breadth (length, diversity) and depth (embeddedness) of syntactic units: T-units, AS-units, sentences, clauses, and phrases (Bulté & Housen, 2012, p. 27). The overwhelming majority of syntactic complexity studies have based their findings on the material of academic L2 English essays written by university students. Ortega's (2003) meta-analysis of 25 pre-corpus era (1975-1998) SLA studies found that values of four measures linearly increased with increasing proficiency: mean length of sentence (MLS), mean length of T-unit (MLTU), mean length of clause (MLC), and clauses per T-unit (C/TU). She also found that for these measures to significantly discriminate between proficiency levels, at least one year of instructed study was necessary.

Ortega (2003) has cautioned that her meta-analysis was limited to global syntactic complexity metrics and invited researchers to explore more specific metrics as well as their dynamic relationships in future studies. Norris and Ortega (2009) later hypothesized that the generally attested developmental increase in values of the most global measures – MLS and MLTU – was caused by growth in different lower-level measures at different L2 proficiency levels: coordination at beginning levels, subordination at intermediate levels, and phrasal elaboration at advanced levels.

Subsequent LCR studies have largely confirmed this hypothesis. Byrnes, Maxim, and Norris (2010) analyzed a corpus of learner essays written at the end of four curricular levels of a US university L2 German program both cross-sectionally and longitudinally. In both designs, they found a significant increase in the sentence length over all levels, whereas subordination increased up to level 3 (intermediate proficiency) and then plateaued, and mean length of clause (attributed to noun phrase complexification) started increasing only at the juncture between levels 2 and 3 (thus marking a transition to more advanced proficiency). While Byrnes et al. (2010) coded their data manually for only three global measures, subsequent LCR studies have employed multiple measures of both global and specific syntactic complexity using NLP tools. Lu (2010) introduced his L2 Syntactic Complexity Analyzer (SCA) for 14 such measures and applied them to a corpus of academic essays of Chinese learners of English. He found that higher proficiency learners produced longer clauses and T-units as a result of increased use of complex phrases (coordinate phrases and complex nominals). These findings were further expanded by multidimensional LCR studies (see section Task Effects / Register Differences).

Lexical Complexity

Lexical complexity (Bulté & Housen, 2012, p. 28) has been defined as lexical breadth, subdivided into density (ratio of lexical to function words) and diversity (type-token ratio, or TTR), and depth, subdivided into compositionality (morphemes or syllables per word) and sophistication (ratio of rare words). TTR was the first automatically calculated lexical measure to be used in L2 complexity research. However, its validity has been questioned because of heavy dependence on text length, and adjusted TTR as well as other diversity measures have

been proposed instead such as the Measure of Textual Lexical Diversity (MTLD, McCarthy & Jarvis, 2010).

Gradually, with the development of NLP tools, researchers started applying a wide range of lexical measures to the study of L2 complexity. The L2 Lexical Complexity Analyzer (LCA; Lu, 2012) automatically computes 25 lexical measures. Lu (2012) applied it to a spoken L2 English (L1 Chinese) corpus and found that several measures of lexical diversity, but not density or sophistication, correlated with holistic measures of text quality. Lexical diversity measures were also stronger proficiency predictors in a cross-sectional learner corpus of L2 French and L2 English (L1 Dutch) oral narratives (De Clercq, 2015) and better short-term development indicators in L2 English written descriptive essays (Bulté & Housen, 2014) than other types of lexical measures. Furthermore, longitudinal LCR studies have shown that lexical diversity develops in a curvilinear rather than a linear fashion: a steep initial increase is followed by a gradual flattening of the curve (Daller, Turlik, & Weir, 2013; Mitchell, Tracy-Ventura, & McManus, 2017).

Combining syntactic and lexical measures

L2 complexity studies that combined both syntactic and lexical complexity measures generally found that "different subcomponents of syntactic and lexical complexity […] develop at a different pace" (Bulté & Housen, 2014, p. 42). In a cross-sectional study of L2 English (L1 Dutch) written narratives with 64 hand-coded complexity and accuracy measures, Verspoor, Schmid, and Xu (2012) found that significant differences between adjacent proficiency levels alternated between lexical and syntactic measures. Longitudinal studies showed similar results. Verspoor, Lowie, and van Dijk (2008) explored the writing of an advanced L2 English learner

over three years from the Complex Dynamic Systems Theory standpoint (see Chapter 14, this volume). They found that, despite an overall increase in global syntactic (MLS) and lexical complexity (TTR), there were considerable peaks and dips in the growth curve. Furthermore, dips in lexical complexity were accompanied by peaks in syntactic complexity and vice versa, and only at the end of the observation period did both types of measures develop in unison.

