Complexity in Writing Development: Untangling Two Approaches to Measuring Grammatical Complexity

CL2017 Pre-conference workshop 2
University of Birmingham, Monday 24th July 2017, 09:30 – 16:30

Workshop convenors

Bethany Gray
Iowa State University
begray@iastate.edu

Shelley Staples
University of Arizona
slstaples@email.arizona.edu

Jesse Egbert
Northern Arizona University
jesse.egbert@nau.edu

Workshop summary

1. Introduction

Linguistic complexity is an often-investigated topic with respect to language development, for both L1 and L2 speakers and writers. The assumption underlying this pairing of complexity and language development is that higher proficiency speakers/writers (in the case of L2) or more advanced speakers/writers (in the case of L1) use more complex language and produce more complex texts. But what constitutes complex language? How do we define complexity? How do we operationalize complexity in order to measure it in language production? How are different approaches to complexity similar and different? How is complexity mediated by proficiency or level, register or genre, and other contextual factors?

This workshop principally focuses on these questions for one type of complexity[1] – grammatical complexity – as it relates to writing development in L1 and L2 writers. The goal of the workshop is to explore how fundamentally distinct measures approach the same underlying construct, to gain a fuller understanding of what the measures capture about language production by carrying out complexity analysis on authentic texts, and to see a selection of these variables applied in research on writing development.

The workshop consists of three parts, each of which is described below:

- *Part 1: Hands-On, Practice-Oriented Session in Coding and Compiling Complexity Variables*
- *Part 2: Research Synthesis on the Development of Complexity in Academic Writing*
- *Part 3: Roundtable Discussion*

---

[1] We return to a discussion of other types of complexity, such as lexical and pragmatic complexity, in Part 3 of the workshop.

2. Approaches to Grammatical Complexity

Grammatical complexity has garnered a great deal of attention in writing development research because it viewed as "an index of language development and progress" (Bulté & Housen, 2014, p. 43; Crossley & McNamara, 2014; Ortega, 2003). In this workshop, we explore two of the major recent approaches to measuring grammatical complexity in writing development research. The first approach relies on ratio- or mean-based measures such as mean length of T-unit, mean number of clauses per T-unit, and mean number of complex nominals, among others, to operationalize grammatical complexity. This approach relies extensively on Hunt's (1966) work on subordination indices. This approach has a long history in SLA research (see Wolfe-Quintero et al., 1998; Ortega, 2003; Norris & Ortega, 2009), and has also been utilized in corpus-based and computational research on language development. The main focus in these studies has been on the amount of subordination (and less commonly coordination) in a text. In recent years, scholars using these T-unit based measures have begun to include other ratio- and mean-based measures to capture complex nominal structures (Ortega, 2015), a point which we return to a bit later. A defining characteristic of this approach is that the measures of complexity are holistic indices, collapsing multiple grammatical features into a single value in order to characterize development in a concise way (see discussion in Biber, Gray & Staples, 2016). We refer to this approach as the holistic (T-Unit) approach.

The second approach has developed more recently out of corpus-based register variation studies (Biber, 1988, 1992; Biber & Gray, 2010; Biber, Gray & Poonpon, 2011), and relies on standardized rates of occurrence for specific grammatical structures (e.g., rate of finite complement clauses per 1,000 words, rate of attributive adjectives per 1,000 words) to operationalize grammatical complexity. This approach is based on register variation studies that have demonstrated the distinct grammatical styles of different registers, and argues for lexico-grammatical measures that can capture the structures that are particularly prevalent in the target register being learned/developed. For academic writing development, this means accounting for more measures that capture complexity in the noun phrase in addition to measures that capture complexity at the clausal level (see Biber, Gray & Poonpon, 2011, 2013). This approach differs from the first in that it retains distinctions between grammatical features, and pays explicit attention to the functions and register distributions of each feature. The approach results in separate values for each grammatical feature of interest (some of which are associated with clausal complexity and others associated with phrasal complexity).[2] We thus refer to this approach as the register/functional approach (Biber, Gray & Staples, 2016).

