Journal of Second Language Writing 29 (2015) 16–27

Contents lists available at ScienceDirect

Journal of Second Language Writing

journal homepage: www.elsevier.com/locate/seclan

Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds

Xiaofei Lu[a,*], Haiyang Ai[b]

[a] Department of Applied Linguistics, The Pennsylvania State University, 304 Sparks Building University Park, PA 16802, USA
[b] School of Education, The University of Cincinnati, Teachers College Suite 615, 2610 McMicken Circle, Cincinnati,OH 45221, USA

ARTICLE INFO

Keywords:
College-level English writing
Difference across L1 groups
Syntactic complexity

ABSTRACT

This paper explores differences in the syntactic complexity in English writing among college-level writers with different first language (L1) backgrounds. We sampled 200 argumentative essays written by native-speaker (NS) U.S. university students from the Louvain Corpus of Native English Essays (LOCNESS; Granger, 1996) and 1400 argumentative essays produced by non-native speaker (NNS) English as a Foreign Language (EFL) learners of seven different L1 backgrounds (200 from each L1 background) from the International Corpus of Learner English Version 2.0 (ICLE 2.0; Granger, Dagneaux, Meunier, & Paquot, 2009). These essays were analyzed using 14 syntactic complexity measures with the L2 Syntactic Complexity Analyzer (Lu, 2010). When the EFL learners' L1 backgrounds were ignored, significant differences emerged in only three of the 14 measures between the NNS group and the NS group. However, when the learners were grouped by their L1 backgrounds, significant differences emerged between the NS group and one or more NNS groups in all 14 measures, and the NNS groups demonstrated drastically varied patterns of difference from the NS group. The implications of such varied patterns for L2 writing research and pedagogy and for automatic native language identification of learner texts are considered.

1. Introduction

The importance of syntactic complexity in second language (L2) writing research and pedagogy has long been recognized, as evidenced in the large number of studies that have examined the relationship of syntactic complexity in L2 writing to L2 proficiency (e.g., Ai & Lu, 2013; Lu, 2011; Norrby & Håkansson, 2007; Ortega, 2000, 2003; Stockwell & Harirington, 2003; Vyatkina, 2013; Wolfe-Quintero, Inagaki, & Kim, 1998) or the quality of L2 writing (e.g., Taguchi, Crawford, & Wetzel, 2013; Yang, Lu, & Weigle, 2015) over the past two decades. Results from such studies have shown that some measures of syntactic complexity may be reliably used to differentiate levels of L2 proficiency, and some to predict the quality of L2 writing. Meanwhile, researchers have also found that syntactic complexity in L2 writing may be affected by various learner-, task-, and context-related factors, such as topic, genre, planning time, and instructional setting, among others (e.g., Ellis & Yuan, 2004; Lu, 2011; Ortega, 2003; Sotillo, 2000; Way, Joiner, & Seaman, 2000; Yang et al., 2015). Collectively, these studies have yielded very useful insights into how the construct of syntactic complexity should be conceived and utilized in L2 writing research and pedagogy.

* Corresponding author. Fax: +1 814 863 7986.
E-mail addresses: xxl13@psu.edu (X. Lu), aihaiyang@gmail.com (H. Ai).

http://dx.doi.org/10.1016/j.jslw.2015.06.003
1060-3743/

One specific learner-related factor that has been surprisingly understudied in the large body of research on syntactic complexity in L2 writing is learners' first language (L1). Meanwhile, there is already a growing body of literature on L1-related differences in various aspects of L2 writing (e.g., Carson & Kuehn, 1992; Edelsky, 1982; Jarvis & Crossley, 2012; Lally, 2000; Lefrançois, 2001; Liu, 2008; Paquot, 2013; Rankin, 2012; Uysal, 2008; van Vuuren, 2013; van Weijen, van den Bergh, Rijlaarsdam, & Sanders, 2009). Examples of the aspects examined include idea generation (Lally, 2000), information structure (van Vuuren, 2013), rhetoric patterns (Liu, 2008; Uysal, 2008), syntactic structures (Rankin, 2012), and lexical bundles (Paquot, 2013). In a comprehensive review, Lefrançois (2001) reported that in addition to orthographic and lexical knowledge, aspects of grammatical and syntactic processing, general strategies, and cultural schemata in the L1 could all influence L2 writing. The studies included in a recent volume edited by Jarvis and Crossley (2012) examined language patterns that are characteristic and distinctive of learners from specific L1 backgrounds, using written data from the International Corpus of Learner English Version 2.0 (ICLE 2.0; Granger, Dagneaux, Meunier, & Paquot, 2009). They showed that patterns of lexical style, *n*-grams, errors, cohesion, syntactic complexity, and conceptual knowledge could all play a role in L1 identification.

Insights from previous research into the influence of learners' L1 and other learner-related factors on L2 writing point to the need for close scrutiny of potential L1-related differences in the syntactic complexity in L2 writing. To date, studies that systematically investigate such differences are scarce. The few studies that compared the syntactic complexity in non-native speaker (NNS) and native-speaker (NS) writing (e.g., Ai & Lu, 2013) did not treat learners' L1 background as an independent variable but either looked at a homogeneous L1 group or treated all NNS learners as one group. In one exceptional study, Crossley and McNamara (2012) examined L2 English writing by four L1 groups and reported significant between-group differences in syntactic complexity. However, they looked at one syntactic complexity measure only, i.e., mean number of words before the main verb. As such, our current understanding of L1-related differences in the syntactic complexity in L2 writing is rather limited.

A solid understanding of such differences, however, would have important implications for L2 writing research and pedagogy. For L2 writing research, this understanding will contribute to the growing body of research on L1-related differences in L2 writing and inform researchers whether and how learners' L1 should be controlled or considered in data collection, analysis and interpretation along with other factors in syntactic complexity research. More importantly, it will allow us to re-examine previous claims about the relationship of syntactic complexity in L2 writing to L2 proficiency made without consideration of the potential influence of the L1. For L2 writing pedagogy, awareness of such differences can help L2 writing teachers understand that patterns of syntactic complexity are not necessarily indicative of L2 proficiency in the same way for learners with different L1 backgrounds. Such awareness can help them develop appropriate pedagogical interventions for learners with different L1 backgrounds to address issues related to syntactic complexity more effectively. A systematic understanding of L1-related differences in the syntactic complexity in L2 writing will also contribute useful insight to research on automatic native language identification of learner texts (Jarvis & Crossley, 2012; Tetreault, Blanchard, & Cahill, 2013).

The goal of the current study is to systematically examine differences in the syntactic complexity in English writing among college-level writers with different L1 backgrounds. Our primary focus here, however, is not to establish causal links between specific L1 features and syntactic complexity patterns in L2 writing but to explore whether such differences do exist and if yes, what they are. The results from this investigation will then be discussed in light of previous claims about the relationship of syntactic complexity in L2 writing to L2 proficiency. The implications of our results for L2 writing pedagogy and for automatic native language identification of learner texts will also be considered.

