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

A B S T R A C T

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.

©2015 Elsevier Inc. All rights reserved.

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).

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