Although both approaches target the underlying construct of grammatical complexity, they are fundamentally different in terms of a number of criteria:
- the nature of quantitative data: ratio- or mean-based measures versus rates of occurrence
- the role of specific grammatical structures: single, holistic variables versus rates of occurrence for multiple specific grammatical structures
- the definition of subordination and of particular grammatical features (e.g., whether or not non-finite clauses are used in operationalizing the extent of subordination)
- the amount of weight given to phrasal versus clausal complexity
- the functional interpretability of the variables across registers and other contexts of language use

It is important to understand what these fundamental differences are, and how these differences impact the results of cross-sectional and longitudinal writing development studies. That is, as consumers of writing development research, we need to understand what each is capturing

---

[2] Biber, Gray & Staples (2016) propose multi-dimensional analysis (Biber, 1988) as a method of consolidating these multiple variables into a more parsimonious model of language production while maintaining grammatical distinctions, a point which we return to in Part 2 of the workshop.

about the nature of the language production to make sense of results of individual studies, and to draw parallels between studies using different measures. As researchers, we need to understand these same issues in order to select an appropriate framework for our research goals and carry out the corresponding methods. The three parts of this workshop work toward increasing participants' understanding of approaches to complexity through hands-on analysis, a research synthesis, and a roundtable discussion.

Part 1: Hands-On, Practice-Oriented Session in Coding and Compiling Complexity Variables

Part 1 is a hands-on, practice-based session in which participants will use manual and automatic corpus tools to analyze authentic texts for a range of complexity measures from both the holistic (T-Unit) and register/functional traditions. The goal of Part 1 of the workshop is to untangle the two approaches and gain a fuller understanding of what the various measures used in these traditions represents linguistically—that is, what each measure tells us about the nature or structure of the language that is being analyzed.

Table 1 summarizes some of the most commonly used measures of clausal and phrasal complexity in the holistic (T-unit) tradition (Norris & Ortega, 2009; Lu, 2011; Bulté & Housen, 2014; Crossley & McNamara, 2014) and in the register/functional tradition (Biber, Gray & Poonpon, 2011; Biber, Gray & Staples, 2016; Staples, Egbert, Biber, & Gray, 2016; Parkinson & Musgrave, 2014).

Table 1. Summary of Common Grammatical Complexity Measures

	Holistic (T-Unit) Tradition	**Register/Functional Tradition**
Clausal Complexity	1. Mean length of T-Unit 2. Mean length of clause 3. Mean number of clauses per T-Unit 4. Mean number of dependent clauses per total clauses	*all are normalized rates of occurrences per X words (usually 100 or 1,000 words depending on the length of texts) 1. Finite adverbial clauses 2. Finite complement clauses (that- and wh-clauses controlled by verbs, adjectives, and nouns) 3. Finite relative clauses 4. Non-finite adverbial clauses 5. Non-finite complement clauses (controlled by verbs, adjectives, and nouns) 6. Non-finite relative clauses 7. Clausal coordination with and/or
Phrasal Complexity	1. Number of complex nominals per T-unit 2. Mean length of noun phrase 3. Mean number of words before main verb	1. Adjectives as nominal pre-modifiers 2. Nouns as nominal pre-modifiers 3. Appositive noun phrases as nominal post-modifiers 4. Prepositional phrases as nominal post-modifiers 5. Nominalizations

In order to develop a concrete understanding of the differing nature of quantitative data (ratio-based versus normalized rate of occurrence) and the differing definitions of grammar structures (e.g., role of non-finite clauses), participants will analyze a small sample of texts from the British Academic Written English (BAWE) corpus (Nesi, Gardner, Thompson, & Wickens, 2008-2010) for several complexity variables from each perspective. Using untagged and tagged texts, participants will code and annotate complexity features in the text, and then quantify their coding using manual and automatic procedures with AntConc 3.4.4 (Anthony, 2014).

Participants will gain practical experience in analyzing three types of features:
1. features which can be analyzed easily through a straight-forward search using AntConc and either lexical or tag information (e.g., attributive adjectives);

2. features which require a more complicated search pattern in AntConc, such as those requiring both lexical and tag information (e.g., nominalizations);
3. features which require manual coding (e.g., appositive noun phrases, prepositional phrases as nominal post-modifiers; mean length of T-unit, number of complex nominal per T-unit).

Issues such as reliability (precision, recall), interrater reliability, and 'fixtagging' will be addressed. For the purposes of this workshop, participants will manually analyze the texts in order to gain a fuller understanding of what each measure is capturing. When possible, the manual analyses will be compared to the results from automatic tools (e.g., Lu, 2010).

Participants will then compile the quantitative data for each of the variables analyzed during the workshop and compare the results for the practice texts. Part 1 then culminates with a case study carried out by the workshop organizers on a larger sub-sample of the BAWE corpus that compares two levels and two disciplines using both variables from both traditions. Part 1 concludes with a discussion of issues and findings raised during the session.

Part 2: Research Synthesis on the Development of Grammatical Complexity in Academic Writing

Part 2 is a research synthesis of recent work by the workshop organizers and colleagues on the development of grammatical complexity in academic writing. This synthesis, which focuses on the register/functional tradition, allows workshop participants to see complexity measures from this perspective practiced in Part 1 applied to the study of writing development. The research synthesis brings together results from both L1 and L2 writers, testing and classroom contexts, and a range of registers, questioning how such variables mediate the development of grammatical complexity in novice writers. References to published studies can be found below and in the reference list; we will also draw from work in progress.