1.1. Syntactic complexity and second language writing

The triad of accuracy, fluency and complexity (CAF) has long been recognized as a principal dimension for gauging L2 proficiency and L2 writing quality (see, e.g., Bulté & Housen, 2012; Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009; Housen, Kuiken, & Vedder, 2012; Norris & Ortega, 2009; Pallotti, 2009; Skehan, 2009; Wolfe-Quintero et al., 1998). The term complexity has been used variably to refer to task, cognitive, or linguistic complexity in the second language acquisition literature (Housen & Kuiken, 2009), and it is its use as linguistic complexity that is most relevant to the discussion of syntactic complexity in the current study. Linguistic complexity may be considered "as a dynamic property of the learner's interlanguage system at large and as a more stable property of the individual linguistic elements that make up the interlanguage system" (Housen & Kuiken, 2009, p. 463). In the latter sense, linguistic complexity has been commonly characterized as "the extent to which language produced in performing a task is elaborate and varied" (Ellis, 2003, p. 340) and has been thought of as consisting of a range of sub-constructs, including lexical, interactional, propositional, and various types of grammatical complexity (Bulté & Housen, 2012; Ellis & Barkhuizen, 2005; see also Ryshina-Pankova, this volume, for a discussion of a meaning-based approach to linguistic complexity in L2 writing). Among these subconstructs, syntactic complexity has perhaps garnered the most attention. In line with the characterization of linguistic complexity, syntactic complexity has been commonly characterized as the range of syntactic structures that are produced and the degree of sophistication of those structures (Ortega, 2003; Pallotti, 2015).

A number of cross-sectional studies have examined the extent to which different syntactic complexity measures reliably index L2 writers' global proficiency (e.g., Ai & Lu, 2013; Bardovi-Harlig & Bofman, 1989; Ferris, 1994; Lu, 2011; Sotillo, 2000) or writing quality (e.g., Taguchi et al., 2013; Yang et al., 2015). For example, Lu (2011) analyzed a large collection of essays

written by Chinese learners of English from different college years using 14 syntactic complexity measures and recommended a subset of those measures as good candidates of proficiency indices. Ai and Lu (2013) reported differences in several dimensions of syntactic complexity between the argumentative writing of NNS students at both low and high proficiency levels and that of NS students, including length of production unit, amount of subordination and coordination, and degree of phrasal sophistication. In a study designed to identify linguistic features that distinguished essays of different quality, Taguchi et al. (2013) analyzed a collection of argumentative essays written by non-native speakers of English and reported that noun phrase modification, along with other measures not related to syntactic complexity, contributed to essay quality.

Several longitudinal studies have also been carried out to track and compare learner development in syntactic complexity over time (e.g., Casanave, 1994; Ishikawa, 1995; Mazgutova & Kormos, this volume; Norrby & Håkansson, 2007; Ortega, 2000, 2003; Stockwell & Harrington, 2003; Vyatkina, 2013; Vyatkina, Hirschmann, & Golcher, this volume). For example, Norrby and Håkansson (2007) tracked adult learners of Swedish over a one-year period and reported that they can be classified into four types: The Careful, The Thorough, The Risk-taker, and The Recycler, based on the relationship between morpho-syntactic development and complexity exhibited in their language production. Specifically, complexity developed hand in hand with grammatical proficiency for The Careful and Thorough, ahead of grammatical proficiency for The Risk-taker, and was avoided despite evidence of grammatical proficiency by The Recycler. Vyatkina (2013) tracked the development of syntactic complexity in the writing of two beginning learners of German over four semesters. She reported that while there was a general developmental trend towards increasing the frequency and range of syntactic complexity features, the two learners diverged from one another in the latter half of the observation period. Taken together, this body of literature revealed that the relationship of syntactic complexity to L2 proficiency or L2 writing quality may vary for different measures or dimensions and that development of syntactic complexity may also vary among learners.

Research has also shown that the syntactic complexity in L2 writing may be affected by various learner-, task- and context-related variables. First, with respect to planning time, learners have been found to produce essay writing with higher syntactic complexity in untimed conditions than in timed conditions (Lu, 2011) and with adequate planning time than with inadequate planning time in narrative essay writing (Ellis & Yuan, 2004) and in computer-mediated communication writing (Sotillo, 2000). Second, with respect to genre and writing tasks, learners have been found to achieve higher syntactic complexity in writing argumentative essays than narrative texts (Lu, 2011), and in descriptive writing tasks than expository tasks (Way et al., 2000). Task complexity, however, has not been found to affect the syntactic complexity in L2 writing (Adams, Newton, & Nik, this volume; Kuiken & Vedder, 2012). Third, with respect to instructional and institutional settings, learners in the English as a Second Language (ESL) instructional setting were found to write with higher syntactic complexity than learners in the English as a Foreign Language (EFL) instructional setting (Ortega, 2003). These findings point to the importance of taking various learner-, task-, and context-related variables into account when analyzing the syntactic complexity of L2 writing. As mentioned above, along with previous findings on L1-related differences in various structural and discourse aspects of L2 writing, these findings motivated us to explore potential L1-related differences in L2 writing syntactic complexity in the current study.

Table 1
Syntactic complexity measures used.

Measure	Code	Definition
Length of production unit		
Mean length of clause	MLC	# of words/# of clauses
Mean length of sentence	MLS	# of words/# of sentences
Mean length of T-unit	MLT	# of words/# of T-units
Amount of subordination		
Clauses per T-unit	C/T	# of clauses/# of T-unit
Complex T-units per T-unit	CT/T	# of complex T-units/# of T-units
Dependent clauses per clause	DC/C	# of dependent clauses/# of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses/# of T-units
Amount of coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases/# of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases/# of T-units
T-units per sentence	T/S	# of T-units/# of sentences
Degree of phrasal sophistication		
Complex nominals per clause	CN/C	# of complex nominals/# of clauses
Complex nominals per T-unit	CN/T	# of complex nominals/# of T-units
Verb phrases per T-unit	VP/T	# of verb phrases/# of T-units
Overall sentence complexity		
Clauses per sentence	C/S	# of clauses/# of sentences

1.2. Measures of syntactic complexity in L2 writing

Syntactic complexity has been measured using a wide range of indices (see, e.g., Bulté & Housen, 2012; Lu, 2011; Norris & Ortega, 2009; Ortega, 2003; Wolfe-Quintero et al., 1998). Probably due to the lack of computational tools for automating syntactic complexity analysis and the labor-intensiveness of manual analysis, most previous studies applied a small number of measures to relatively small amounts of data, with few exceptions. This is problematic for at least two reasons. First, as has been shown in many of the studies discussed above, the relationships of different syntactic complexity measures to L2 proficiency vary substantially. As such, it is difficult to generalize results pertaining to any specific measure to the general construct of syntactic complexity. Second, and more importantly, the construct of syntactic complexity is increasingly thought of as a multi-dimensional construct, with each dimension requiring one or more different measures appropriate for that dimension (Bulté & Housen, 2012; Lu, 2011; Norris & Ortega, 2009). Norris and Ortega (2009), for example, recommended that researchers should at least incorporate measures for global or general complexity, complexity by subordination, complexity via phrasal elaboration, and possibly coordination. While they offered some example measures for some of the key dimensions, they did not propose a comprehensive, directly usable set of measures for all dimensions.