Morphological complexity

Investigations of morphological complexity in L2 corpora have been taken up only recently. One such study is De Clercq and Housen (2019), which is presented in the section Representative Corpora and Research. Another study is Brezina and Pallotti (2019), who built on the Morphological Complexity Index (MCI): the diversity of morphological exponents (the forms taken by lexemes to express grammatical categories and functions) in a text sample (Pallotti, 2015). Brezina and Pallotti developed an automated, web-based version of the MCI analyzer which they applied to two corpora of argumentative written texts from native and non-native speakers of Italian and English (respectively a morphologically rich and poor language). They found that MCI varied between native and non-native speakers of Italian and was significantly lower in lower proficiency learner samples. In L2 Italian, MCI was strongly correlated with proficiency, and also showed significant correlations with other measures of complexity, such as lexical diversity and sentence length. However, when applied to a corpus of advanced L2 English learners, MCI remained constant across natives and learners, and did not correlate with other complexity measures. These results suggest that morphological L2 complexity is a function of speakers' proficiency and the specific language under investigation: for languages with a relatively simple inflectional morphology, such as English, learners will quickly reach a threshold level after which inflectional morphological diversity remains

constant.

Lexico-grammatical complexity

Whereas the studies reviewed above consider complexity measures at different linguistic levels as tapping into distinct constructs, others have assumed an integrated view of complexity as a lexico-grammatical construct. Paquot (2019) has introduced the construct of *phraseological* complexity, operationally defined as the diversity of phraseological units in text samples (type-token ratio of specific words co-occurring in grammatical relations, e.g., adjective + noun, verb + direct object) and the sophistication of those units (the mutual information score of these units). Paquot found that these lexico-grammatical measures reliably discriminated between upper-intermediate to advanced proficiency levels in a written L2 English corpus, unlike separate syntactic and lexical complexity measures.

Researchers working in the usage-based *construction grammar* paradigm (see Chapter 14, this volume) also argue that complexity measures should integrate both grammatical and lexical properties. Kyle (2016) and Kyle, Crossley, and Berger (2017) developed TAASSC and TAALES - automated syntactic and lexical complexity analyzers for English - that offer, in addition to traditional measures (such as ones in Lu's SCA and LCA), an impressive range of usage-based measures. What sets these measures apart from traditional ones is their interface nature: they go beyond length, frequency, and ratios of discrete syntactic and lexical elements and take into account frequency of occurrence of lexical items within specific syntactic constructions as well as strength of association between lexical items and constructions. In a series of recent studies, Kyle and colleagues applied TAASSC and TAALES to the learner

corpus of TOEFL exams and found that usage-based measures, especially at the phrasal level, predicted learner writing quality better than traditional measures (Kyle & Crossley, 2018; Kim, Crossley & Kyle, 2018). These usage-based studies were in part informed by functional multi-dimensional approaches to corpus research reviewed in the next section.

Task effects

The definition of task (not unlike complexity) has been the subject of much SLA debate that is beyond the scope of this chapter (see, e.g., Robinson, 2011; Skehan, 1998). For our purposes here, we consider task effects associated with several design features of learner language data elicitation procedures applied in L2 complexity research: *register* and *cognitive task complexity*.

Register differences

Register differences have been extensively studied by Biber and colleagues in their L1 corpus research spanning several decades since 1980s (Biber, 1988). Their approach was termed functional and multidimensional because it showed that multiple heterogeneous linguistic units (e.g., syntactic units, parts-of-speech, semantic word classes, lexical items) co-occur in specific registers to express certain communicative functions (see Chapter 19, this volume). A designated NLP tool, the *Biber Tagger* (Biber, 1988, see also Nini, 2014), was developed for annotating corpora for a wide range of these features. As far as linguistic complexity is concerned, the researchers have argued that its nature depends on the register: e.g., longer phrases and sentences are needed to express arguments, whereas multiple and varied subordinate clauses are needed in narration (e.g., Biber & Gray, 2010). Based on these findings