Recognizing the importance to measure syntactic complexity as a multidimensional construct, in the current study we used the full set of 14 measures provided in the L2 Syntactic Complexity Analyzer (L2SCA; Lu, 2010), the computational tool used to analyze the syntactic complexity of the writing samples in our dataset. Each of these measures gauges one of the following five dimensions of syntactic complexity: length of production unit, amount of subordination, amount of coordination, degree of phrasal sophistication, and overall sentence complexity (Ai & Lu, 2013; Lu, 2010, 2011; Ortega, 2003, 2009; Wolfe-Quintero et al., 1998). The measures, their definitions, and the dimensions of syntactic complexity they gauge are summarized in Table 1, adapted from Lu (2010).

The use of this large set of measures is motivated by several considerations. First, the 14 measures incorporated in L2SCA were recommended by Wolfe-Quintero et al.'s (1998) early review of L2 writing measures and still well represent the range of measures in current use in L2 writing research (see, e.g., Lu, 2010, 2011; Ortega, 2003). Second, we are aware that some measures in this set may be stronger indicators of L2 proficiency than others (e.g., Ortega, 2003; Lu, 2011) and some of them may be partially redundant of each other (e.g., Norris & Ortega, 2009). Indeed, the field as a whole is still in search of the best set of non-overlapping measures that are consistently indicative of L2 proficiency or L2 writing quality and that capture all major dimensions of syntactic complexity. As will become clear in Section 5, in this case, the benefit of including a larger set of potentially useful measures outweighs the concern for potentially superfluous information, as it allows us to problematize a number of previous claims and to empirically examine which measures are indeed redundant.

2. Research question

The central research question we seek to answer in the current study is: Are there systematic differences in the syntactic complexity of English writing among college-level writers with different L1 backgrounds and, if yes, what are these differences? Specifically, using a large collection of writing samples produced by native-speaker U.S. university students and college-level EFL learners with seven different L1 backgrounds, we first examine whether and how syntactic complexity differs between NS and NNS college students' English writing; we then investigate whether and how syntactic complexity differs between the NS group and each of the seven NNS groups. Note that the use of the NS group as a comparison point does not imply a deficit orientation of L2 writing wherein any NNS departure from NS usage is construed as a problem. Rather, it serves as a convenience to allow us to discuss L1-related difference in L2 writing syntactic complexity while constraining the large number of pair-wise comparisons that we would otherwise need to make. Based on previous studies into L1-related differences in L2 writing and the effect of learner-related factors on the syntactic complexity in L2 writing, we hypothesize that treating all EFL learners as one homogeneous group would conceal a great deal of variability among learners of different L1 backgrounds and that EFL learners with different L1 backgrounds would show varied patterns of difference in syntactic complexity from the NS group.

3. Method

3.1. Corpus data

The data used in the current study consisted of 1400 English argumentative essays produced by college-level EFL learners with seven different L1 backgrounds (200 essays from each L1 group) and 200 English argumentative essays produced by native-speaker U.S. university students. The NNS data and the NS data were randomly sampled from the International Corpus of Learner English Version 2.0 (ICLE 2.0; Granger et al., 2009) and the Louvain Corpus of Native English Essays (LOCNESS; Granger, 1996), respectively. While not fully controlled for all relevant variables (see Section 5), the two corpora have been considered largely comparable because of their similarity in text type and range of topics (Altenberg & Granger, 2001), and they have been widely used to compare NNS and NS writing or examine L1-related differences in L2 writing (e.g., Aijmer, 2002; Altenberg & Granger, 2001; Crossley & McNamara, 2012; Flowerdew, 2010; O'Donnell, Römer, & Ellis, 2013; Paquot, 2013; Siyanova & Schmitt, 2008).

Table 2
Summary of the dataset.

L1	Essays	Words per essay		Total words	Proficiency ratings and rankings			
		Mean	SD		Ranking	B2 (or lower)	C1	C2
English	200	823.65	468.84	164,729	N/A	N/A	N/A	N/A
German	200	525.43	234.34	105,085	1	1	12	7
Bulgarian	200	666.68	284.00	133,335	2	2	16	2
French	200	598.21	143.52	119,641	3	3	11	6
Russian	200	852.21	393.74	170,441	4	3	15	2
Tswana	200	375.29	131.65	75,057	5	18	0	2
Japanese	200	542.77	95.28	108,553	6	18	2	0
Chinese	200	514.81	106.62	102,961	7	19	1	0
All	1600	612.38	307.80	979,802				

Note: proficiency ratings and rankings of the NNS groups are based on ratings of a random sample of 20 eassys from each NNS group on the basis of the *Common European Framework of Reference for Languages* descriptors for writing (Granger et al., 2009). B2 denotes upper intermediate level; C1 and C2 denote advanced levels. The rankings of the NNS groups are based on the number of essays rated advanced (i.e., C1 or C2) and in the case of a tie, the number of essays rated C2.

ICLE 2.0 is a 3.7-million-word corpus containing argumentative essays written by college-level EFL learners from 16 different L1 backgrounds. The essays were collected from a range of international partner universities under the leadership of Université Catholique de Louvain following the same guidelines.[1] The guidelines specified three data collection stages: students were first requested to fill in a learner profile; they then wrote an essay on a suggested topic (either timed or untimed); the essays were then collected, formatted, and sent to the corpus compilation team. The learner profile contains a number of useful metadata fields.[2] Some of the most relevant ones include essay type, title, writing condition (whether the essay is timed), age, gender, nationality, native language, languages spoken at home, medium of instruction, current studies (discipline), current year of study, current institution, years of English at school and at university, stay in an English-speaking country, and other foreign languages spoken. The corpus collection guidelines include 14 suggested essay topics, such as "Crime does not pay" and "Feminists have done more harm to the cause of women than good." These topics are generally of the same type, i.e., asking students to express an opinion on a statement and explain or argue for that opinion. The guidelines also indicate that essays are expected to be 500–1000 words long and that the target size of the subcorpus for each L1 group is 200,000 words.

LOCNESS is a 324,304-word corpus of native English essays compiled by the Center for English Corpus Linguistics (CECL), Université Catholique de Louvain, specifically as a control corpus for comparing college-level NNS writing in ICLE. This corpus consists of 436 essays produced by fully native speakers of American English (from the United States) or British English (from the UK), although four students indicated that one or both of their parents had a mother tongue other than English. These four students' essays were not included in our NS data. Among these essays, 232 were written by U.S. university students (totaling 168,400 words), 90 by British university students (totaling 95,695 words), and 114 by British A-Level (General Certificate of Education Advanced Level) students (totaling 60,209 words). A small number of the essays written by U.S. and British university students are literary, but the majority are argumentative. Both timed and untimed essays are included.