from L1 English corpora, Biber, Gray, and Poonpon (2011) also hypothesized that L2 writing should exhibit spoken complexity features at beginning proficiency levels and written complexity features at later stages (which was in line with SLA studies like Norris & Ortega, 2009, discussed in 2.2.1 above). This assumption was confirmed by Taguchi, Crawford, and Wetzel (2013) and Parkinson and Musgrave (2014), who found more noun modifiers and fewer dependent clauses in the writing of more proficient L2 English learners. Biber, Gray, and Staples (2016) explored the corpus of spoken and written TOEFL exam productions and found that not only mode and register but also L2 proficiency were robust predictors of two sets of lexico-grammatical features, one characterizing the orate and the other the literate dimension. For example, written texts, informational summaries, and higher-rated texts had more nouns, nominalizations, and noun phrase modifiers; whereas spoken texts, personal accounts, and lower-rated texts had more finite adverbial clauses and adverbs. Yoon and Polio (2017) found similar register-related effects in a written ESL corpus: namely, MLS and phrasal complexity was higher in argumentative than in narrative essays. Furthermore, lexical sophistication and richness was higher in argumentative essays, whereas lexical diversity was higher in narrative essays.

Cognitive task complexity

The effects of cognitive task complexity (Robinson, 2011; Skehan, 1998) on the complexity of language production have received considerable attention in Task-Based Language Teaching (TBLT) SLA studies (e.g., Ellis, 2005; Kuiken & Vedder, 2012; Lambert & Kormos, 2014; Révész, Kourtali & Mazgutova, 2017; Vasylets, Gilabert & Manchón, 2017). Collectively, this research indicated that cognitively more complex tasks (in terms of reasoning demands made

on the learner), written tasks, and tasks that allow for pre-task or online planning tend to result in syntactically and lexically more complex language (e.g., more subordination, longer clauses and sentences, more diverse and sophisticated vocabulary). On the other hand, some results were inconsistent across studies, as shown in Bulté and Housen's (2012) review of 40 TBLT studies on complexity.

Until recently, there has been little overlap between TBLT research and LCR because data elicitation tasks in the available larger learner corpora have not been differentiated by nuanced task complexity conditions (Alexopoulou, Michel, Murakami & Meurers, 2017). Alexopoulou et al. (2017) classified essay prompts in the EFCAMDAT corpus from three registers *post hoc* according to the level of cognitive load each task imposed on the learner. They found the register effect to be stronger than the cognitive load effect: i.e., complexity values were similar within one register for both the simple and difficult prompt. More specifically, the authors found that narratives were more complex than descriptions based on three global syntactic complexity measures (MLS, MLC, C/TU) and the MTLD (lexical measure).

L1 and L2 effects

Most L2 complexity studies reviewed above have been conducted on ESL learner productions with only a few representatives from different L1s which made it impossible to explore L1 and L2 effects. Crossley and McNamara (2012) was one of the first LCR studies that focused on the L1 variable. They found that lexical complexity measures were the strongest predictor for L1 identification, but that syntactic phrasal complexity also distinguished between four L1 groups from the ICLE corpus. Lu and Ai (2015) conducted a comparative study on a larger scale and found that all 14 syntactic measures from Lu's SCA tool reliably discriminated

among ICLE essays from seven L1 groups as well as L1 English essays. Research of the L2 effect also shows that complexity develops differently in the acquisition of different languages (see Housen, De Clercq, Kuiken & Vedder, 2019). Future LCR studies are needed that would tie L1 and L2 effects on complexity to differences in linguistic typology (Miestamo, 2008).

Main research methods

Methods

As can be inferred from the review above, learner corpus complexity research has typically been quantitative in nature. Most studies have been cross-sectional, although a recent trend toward more longitudinal research can be observed. Early L2 complexity studies have applied t-tests and ANOVAs to compare cross-sections of L2 data, as was typical of SLA research in general. More recently, there has been a gradual shift in both SLA and LCR to multifactorial methods (see Chapters 9 and 10, this volume, for more information). For example, Biber et al. (2016) used a factor analysis with General Linear Models to group complexity features and to explore which independent variables (mode, register, proficiency) predicted co-occurrence of these features. Kim et al. (2018) used a Principal Component Analysis to cluster complexity measures that were highly correlated with each other into groups and to then reduce the variables to a smaller set of derived variables. In a longitudinal study, Chan, Verspoor and Vahtrick (2015) applied the Hidden Markov Model to explore the development of complexity in L2 speech and writing of two identical twins. With the help of this non-linear method, the authors found that values for specific complexity measures changed dynamically and that developmental paths of the twins differed drastically (contrary to expectations).

Tools

One distinctive characteristic of LCR is the use of NLP tools (Meurers & Dickinson, 2017) for extracting and computing frequencies of linguistic features. The number and variety of these features has been growing exponentially, as reflected in Table 1. The use of some tools requires considerable computational expertise while others, especially more recent ones, can be used by a regular "teacher, textbook writer, or second language acquisition researcher" due to their "user-friendly interface and visualization features" (Chen & Meurers, 2016, pp. 113-114). It is important to note that, like other NLP tools, complexity analyzers have originally been developed for and trained on texts produced by native speakers, but many have also been applied to learner texts. The accuracy of NLP tools for learner data (i.e., non-standard language data) has been a subject of debate, and complexity analyzers are no exception. For example, while MLS or MLTU can be fairly reliably calculated in learner written texts, lexical diversity ratios may be inflated due to spelling errors or the number of subordinate clauses miscalculated due to lexical or syntactic errors (Meurers & Dickinson, 2017). In a much-needed study of reliability and validity of complexity analyzers, Polio and Yoon (2018) explored the SCA (Lu, 2010) and Coh-Metrix (McNamara, Graesser, McCarthy & Cai, 2014) tools. They found that the majority of syntactic complexity measures in both tools served as reliable and valid identifiers of register differences in L2 written data, although some measures were not transparent and consistent. Automatic coding and analysis problems are augmented in the case of spoken learner data, which, depending on the transcription format used, typically need to undergo multiple steps of preliminary data segmentation and 'cleaning'.

Table 1. List of automated tools for text complexity analysis in chronological order of publication

Title	Website	Target language(s)	Complexity features	Reference
Computerized Language Analysis (CLAN)	https://talkbank.org	English, French, German	Some basic lexical and syntactic (TTR, VOCD, MLT, MLU)	MacWhinney (2000)
The Compleat Lexical Tutor	http://www.lextutor.ca	English, French	Lexical rarity (a sophistication measure)	Cobb (n. d.)
Text Inspector	http://textinspector.com	English	Lexical diversity	Bax (n. d.)
L2 Syntactic Complexity Analyzer	http://www.personal.psu.edu/xxl13/downloads/l2sca.html	English	14 syntactic	Lu (2010)
Lexical Complexity Analyzer	http://www.personal.psu.edu/xxl13/downloads/lca.html	English	25 lexical	Lu (2012)
Coh-Metrix	http://cohmetrix.com	English	100+ lexico-grammatical	McNamara et al. (2014)

T-Scan	https://github.com/proycon/tscan	Dutch	lexical, syntactic, morphological	Pander Maat et al. (2014)
Multidimensional Analysis Tagger	https://sites.google.com/site/multidimensionaltagger/	English	Multiple lexico-grammatical, based on Biber (1988)	Nini (2014)
Lancaster Vocab Analysis Tool (LancsLex)	http://corpora.lancs.ac.uk/vocab/index.php	English	Lexical diversity and rarity	Brezina and Gablasova (2015)
CorpusExplorer	http://corpusexplorer.de	German	45 lexical	Rüdiger (2016)
Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC)	https://www.linguisticanalysistools.org/taassc.html	English	Multiple lexico-grammatical	Kyle (2016)
Common Text Analysis	http://www.ctapweb.com/home.html	English (German in development)	170+ lexical and syntactic	Chen and Meurers (2016)

Platform (CTAP)				
Tool for the Automatic Analysis of Lexical Sophistication (TAALES)	https://www.linguisticanalysistools.org/taales.html	English	400+ lexical	Kyle et al. (2017)
Morpho complexity tool	http://corpora.lancs.ac.uk/vocab/analyse_morph.php	English and Italian (French, German, Spanish in development)	Morphological (one index)	Brezina and Pallotti (2019)

Representative corpora and research

Most research reviewed above was conducted on written L2 English data. To highlight some other corpora and associated studies, this section presents an L2 German corpus and a spoken corpus of three L2s - English, French, and Dutch – that have been analyzed in terms of linguistic complexity.