Care was taken to maximize homogeneity of the NS data and comparability of the NS and NNS data. Specifically, to avoid potential effects of English variety (British vs. American English) and school level (A-Level students vs. university students) in the NS data, we decided to only sample essays produced by U.S. university students from LOCNESS. Furthermore, to avoid potential effects of genre and sample size, we decided to sample the same number of only argumentative essays for each L1 group. Based on these considerations and the number of essays from different L1 groups meeting our selection criteria in the two corpora, we randomly selected 200 argumentative essays produced by U.S. university students from LOCNESS and 200 argumentative essays from ICLE for each of the following seven L1 groups: Bulgarian, Chinese, French, German, Japanese, Russian, and Tswana. The final dataset thus consisted of 1600 essays, with 200 essays for each of the eight L1 groups (one NS group and seven NNS groups) being considered. These L1 groups represent four different language families: Sino-Tibetan (Chinese), Japonic (Japanese), Niger-Congo (Tswana), and Indo-European (Bulgarian, English, French, German, Russian). Within the Indo-European language family, three different subgroups are represented: Germanic (English and German), Romance (French), and Slavic (Russian and Bulgarian). Among these, German and English are closest to each other. Languages from different families and subgroups were selected to help ensure generalizability of the findings from the current study.

It is important to note that, in order to obtain a large enough sample size, not all pertinent variables could be controlled to achieve full homogeneity of the data for each L1 group or full comparability of the data between the NS group and the NNS groups. First, the essays in the subcorpora were on a number of different topics within the argumentative genre, although

[1] The corpus collection guidelines can be found at http://www.uclouvain.be/en-317607.html.
[2] See http://www.uclouvain.be/cps/ucl/doc/cecl/documents/LEARNER_PROFILE.txt.

Table 3
Mean values and standard deviations of 14 syntactic complexity indices of the NNS and NS groups.

Measure	Code	NNS Group Mean (SD)	NS Group Mean (SD)	*t*	*p*
Length of production unit					
Mean length of clause	MLC	9.691 (1.809)	10.092 (1.624)	−2.968	.003*
Mean length of sentence	MLS	19.417 (6.024)	19.602 (4.321)	−.537	.592
Mean length of T-unit	MLT	16.763 (4.371)	17.308 (3.210)	−2.136	.033
Amount of subordination					
Clauses per T-unit	C/T	1.743 (.408)	1.734 (.310)	.296	.767
Complex T-units per T-unit	CT/T	.490 (.151)	.505 (.137)	−1.330	.184
Dependent clauses per clause	DC/C	.392 (.099)	.404 (.091)	−1.637	.102
Dependent clauses per T-unit	DC/T	.716 (.361)	.726 (.283)	−.373	.709
Amount of coordination					
Coordinate phrases per clause	CP/C	.236 (.125)	.253 (.113)	−1.932	.054
Coordinate phrases per T-unit	CP/T	.407 (.224)	.429 (.175)	−1.609	.109
T-units per sentences	T/S	1.153 (.143)	1.132 (.121)	2.250	.025
Degree of phrasal sophistication					
Complex nominals per clause	CN/C	1.131 (.347)	1.222 (.330)	−3.503	.000*
Complex nominals per T-unit	CN/T	1.953 (.691)	2.087 (.565)	−3.031	.003*
Verb phrases per T-unit	VP/T	2.358 (.594)	2.342 (.434)	.475	.635
Overall sentence complexity					
Clauses per sentences	C/S	2.023 (.621)	1.968 (.444)	1.192	.223

Note: *statistical significance at the level $p < .0036$.

Altenberg and Granger (2001) considered the topics covered in ICLE and LOCNESS to have comparable language demands. Second, both timed and untimed essays were included in all subcorpora. Third, while data for the Bulgarian and Russian groups came from one institution, texts for other L1 groups came from two or more institutions (five for the U.S. group). Information about the students' prior writing experience was not available in either ICLE or LOCNESS. As such, the data is inadequate for addressing the effects of such variables or the effects of their interactions with learners' L1 on syntactic complexity. However, following the general practice in previous studies using ICLE and/or LOCNESS (e.g., Altenberg & Granger, 2001; O'Donnell et al., 2013; Paquot, 2013), we expect the large sample size to reduce or diminish the impact of such confounding variables.

A note on the proficiency level of the different NNS groups is also in order. While the aim of the ICLE project was to collect data from advanced learners, Granger et al. (2009) noted that differences existed among the essays both within and across the subcorpora. Based on ratings of a random sample of 20 eassys from each subcorpus on the basis of the Common European Framework of Reference for Languages (CEFR) descriptors for writing[3], they reported that some subcorpora are in the upper intermediate range while others are advanced (see Table 2). Bestgen, Granger, and Thewissen (2012) rated a larger set of essays from the Spanish, German, and French subcorpora and found that essays from the German subcorpus were rated the highest and the French subcorpus was primarily at the advanced level as well. Based on the proficiency ratings provided by Granger et al. (2009) and Bestgen et al. (2012), we will characterize the German, Bulgarian, French and Russian groups as advanced and the Tswana, Japanese and Chinese groups as upper intermediate.

Table 2 summarizes the details of the composition of the final dataset. Although there are considerable differences in mean length of essays among the different L1 groups, these differences do not affect the types of comparisons being pursued in our analysis, as the syntactic complexity indices are all calculated as mean length of production units or ratios of the frequency of one syntactic structure to that of another in individual texts.

3.2. Analytical procedure

Each essay in the final dataset was analyzed using the L2 Syntactic Complexity Analyzer (Lu, 2010), a computer program designed to analyze the syntactic complexity of English writing samples using the 14 measures discussed above. L2SCA was chosen because of its free availability, its integration of a large set of measures that are viable candidates for syntactic complexity research, its capability to process files in batches, and its high reliability. For each writing sample, L2SCA produces frequency counts for the following nine structural units: words, sentences, verb phrases, clauses, dependent clauses, T-units, complex T-units, coordinate phrases, and complex nominals; it also returns 14 indices of syntactic complexity calculated using the frequency counts. Lu (2010) reported accuracies ranging from .830 to 1.000 for structural

[3] These CEFR descriptors for writing do not explicitly mention syntactic complexity.

unit identification and correlations ranging from .834 to 1.000 between the syntactic complexity scores computed by human annotators and L2SCA.

After the syntactic complexity indices have been obtained for each essay in the dataset, a set of independent samples *t* tests were run to compare differences between the NNS group (all EFL learners combined) and the NS group for each of the 14 syntactic complexity measures. One-way ANOVAs were then run to compare the seven EFL groups (based on L1) against the NS group for each of the 14 syntactic complexity measures, followed by post hoc pairwise comparisons using Dunnett's *t*-tests (see Howell, 2006) where appropriate.