Falko (FehlerAnnotiertes LernerKOrpus): an error-annotated learner corpus of German as a Foreign Language

Falko is available with free and open access on the website maintained by the corpus linguistics team at the Humboldt University Berlin (https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko). Falko consists of several German as a Foreign Language subcorpora, all of them comprising academic essays written by learners of German at different universities around the world. What sets Falko apart from other error-annotated learner corpora (e.g., ICLE) is that its texts are annotated on multiple layers including the so-called "target hypotheses", or explicit corrections of learner errors by annotators. This unique feature of Falko allows for analyzing learner language complexity separately from accuracy. To illustrate this, we review an L2 complexity study conducted on KANDEL (Kansas Developmental Learner corpus; Vyatkina, 2016). KANDEL is a Falko subcorpus that comprises short essays collected at dense time intervals from American learners of German as they progressed from no or minimal L2 proficiency to intermediate proficiency over two years of instructed collegiate study.

Vyatkina, Hirschmann, and Golcher (2015) explored KANDEL for the development of a specific type of syntactic complexity, operationally defined as syntactic modification. Modifiers were defined as optional elements that attach to the heads of noun phrases, verb phrases, or sentences. To find surface proxies for modifiers, the corpus was searched for POS tags for attributively and predicatively used adjectives, ordinal numerals, adverbs, subordinating conjunctions, and relative pronouns. However, automated POS-tagging was not applied to raw texts here because many spelling and grammatical mistakes made by beginning learners were expected to lead to many tagger errors. Instead, the "Target Hypothesis 1"

annotation layer and its automatically assigned POS-tags were used. These annotations contain minimal corrections of spelling and morpho-syntactic errors to form a grammatical German sentence, while keeping it as close to the original surface structure as possible and leaving semantic, pragmatic, and discourse errors uncorrected (Lüdeling & Hirschmann, 2015).

The study found that learners started using syntactic modification from their very first essay written in German, but the frequencies of different modifiers changed as they progressed in their studies. In particular, the increase in more structurally complex (inflected, clausal) modifiers was accompanied by a decrease in less complex modifiers. Furthermore, the growth trends were not linear but curvilinear, with smoother or abrupt curves for different measures. Finally, instruction did not appear to cause the onset of development or to change its direction but instead to modulate growth curves.

Brussels Instructed Language Learner Corpus (BILLC) and the cross-linguistic development of L2 complexity

The Brussels Instructed Language Learner Corpus (BILLC) consists of oral and written L2 English, French and Dutch production data, collected both cross-sectionally and longitudinally, from L2 learners from a range of L1 backgrounds (mainly Dutch, French, English, Italian, and German), age levels (child, adolescent and adult learners) and learning contexts (traditional foreign-language classrooms, CLIL/immersion, European School). This multilingual corpus contains narrative, descriptive, and argumentative texts. The data have been transcribed in the CHAT format (MacWhinney, 2000), segmented into clausal units, and parts of the corpus were error and/or POS annotated. BILLC will be freely online accessible via a Vrije Universiteit Brussel server in the foreseeable future. Studies that have used this corpus for complexity

analyses were carried out on a parallel bilingual subcorpus of 100 L2 French and 100 L2 English oral narratives representing four different proficiency levels and collected cross-sectionally from Dutch-speaking 12-18-year-old students learning these two L2s in the Flemish education system. Data from matched native speakers of French and English served as benchmarks for the complexity analyses. De Clercq (2015) and De Clercq and Housen (2017) focused on lexical complexity and syntactic complexity, respectively (see section Core Issues and Topics). De Clercq and Housen (2019) focused on morphological complexity: this study will be reviewed here as an illustration.

De Clercq and Housen (2019) examined how and when cross-linguistic differences in the complexity of morphological systems surface in oral L2 productions and to what extent measures of morphological complexity can function as cross-linguistically reliable indices of linguistic development. Additionally, the study compared three recently proposed measures that operationalize the notion of complexity in different ways: the Morphological Complexity Index (MCI; Pallotti, 2015), the Types per Family index (T/F-I; Horst & Collins, 2006), and the measure of Inflectional Diversity (ID; Malvern, Chipere, Richards, & Durán, 2004). While all three measures operationalize complexity as diversity, they differ considerably in scope and in their mathematical operationalization. The study aimed at verifying whether these measures function more effectively as metrics of development in French, a morphologically richer language, than in English, a morphologically poorer language. All three measures were found to increase stronger and more significantly across the four proficiency levels in French than in English. Contrary to expectations, learners in both languages eventually reached native speaker levels of morphological diversity, although at lower proficiency levels in English than in French.