4. Results

4.1. Differences between the combined NNS group and the NS group

In the first part of the analysis, we aimed to determine whether significant differences existed in the syntactic complexity in NNS and NS university students' English writing and, if yes, what they were. To this end, the writing samples produced by NNS writers were collapsed into one group and compared against the writing samples produced by NS writers. As Table 3 shows, the NNS group had lower mean values than the NS group for 10 out of the 14 syntactic complexity measures. Independent-samples *t* tests were run to determine which differences between the NNS and NS groups were statistically significant. Given that we were conducting 14 tests on the same dataset simultaneously, we employed the Bonferroni correction to control the familywise error rate and adjusted the alpha value for each comparison to .05/14, or .0036, where .05 is the significance level for the complete set of tests, and 14 is the number of tests performed.

The results suggest that the two groups differed significantly in one length of production unit measure (mean length of clause) and two phrasal sophistication measures (complex nominals per clause and per T-unit). No significant differences were found for other measures. This finding is in line with those reported by Ai and Lu (2013), who found significant differences in these measures between NNS and NS writings as well. However, they also reported significant differences between NNS and NS writings in mean length of sentence and T-unit, dependent clauses per clause and per T-unit, and coordinate phrases per T-unit. This discrepancy might be attributed to the fact that the NNS data in Ai and Lu's (2013) study consisted of writing samples produced by English learners of L1 Chinese background only, whereas in the current study the NNS data were produced by English learners of seven different L1 backgrounds. As will become clear in the second part of the analysis below, pooling data from learners with heterogeneous L1 backgrounds may conceal differences related to the individual L1s.

4.2. Differences between the seven NNS groups and the NS group

In the second part of the analysis, we compared each of the seven NNS groups against the NS group. One-way ANOVAs revealed significant between-group differences in all 14 syntactic complexity measures. In what follows, we discuss the between-group differences found in each of the following five dimensions of syntactic complexity: length of production units, amount of subordination, amount of coordination, degree of phrasal sophistication, and overall sentence complexity. We specifically focus on the significant differences found between the NS group and one or more of the NNS groups, as established by Dunnett's *t*-tests. Note that in Table 4 through Table 7, the L1 groups are ordered in ascending manner by their mean values of each measure.

4.2.1. Length of production unit

Table 4 summarizes the means and standard deviations of the three length measures as well as the significant differences between the NS group and one or more of the NNS groups in these measures. The results show that the NS group did not consistently produce longer production units than all NNS groups. Rather, the seven NNS groups compared very differently

Table 4
Mean values and standard deviations of MLS, MLC, and MLT across different L1 groups.

MLS			MLC			MLT		
L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.
Japanese	14.672 (3.574)	–	Japanese	8.307 (1.217)	–	Japanese	13.115 (2.574)	–
Chinese	17.636 (3.220)	–	Tswana	8.636 (1.429)	–	Russian	15.984 (2.585)	–
Russian	18.742 (5.160)		Russian	9.657 (1.637)	–	Chinese	16.153 (4.111)	–
English	19.602 (4.321)		Bulgarian	9.935 (1.563)		English	17.308 (3.210)	
Bulgarian	20.523 (4.622)		English	10.092 (1.624)		French	17.539 (4.121)	
French	20.809 (5.611)		Chinese	10.171 (1.482)		Bulgarian	17.606 (3.661)	
Tswana	21.225 (8.955)	+	French	10.508 (1.492)		Tswana	17.961 (5.515)	
German	22.309 (5.584)	+	German	10.624 (1.984)	+	German	18.982 (4.612)	+

Note: see Table 1 for definitions of MLS, MLC and MLT. One-way ANOVAs revealed significant between-group differences for MLS ($F(7, 1592) = 40.255$, $p < .001$), MLC ($F(7, 1592) = 55.032$, $p < .001$) and MLT ($F(7, 1592) = 41.424$, $p < .001$). – and + denote a value that is significantly lower or higher than the value for the U.S. group ($p < .05$), respectively, as determined by Dunnett's *t*-tests (2-sided).

to the NS group on the three length measures. While some groups (i.e., German and Japanese) showed significantly positive or negative differences from the NS group across all three measures, some groups differed significantly from the NS groups on some measures but not others (i.e., Chinese, Russian, and Tswana), and still others (i.e., Bulgarian and French) did not differ significantly from the NS group on any of the three measures.

4.2.2. Amount of subordination

Table 5 summarizes the means and standard deviations of the four subordination measures and the significant differences between the NS group and one or more of the NNS groups in these measures. These results are in sharp contrast to those reported in the first part of the analysis, in which we found no significant difference in the subordination measures between the NNS and NS groups. In other words, once L1 was factored in, we observed substantial variations with respect to the amount of subordination produced across different NNS groups. While the Chinese, Japanese, and Russian groups produced significantly less subordination than the NS group, the Tswana group produced significantly more subordination than the NS group.

4.2.3. Amount of coordination

Table 6 summarizes the means and standard deviations of the three coordination measures as well as significant differences between the NS group and one or more of the NNS groups in these measures. These results again contrast with the findings from the first part of the analysis, which showed no significant difference between the NNS and NS groups in the coordination measures. Substantial variation was observed in the amount of coordination produced by learners of different L1 backgrounds, and every NNS group differed significantly from the NS group in at least one coordination measure.

4.2.4. Phrasal and overall sentence complexity

Table 7 summarizes the means and standard deviations of the phrasal and overal sentence complexity measures as well as significant differences between the NS group and one or more NNS groups in these measures. Consistent with the results for other dimensions of syntactic complexity, variation abounds among the NNS groups with different L1 backgrounds in the dimensions of phrasal and overall sentence complexity. Two observations are especially worth noting. The Japanese group continued to show significantly lower syntactic complexity than the NS group. The Tswana group continued to show significantly higher levels of production of verb phrases and clauses than the NS group but significantly lower usage of complex phrases than the NS group.

5. Discussion

This study has examined differences in the syntactic complexity in college-level English writing among writers with different L1 backgrounds. When all NNS writers with heterogeneous L1 backgrounds were treated as a single group, they differed significantly from the NS group in only three out of the 14 syntactic complexity measures. These results differed substantially from those reported by Ai and Lu (2013), which showed significant differences between a homogeneous NNS group (i.e., Chinese learners of English) and an NS group in several additional measures. This difference constitutes a first piece of evidence that combining NNS speakers with heterogeneous L1 backgrounds may conceal L1-related differences in the syntactic complexity in L2 writing.

When the seven NNS groups with different L1 backgrounds were compared with the NS group individually, significant differences emerged between the NS group and one or more NNS groups in all 14 measures, and the NNS groups showed drastically varied patterns of difference from the NS group. As mentioned in Section 1, L1-related differences in various structural and discourse aspects of L2 writing have been well documented in the L2 writing literature (e.g., Carson & Kuehn, 1992; Edelsky, 1982; Jarvis & Crossley, 2012; Lally, 2000; Lefrancois, 2001; Liu, 2008; Rankin, 2012; Uysal, 2008; van Vuuren,

Table 5
Mean values and standard deviations of C/T, CT/T, DC/C, and DC/T measurements across different L1 groups.