When comparing the findings of the three measures, considerable differences in attained levels of morphological complexity were found. The range of scores and developmental trends for the ID and MCI measures were the most similar for both target languages, although MCI scores covered a more comparable range in English and French, while scores for ID were overall higher in the French data than in the English data. The scores obtained by the TF-I differed from the other two measures and indicated few differences across proficiency levels and languages. The authors conclude that MCI seems to be best suited for corpora similar to BILLC, especially because it works for shorter texts with a low ratio of word forms to lemmas.

Future directions

This chapter has highlighted the need for further conceptual clarification and operational refinement of L2 complexity constructs, reliable and valid corpus annotation, and automated retrieval of non-redundant measures. First, existing measures need to be evaluated for their capacity to validly and reliably assess complexity of relatively small text samples, which are typical of L2 data and L2 corpora. Many learner corpora may now be sufficiently large (in terms of the total number of words they contain) but they are frequently made up of short text samples (e.g. only 50-500 words long or even shorter), which detracts from the reliability and the usability of many of the currently available complexity measures. This specific issue of data 'scarcity' in LCR needs to be addressed. On the other hand, LCR is uniquely positioned to take advantage of big data and associated research methods. An example of how large corpora can be used for comparative LCR purposes is Alexopoulou et al. (2017) described in the section Core Issues and Topics. Furthermore, in order to better understand the cross-

linguistic manifestations of complexity in the course of L2 development, there is a need for massive parallel (or multilingual) learner corpora and computational tools for their analysis. Such massive parallel corpora are now being compiled in language typology research (Cysouw & Wälchli, 2007).

Next, new conceptualizations of L2 linguistic complexity and new measures are needed, including measures based on information-theoretic definitions of complexity (Ehret & Szmrecsanyi, 2019). There is also need to go beyond syntactic and lexical complexity to investigate its other types (morphological complexity, phonological complexity, complexity at the interfaces). While the recent inclusion of morphological diversity measures (such as the MCI, TF-I, and ID discussed in this chapter), syntactic diversity measures (as in De Clercq & Housen, 2019), and phraseological diversity and sophistication (as in Paquot, 2019) is already an important step toward finding valid and cross-linguistically representative and reliable indicators, future research should consider measures beyond diversity and sophistication. For instance, with respect to morphological complexity, L2 research could implement approaches developed in language typology, where studies have also looked at the degree of deviation from a base form, the number of morphological elements in a word (Dammel & Kürschner, 2008), or morphological productivity (Lüdeling, Hirschmann, & Shadrova, 2017). Such a broadening of perspective and scope could also prove informative for the study of a typologically wider range of L2s. Future research should also reach out beyond linguistic complexity and explore other manifestations such as propositional and interactional complexity (Bulté & Housen, 2012; Vasylets et al., 2017).

Finally, given the multicomponential and multidimensional nature of L2 complexity, the restricted scope of individual complexity measures, the non-linear character of individual

developmental paths provided on the basis of individual measures, and the complex relationships between the individual measures and components, it may be worthwhile to combine different complexity measures in order to obtain a perhaps less detailed and fine-grained, but more comprehensive picture of overall L2 complexity and its development. Particularly when the aim is to assess (overall) L2 development or proficiency, combined measures have the potential of outperforming individual measures, since they offer a more global picture of L2 complexity. Such an undertaking necessarily involves the reduction of multiple dimensions into a single one, and losing detailed information, but on the other hand, single scores have the advantage of being easier to interpret than multiple scores, especially when many distinct dimensions are involved.

To conclude, corpus-based research on L2 complexity has a rich potential. Beyond providing yardsticks for L2 development and descriptors of L2 performance, this research can make significant contributions to theory construction in SLA by capturing relevant properties of interlanguage at each developmental stage. LCR studies can also help identify the developmental mechanisms for how and why L2 learners develop L2 features in a particular order. With larger and more representative corpora, with a wider range of L2s, learner and task variables, and the application of longitudinal and multidimensional designs and data analysis methods, both LCR and SLA can be propelled into new, mutually enriching directions.