C/T			CT/T			DC/C			DC/T		
L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.
Japanese	1.583 (.238)	–	Chinese	.436 (.127)	–	Chinese	.345 (.078)	–	Chinese	.567 (.192)	–
Chinese	1.600 (.227)	–	Japanese	.440 (.136)	–	Japanese	.355 (.088)	–	Japanese	.580 (.224)	–
Russian	1.653 (.280)		Russian	.446 (.143)	–	Russian	.363 (.094)	–	Russian	.623 (.254)	–
French	1.678 (.311)		French	.481 (.136)		French	.387 (.091)		French	.673 (.285)	
English	1.734 (.310)		English	.505 (.137)		English	.404 (.091)		English	.726 (.283)	
Bulgarian	1.779 (.294)		Bulgarian	.519 (.133)		Bulgarian	.413 (.079)		Bulgarian	.755 (.252)	
German	1.802 (.369)		German	.530 (.153)		German	.425 (.096)		German	.795 (.332)	
Tswana	2.108 (.685)	+	Tswana	.576 (.168)	+	Tswana	.456 (.110)	+	Tswana	1.019 (.599)	+

Note: see Table 1 for definitions of C/T, CT/T, DC/C, and DC/T. One-way ANOVAs revealed significantly between-group differences for C/T ($F(7, 1592) = 41.891$, $p < .001$), CT/T ($F(7, 1592) = 24.830$, $p < .001$), DC/C ($F(7, 1592) = 34.937$, $p < .001$) and DC/T ($F(7, 1592) = 40.735$, $p < .001$). – and + denote a value that is significantly lower or higher than the value for the U.S. group ($p < .05$), respectively, as determined by Dunnett's t-tests (2-sided).

Table 6
Mean values and standard deviations of CP/C, CP/T, and T/S measurements across different L1 groups.

CP/C			CP/T			T/S		
L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.
Japanese	.168 (.087)	–	Japanese	.263 (.139)	–	Chinese	1.092 (.095)	–
Tswana	.185 (.117)	–	Chinese	.360 (.142)	–	Japanese	1.115 (.127)	
Chinese	.229 (.093)		Tswana	.380 (.236)		English	1.132 (.121)	
French	.246 (.122)		French	.410 (.220)		Tswana	1.163 (.189)	
Russian	.252 (.118)		Russian	.419 (.221)		Bulgarian	1.167 (.121)	
English	.253 (.113)		English	.429 (.175)		Russian	1.171 (.122)	+
German	.274 (.132)		German	.486 (.230)	+	German	1.180 (.146)	+
Bulgarian	.301 (.142)	+	Bulgarian	.528 (.249)	+	French	1.186 (.156)	+

Note: see Table 1 for definitions of CP/C, CP/T and T/S. One-way ANOVAs revealed significantly between-group differences for CP/C ($F(7, 1592) = 28.534$, $p < .001$), CP/T ($F(7, 1592) = 30.356$, $p < .001$) and T/S ($F(7, 1592) = 12.114$, $p < .001$). – and + denote a value that is significantly lower or higher than the value for the U.S. group ($p < .05$), respectively, as determined by Dunnett's *t*-tests (2-sided).

Table 7
Mean values and standard deviations of CN/C, CN/T, VP/T, and C/S measurements across different L1 groups.

CN/C			CN/T			VP/T			C/S		
L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.	L1	Mean (SD)	Sig.
Tswana	.913 (.272)	–	Japanese	1.499 (.477)	–	Japanese	2.075 (.396)	–	Chinese	1.748 (.301)	–
Japanese	.944 (.249)	–	Tswana	1.914 (.803)	–	Russian	2.189 (.415)	–	Japanese	1.768 (.349)	–
Bulgarian	1.161 (.308)		Russian	1.934 (.710)		Chinese	2.194 (.351)	–	Russian	1.943 (.420)	
Russian	1.162 (.339)		Chinese	2.010 (.545)		French	2.284 (.472)		English	1.968 (.444)	
German	1.188 (.372)		Bulgarian	2.058 (.622)		English	2.342 (.434)		French	1.994 (.470)	
English	1.222 (.330)		English	2.087 (.565)		Bulgarian	2.369 (.411)		Bulgarian	2.080 (.429)	
Chinese	1.265 (.325)		German	2.117 (.713)		German	2.536 (.579)	+	German	2.129 (.513)	+
French	1.281 (.357)		French	2.140 (.709)		Tswana	2.862 (.934)	+	Tswana	2.496 (1.124)	+

Note: see Table 1 for definitions of CN/C, CN/T, VP/T and C/S. One-way ANOVAs revealed significant between-group differences for CN/C ($F(7, 1592) = 37.440$, $p < .001$), CN/T ($F(7, 1592) = 20.213$, $p < .001$), VP/T, ($F(7, 1592) = 40.635$, $p < .001$), and C/S ($F(7, 1592) = 35.260$, $p < .001$). – and + denote a value that is significantly lower or higher than the value for the U.S. group ($p < .05$), respectively, as determined by Dunnett's *t*-tests (2-sided).

2013; van Weijen et al., 2009). The varied patterns among the three upper intermediate NNS groups (Chinese, Japanese, and Tswana) and those among the four advanced NNS groups (Bulgarian, French, German, and Russian) suggest that the intergroup variation in syntactic complexity cannot be accounted for by proficiency alone but that learners' L1 may play a role in the syntactic complexity in their L2 writing as well.

Previous research has made a number of claims about the relationship of syntactic complexity to L2 proficiency and about potential trade-offs between different dimensions of syntactic complexity as learners' proficiency levels advance. Specifically, coordination and subordination are considered to be the most indicative and useful indices of complexification at beginning and intermediate levels, respectively, while at the advanced level, the role of subordination is subdued as phrasal-level complexification becomes the most pervasive means of syntactic complexity (Bardovi-Harlig, 1992; Norris & Ortega, 2009; Ortega, 2003). Longer production units have also been found to correlate with higher levels of proficiency (e.g., Lu, 2011; Ortega, 2003; Wolfe-Quintero et al., 1998). In what follows, we discuss some of the patterns

Table 8
Summary of significant differences between the NNS groups and the NS group in all 14 syntactic complexity measures.

L1	Length of production unit			Amount of subordination				Amount of coordination			Degree of phrasal sophistication			Sentence complexity
	MLS	MLC	MLT	C/T	CT/T	DC/C	DC/T	CP/C	CP/T	T/S	CN/C	CN/T	VP/T	C/S
Japanese	–	–	–	–	–	–	–	–	–		–	–	–	–
Chinese	–		–	–	–	–	–		–	–			–	–
Russian		–	–		–	–	–			+			–	
Tswana	+	–		+	+	+	+	–			–	–	+	+
French										+				
Bulgarian								+	+					
German	+	+	+						+	+			+	+

Note. See Table 1 for definitions of the 14 measures. – denotes a value that is significantly lower than the value for the U.S. group ($p < .05$) and + denotes a value that is significantly higher than the value for the U.S. group ($p < .05$), as determined by Dunnett's *t*-tests (2-sided).

observed in the current study in light of such previous claims and findings. To facilitate this discussion, we have summarized the significant differences found in all 14 syntactic complexity measures between the NS group and the NNS groups in Table 8, with the NNS groups arranged in order of decreasing negative difference followed by increasing positive difference from the NS group.