Further reading

Connor-Linton, J., &, Polio, C. (Eds.). (2014). Comparing perspectives on L2 writing: Multiple

analyses of a common corpus [Special issue]. *Journal of Second Language Writing, 26.*

This special issue presents five studies of a common corpus of L2 English writing - the Michigan State University (MSU) corpus. The corpus comprises descriptive essays collected from students in an intensive English program at three time points over an academic semester. What sets this volume apart is that the same language data was analyzed from five different perspectives that focus, respectively, on accuracy, collocational strength, manually and automatically annotated complexity measures, and multidimensional analysis. All five studies also explored learner development over time and correlations of linguistic measures with holistic writing quality ratings.

Housen, A., De Clercq, B., Kuiken, F., & Vedder, I. (Eds.). (2019). Linguistic complexity [Special Issue]. *Second Language Research, 35*(1).

This recently published special issue charts new directions in L2 complexity research. Its editorial provides an overview of current methodological practices, discusses interfaces of L2 complexity research with theoretical and typological linguistics, and problematizes existing definitions and operationalizations of the complexity construct. The volume's five empirical studies focus on previously underexplored aspects of L2 complexity: morphological complexity, phraseological variety and sophistication, L1 effects, and compressibility of texts (an information-theoretical concept).

Vyatkina, N. (Ed.). (2015). New developments in the study of L2 writing complexity [Special

issue]. *Journal of Second Language Writing, 29.*

The five empirical contributions to this special issue all focus on L2 writing and address the development in lexical and syntactic complexity in L1 English and L2 German, L1 identification of L2 English essays based on complexity measures, complexity in CMC data from the TBLT perspective, as well as advanced writing complexity from the Systemic-Functional Linguistics perspective. Ortega's (2015) extended commentary synthesizes the accomplishments in the field and proposes future research directions.

Related Topics

Chapters 5, 8, 9, 13, 14, 18, 22, and 23.

References

Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning, 67*(S1), 180-208.

Andersen, R. (1983). *Pidginization and creolization as language acquisition.* Rowley, MA: Newbury House Publishers.

Bax, S. (n. d.). *Text Inspector* [Computer software]. Retrieved from http://textinspector.com

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University

Press.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*, 2-20.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use the characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*, 5-35.

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, *37*(5), 639-668.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics, 36*(1), 1-22.

Brezina, V., & Pallotti, G. (2019). Morphological complexity in written L2 texts. *Second Language Research, 35*(1), 99-119.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken & I. Vedder (Eds.). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42-65.

Byrnes, H., Maxim, H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment [Monograph]. *Modern Language Journal*, *94*(S1).

Chan, H., Verspoor, M., & Vahtrick, L. (2015). Dynamic development in speaking versus writing in identical twins. *Language Learning, 65*(2), 298-325.

Chen, X. B., & Meurers, D. (2016). CTAP: A web-based tool supporting automatic complexity analysis. *Proceedings of The Workshop on Computational Linguistics for Linguistic Complexity*. Osaka, Japan: The International Committee on Computational Linguistics.

Retrieved from https://www.aclweb.org/ anthology/W16-4113/

Cobb, T. (n. d.). *Compleat Lexical Tutor* [Computer software]. Retrieved from http://www.lextutor.ca

Connor-Linton, J., & Polio, C. (Eds.). (2014). Comparing perspectives on L2 writing: Multiple analyses of a common corpus [Special issue]. *Journal of Second Language Writing, 26*.

Crossley, S. A., & McNamara, D. S. (2012). Detecting the first language of second language writers using automated indices of cohesion, lexical sophistication, syntactic complexity and conceptual knowledge. In S. Jarvis & S. Crossley (Eds.), *Approaching language transfer through text classification: Explorations in the detection-based approach* (pp. 106-126). Bristol: Multilingual Matters.

Cysouw, M., & Wälchli, B. (2007). Parallel texts: using translational equivalents in linguistic typology. *Language Typology and Universals 60*(2), 95-99.

Daller, M., Turlik, J., & Weir, I. (2013). Vocabulary acquisition and the learning curve. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 185-217). Amsterdam/Philadelphia: John Benjamins.

Dammel, A., & Kürschner, S. (2008). Complexity in nominal plural allomorphy: A contrastive survey of ten Germanic languages. In M. Miestamo, K. Sinnemäki & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 243-261). Amsterdam: John Benjamins.

De Clercq, B. (2015). The development of lexical complexity in second language acquisition: A cross-linguistic study of L2 French and English. *EuroSLA Yearbook, 15*, 69-94.

De Clercq, B., & Housen, A. (2017). A cross-linguistic perspective on syntactic complexity in L2 development: Syntactic elaboration and diversity. *Modern Language Journal, 101*(2), 315-334.