Consider the results on sentential coordination in light of the previous claim that greater use of coordinate sentences is indicative of beginning or lower levels of L2 proficiency (Bardovi-Harlig, 1992; Norris & Ortega, 2009). Our results revealed, however, that the French, German, and Russian groups (rated at the advanced level) all produced significantly more sentential coordination than the NS group, while this was not the case for any of the three upper intermediate groups (Chinese, Japanese, and Tswana). In fact, despite being the lowest ranked group, the Chinese group was the only group that produced significantly less sentential coordination than the NS group. This observation coincides with the fact that in Chinese, coordination of two independent clauses can be indicated by punctuation without using a coordinating conjunction (Li & Thompson, 1981). Further research is needed to confirm whether this feature of Chinese is directly linked to the lower usage of sentential coordination by L1 Chinese learners of English. However, taken together, our results strongly suggest that previous claims about the relationship of sentential coordination to L2 proficiency may need to be reexamined by taking into account the potential influence of the L1. At the very least, the interaction between learners' L1 and proficiency needs to be considered.

Now consider the results on subordination and phrasal sophistication (as indicated by the complex nominal measures). Previous claims about their relations to L2 proficiency would predict that the upper intermediate groups (Chinese, Japanese, and Tswana) should use comparable and possibly more subordination but fewer complex nominals than the NS group, while the advanced groups (Bulgarian, French, German, and Russian) should exhibit comparable usage of both subordination and complex nominals as the NS group. In our results, one upper intermediate group (Tswana) and three advanced groups (Bulgarian, French, and German) appear to meet this prediction perfectly. However, one advanced group (Russian) and two upper intermediate groups (Chinese and Japanese) all diverge from this prediction. The Russian group has achieved the same level of phrasal complexity as the other advanced NNS groups and the NS group but has also a significantly lower amount of subordination than those groups. Given that the Russian group is ranked the lowest among the four advanced levels, the cause of this additional reduction in clausal complexity cannot be readily attributed to proficiency but deserves further investigation. More interestingly, the Chinese group exhibited the same pattern as the Russian group. This goes against the proficiency-based prediction, as clausal complexity seems to lag behind phrasal complexity for the Chinese group, instead of the other way round. Finally, the Japanese group differed from both of the other two upper intermediate groups in that it used both less subordination and fewer complex nominals than the NS group, suggesting that it has not fully developed either clausal or phrasal complexity. This again problematizes a purely proficiency-based explanation, as the Japanese group has already advanced to the upper intermediate level based on the sample essay ratings.

Next, consider the results on the length measures for the advanced groups. As mentioned above, previous research has shown that longer production units are generally indicative of a higher proficiency level (e.g., Lu, 2011; Ortega, 2003; Wolfe-Quintero et al., 1998). However, Pallotti (2009) also argued that, "[i]f sociopragmatic adequacy were taken into account, for example, by looking at native speakers' baseline data, it would appear that after a certain point a decrease in syntactic complexity might be interpreted as a sign of higher proficiency" (pp. 598–599). Together, these claims would predict that, on average, the advanced NNS groups should produce sentences, clauses, and T-units of more or less comparable length as the NS group. This was indeed the case for the French and Bulgarian groups. The Russian group, however, produced significantly shorter clauses and T-units than the NS group, while the German group produced significantly longer sentences, clauses, and T-units than the NS group. The case of the German group is especially interesting. Given that it represents the highest proficiency level among the seven NNS groups, this high level of length of production units appears to be excessive in comparison to both the NS group and the other advanced groups. In fact, as shown in Table 8, the German group showed a higher level of complexity than the NS group in several other measures as well. The high use of both sentential and phrasal coordination may have contributed to the longer production units. While additional research is needed to confirm an L1 transfer effect, we note that the German group's significantly higher complexity in length of production unit coincides with "a well-known fact . . . that German sentences tend to be longer than English sentences" (Ziegler, 1991, p. 147).

It appears that the typological similarities and differences between the seven non-English L1s and English may have some influence on the syntactic complexity patterns observed for the seven NNS groups, but such a claim is not straightforward and at best tentative. Although the ICLE subcorpora were all sampled from university undergraduate students, the four advanced groups (Bulgarian, French, German, and Russian) all belong to the same language family as English (i.e., Indo-European), while the three upper intermediate groups belong to three different language families. The four advanced groups also exhibited closer patterns to the NS group than the upper intermediate groups, but it is possible that proficiency played a role here. Among the four advanced groups, however, the groups that were most similar to the NS group were the French and Bulgarian groups instead of the German group, despite the fact that German is typologically the closest to English and that the German group supposedly had the highest proficiency level among all of the NNS groups.

Finally, our results appear to confirm Norris and Ortega's (2009) claim that some syntactic complexity measures might be highly correlated and thus redundant of each other, a claim discussed earlier in Section 1. Specifically, the following three

subordination measures: complex T-units per T-unit, dependent clauses per clause, and dependent clauses per T-unit demonstrated identical patterns across all L1 groups. Additionally, two of the phrasal complexity measures, i.e., complex nominals per clause and per T-unit, also showed identical patterns across all L1 groups. As such, it may be adequate to consider only one of these subordination and phrasal complexity measures in future research.

6. Conclusions

This study revealed significant and varied patterns of differences in multiple dimensions of syntactic complexity in college-level English writing among writers with different L1 backgrounds, although further research is needed to establish any causal links between specific L1 features and syntactic complexity patterns in L2 writing. These results call for the need to take learners' L1 into account in assessing the relationship of syntactic complexity to L2 proficiency.

In addition to the implications for L2 writing research discussed above, our findings have useful implications for L2 writing pedagogy as well. Our results show that in terms of syntactic complexity development, learners with different L1 backgrounds, even for those at the same or comparable proficiency levels, may not develop in the same ways in all areas. For example, among the upper intermediate groups, the Japanese group appears to lag behind other groups in all areas other than sentential coordination, the Tswana group shows a higher level of clausal complexity than phrasal complexity, while the Chinese group shows a higher level of phrasal complexity than clausal complexity. Awareness of and attention to the patterns revealed here can help L2 writing teachers understand differences in syntactic complexity that may surface in writing among their students. Our results also show that advanced writers from certain L1 backgrounds may be more prone to produce longer and more complex sentences in English compared to writers from other L1 backgrounds (including English L1), and it may be useful for writing teachers to be aware of this propensity, should it lead to writing that may be considered problematic in some cases.

Finally, the findings reported here also have practical implications for research on automatic native language identification of learner text, a growing subfield of natural language processing (Jarvis & Crossley, 2012; Tetreault et al., 2013). Previous research in this area has attempted to identify language features that are useful for distinguishing learners with different L1 backgrounds. The significant differences in syntactic complexity found among the NNS groups can be used along with those in other language features to improve the performance of native language identification systems. Among other applications, such systems can be integrated in educational applications that aim to provide tailored assistance to learners with different L1 backgrounds.

Before we conclude, it is important to acknowledge several limitations of the current study, some of which may be addressed in future research. First, we did not attempt to establish any causal links between specific L1 features and the syntactic complexity patterns observed in the current study. Future research could examine the existence and nature of such links. Second, as discussed in Section 3, we relied on large samples randomly selected from the subcorpora of ICLE to alleviate the potential effects of various learner- and task-related factors that could not be controlled. It would be desirable to use more homogeneous datasets in future research to completely eliminate such effects. Third, the characterization of the proficiency levels of the NNS groups was based on ratings of a relatively small sample of 20 essays from each group. Furthermore, while the CERF writing descriptors used to rate the essays do not explicitly mention syntactic complexity, it is unclear as to what extent syntactic complexity was taken into account subconsciously by the raters. In the absence of an independent measure of L2 proficiency, there is a risk of co-dependence between the proficiency ratings and the syntactic complexity measures given that they were derived from the same writing samples. Future research could validate our interpretation of the results using datasets in which multiple NNS groups have been more reliably differentiated using an independent measure of L2 proficiency. Finally, to obtain a more complete picture of the effect of learners' L1 on syntactic complexity development, it would be desirable to track learners' syntactic complexity development over time and systematically compare the trajectories of development across learners of different L1 backgrounds.

References

Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in nns and ns university students writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249–264). Amsterdam: John Benjamins.

Aijmer, K. (2002). Modality in advanced swedish learners' written interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55–76). Amsterdam: John Benjamins.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of make in native and non-native student writing. *Applied Linguistics*, *22*, 173–195.

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, *11*, 17–34.

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, *26*, 390–395.

Bestgen, Y., Granger, S., & Thewissen, J. (2012). Error patterns and automatic L1 identification. In S. Jarvis, & S.A. Crossley (Eds.), *Approaching language transfer through text classification: Explorations in the detection-based approach* (pp. 127–153). Bristol: Multilingual Matters.

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.

Carson, J. E., & Kuehn, P. A. (1992). Evidence of transfer and loss in developing second language writers. *Language Learning*, *42*, 157–179.

Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing*, *3*, 179–201.

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.

Edelsky, C. (1982). Writing in a bilingual program: The relation of L1 and L2 texts. *TESOL Quarterly*, *16*, 211–228.

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.
Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*, 59–84.
Ferris, D. R. (1994). Rhetorical strategies in student persuasive writing: Differences between native and non-native English speakers. *Research in the Teaching of English*, *28*, 45–65.
Flowerdew, J. (2010). Use of signaling nouns across L1 and L2 writer corpora. *International Journal of Corpus Linguistics*, *15*, 36–55.
Granger, S. (1996). From ca to cia and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast: Paper from a symposium on text-based cross-linguistic studies. Lund studies in English:* (Vol. 88. pp. 37–51). Lund: Lund University Press.
Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International corpus of learner English (Version 2.0)*. Louvain-la-Neuve: Presses universitaires de Louvain.
Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*, 461–473.
In A. Housen, F. Kuiken, & I. Vedder (Eds.), (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.
Howell, D. C. (2006). *Statistical methods for psychology*. Belmont, CA: Wadsworth Publishing.
Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, *4*, 51–69.
In S. Jarvis, & S.A. Crossley (Eds.), (2012). *Approaching language transfer through text classification: explorations in the detection-based approach*. Bristol: Multilingual Matters.
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.
Lally, C. G. (2000). First language influences in second language composition: The effect of pre-writing. *Foreign Language Annals*, *33*, 428–432.
Lefrançois, P. (2001). Le Point sur les transferts dans l'ecriture en langue seconde [Transfer in second-language writing]. *Canadian Modern Language Review*, *58*, 223–245.
Li, C. N., & Thompson, S. A. (1981). *Mandarin chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
Liu, J. (2008). The generic and rhetorical structures of expositions in English by Chinese ethnic minorities: A perspective from intracultural contrastive rhetoric. *Language and Intercultural Communication*, *8*, 2–20.
Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*, 474–496.
Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*, 36–62.
Norrby, C., & Håkansson, G. (2007). The interaction of complexity and grammatical processability: The case of Swedish as a foreign language. *IRAL—International Review of Applied Linguistics in Language Teaching*, *45*, 45–68.
Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*, 555–578.
O'Donnell, M. B., Römer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing: Investigating effects of frequency, association, and native norm. *International Journal of Corpus Linguistics*, *18*, 83–108.
Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners. Unpublished doctoral dissertation*. Honolulu: University of Hawaii.
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. (2009). Studying writing across efl contexts: Looking back and moving forward. In R.M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 232–255). Clevedon, UK: Multilingual Matters.
Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*, 590–601.
Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, *31*, 117–134.
Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics*, *18*, 391–417.
Rankin, T. (2012). The transfer of V2: Inversion and negation in German and Dutch learners of English. *International Journal of Bilingualism*, *16*, 139–158.
Siyanova, A., & Schmitt, N. (2008). L2 Learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*, 429–458.
Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, *30*, 510–532.
Sotillo, S. M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning and Technology*, *4*, 82–119.
Stockwell, G., & Harrington, M. (2003). The incidental development of L2 proficiency in NS-NNS email interactions. *CALICO Journal*, *20*, 337–359.
Taguchi, N., Crawford, B., & 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.
Tetreault, J., Blanchard, D., & Cahill, A. (2013). A report on the first native language identification shared task. *Proceedings of the eighth workshop on innovative use of NLP for building educational applications* (pp. 48–57)..
Uysal, H. H. (2008). Tracing the culture behind writing: Rhetorical patterns and bidirectional transfer in L1 and L2 essays of Turkish writers in relation to educational context. *Journal of Second Language Writing*, *17*, 183–207.
van Vuuren, S. (2013). Information structural transfer in advanced Dutch EFL writing: A cross-linguistic longitudinal study. *Linguistics in the Netherlands*, *30*, 173–187.
van Weijen, D., van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2009). L1 use during L2 writing: An empirical study of a complex phenomenon. *Journal of Second Language Writing*, *18*, 235–250.
Vyatkina, N. (2013). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, *97*, 11–30.
Way, D. P., Joiner, E. G., & Seaman, M. A. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *The Modern Language Journal*, *84*, 171–184.
Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.
Yang, W., Lu, X., & Weigle, S. A. (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.
Ziegler, H. (1991). On translating the sunday drive. *The Review of Contemporary Fiction*, *11*, 137–149.

Xiaofei Lu is Gil Watz Early Career Professor in Language and Linguistics and Associate Professor of Applied Linguistics and Asian Studies at The Pennsylvania State University. His research interests are primarily in computational linguistics, corpus linguistics, and intelligent computer-assisted language learning. He is the author of *Computational Methods for Corpus Annotation and Analysis* (2014, Springer).

Haiyang Ai is an Assistant Professor at the University of Cincinnati. His research interests include corpus linguistics, computational linguistics, intelligent computer-assisted language learning, and second language writing.