De Clercq, B., & Housen, A. (2019). The development of morphological complexity: A cross-

linguistic study of L2 French and English. *Second Language Research, 35*(1), 71-97.

Ehret, K., & Szmrecsanyi, B. (2019). Compressing learner language: An information-theoretic measure of complexity in SLA production data. texts. *Second Language Research, 35*(1), 23-45.

Ellis, R. (Ed.). (2005). *Planning and task performance in a second language.* Amsterdam: John Benjamins.

Horst, M., & Collins, L. (2006). From faible to strong: How does their vocabulary grow? *Canadian Modern Language Review / La revue canadienne des langues vivantes*, *63*(1), 83-106.

Housen, A., De Clercq, B., Kuiken, F., & Vedder, I. (2019). (Eds.). Linguistic complexity [Special Issue]. *Second Language Research, 35*(1).

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: relations to second language lexical proficiency, development, and writing quality. *Modern Language Journal*, *102*(1), 120-141.

Kuiken, F., & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, Accuracy and Fluency in SLA* (pp. 143-170). Amsterdam: John Benjamins.

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Doctoral Dissertation). Retrieved from http://scholarworks.gsu.edu/alesl_diss/35

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *Modern Language Journal, 102*(2), 333-349.

Kyle, K., Crossley, S. A., & Berger, C. (2017). The tool for the analysis of lexical sophistication

(TAALES): Version 2.0. *Behavior Research Methods*, 1-17.

Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics, 35*(5), 606-614.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474-496.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *Modern Language Journal*, *96*(2), 190-208.

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.

Lüdeling, A., & Hirschmann, H. (2015). Error annotation systems. In S. Granger, F. Meunier & G. Gilquin (Eds.), *The Cambridge handbook of learner corpus research* (pp. 135-158). Cambridge: Cambridge University Press.

Lüdeling, A., Hirschmann, H., & Shadrova, A. (2017). Linguistic models, acquisition theories, and learner corpora: Morphological productivity in SLA research exemplified by complex verbs in German. *Language Learning, 67*(S1), 96-129.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. 3rd Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Malvern, D., Chipere, N., Richards, B., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills: Palgrave Macmillan.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*, 381-392.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2012). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

Meurers, D., & Dickinson, M. (2017). Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning, 67*(S1), 66-95.

Miestamo, M. (2008). Grammatical complexity in a cross-linguistic perspective. In M. Miestamo, K. Sinnemäki & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 23-42). Amsterdam: John Benjamins.

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships and language learning*. New York: Routledge.

Nini, A. (2014). *Multidimensional Analysis Tagger 1.2 - Manual.* Retrieved from: http://sites.google.com/site/multidimensionaltagger

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*, 492-518.

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing, 29*, 82-94.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*(1), 117-134.

Pander Maat, H., Kraf, R., van den Bosch, A., Dekker, N., van Gompel, M., Kleijn, S., Sanders, T., & van der Sloot, K. (2014). T-Scan: a new tool for analyzing Dutch text. *Computational Linguistics in the Netherlands Journal, 4*, 53-74.

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes,*

14, 48-59.

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, *35*(1), 121-145.

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*, 165-188.

Révész, A., Kourtali, N.-E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, *67*(1), 208-241.

Robinson, P. (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. Amsterdam: John Benjamins.

Rüdiger, J. O. (2016). *Corpus Explorer V2.0* [Computer software]. Retrieved from http://corpusexplorer.de

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

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, *47*, 420-430.

Vasylets, O., Gilabert, R., & Manchón, R. M. (2017). The effects of mode and task complexity on second language production. *Language Learning*, *67*(2), 394-430.

Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, *92*(2), 214-231.

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

Vyatkina, N. (Ed.). (2015). New developments in the study of L2 writing complexity [Special issue]. *Journal of Second Language Writing, 29.*

Vyatkina, N. (2016). KANDEL: A developmental corpus of learner German. *International Journal of Learner Corpus Research, 2*(1), 102-120.

Vyatkina, N., Hirschmann, H., & Golcher, F. (2015). Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study. *Journal of Second Language Writing, 29*, 28-50.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: National Foreign Language Resource Center.

Yoon, H., & Polio, C. (2017). ESL students' linguistic development in two written genres. *TESOL Quarterly, 51*(2), 275-301.