We explore the following research questions in our synthesis:
1. What measures of complexity most reflect developmental increases across a variety of contexts?
2. What measures seem to be mediated by differences across L1 and L2 writers?
3. What measures seem to be mediated by differences across assessment vs. classroom contexts?
4. What measures seem to be mediated by differences across genres and contexts?
5. Does a holistic approach (i.e., MD analysis) allow us to capture these differences more effectively?

In order to examine these research questions, we present findings from six corpora of student academic writing. First, we utilize the British Academic Writing in English (BAWE) corpus, which focuses on four academic levels (three years of undergrad and grad), as well as a variety of genres (e.g., argumentative essays, lab and research reports), and disciplines (arts and humanities, social sciences, life sciences, physical sciences). BAWE also allows for investigation across L1 and L2 writers. Second, we use two corpora of first year writing, the Freshman Academic Composition in English (FACE) corpus, which consists of both L1 and L2 writers producing writing in two genres (Rhetorical Analysis and Argumentative Essay), and the Purdue Second Language Writing Corpus (PSLW), which comprises a much larger L2 writing corpus across five genres (Literacy Narrative, Proposal, Literature Review, Interview Report, and Argumentative Essay). Finally, we draw on three assessment corpora based on 1) TOEFL iBT 2) the English Certificate of Proficiency in English writing test (ECPE) from Cambridge Michigan Language Assessments, and 3) the English Placement Test (EPT), an in-house writing placement test at Iowa State University.

Our exploration of complexity is primarily focused on the grammatical structures practiced in Part 1 (see Table 1), but does include some additional lexico-grammatical features (e.g., lexical bundles, semantic categories or lexical realizations of grammatical features), broadening the definition of complexity, which will lead into our discussion in Part 3.

We will present the results of our analyses across corpora in relation to the five research questions. First, we will illustrate that, overall, academic writing shows an increase in the use of phrasal features and a decrease in the use of clausal features as writers develop. This finding is supported by our results from analysis of the BAWE corpus (academic levels 1-4) and results from the ECPE test (proficiency levels E-A) (Staples, Egbert, Biber, & Gray, 2016; Yan & Staples, in press). We also show the similar individual features that are associated with this pattern of development. Next, we will illustrate that for some phrasal or clausal features there may be differing patterns of development based on whether the writers have an L1 or L2 background. For example, there is some evidence from our studies to suggest that L2 writers may use more noun-noun sequences than L1 writers at lower levels of development, perhaps due to increased reliance on formulaic chunks of language (Staples & Reppen, 2016). This reliance on formulaic chunks may also form a key difference between the development we see in classroom and assessment contexts, with less proficient/developed writers in assessment contexts using this strategy to an even greater extent (Yan & Staples, in press).

Fourth, we will explore the variation in complexity across genre, and to some extent discipline, in relation to development and L1 background in classroom contexts. Findings from our analyses show that while the general trend for using more phrasal features increases as academic level increases, the use of particular phrasal features varies across genres and disciplines. For example, noun-noun sequences are used much more in life and physical sciences as well as critiques when compared to arts and humanities and argumentative essays. Thus, the general findings on complexity need to be moderated based on the contexts in which students are writing (Staples et al., 2016).

Finally, we will briefly illustrate the advantages of using multi-dimensional (MD) analysis, which allows us to explore the co-occurrence patterns among the individual complexity features. For example, research on the TOEFL iBT shows few relationships between score level and individual variables but significant differences across score level for the first dimension of an MD analysis (Biber, Gray, & Staples, 2016).

Part 3: Roundtable Discussion

Part 3 provides an opportunity for extended discussion between workshop participants on the issues practiced and discussed in Parts 1 and 2. The following questions represent possible areas of discussion, although we anticipate that the roundtable will be guided by topics nominated by workshop participants:

Conceptual Questions
1. What is the definition of linguistic complexity? Should complexity be defined in absolute or relativistic terms?
2. This workshop has focused on grammatical/syntactic complexity. What are some other ways that language can be 'complex'? How can we measure these other forms of complexity in language?
3. What is the difference between "linguistic complexity" and "text complexity"?

Synthesis Questions
4. What are the fundamental differences between the two major approaches to complexity research introduced in this workshop? What is the impact of these differences on how we interpret research findings from these two approaches?
5. How can we directly compare or synthesize complexity research based on fundamentally different measures of complexity?

Extension Questions

6. Can/should the measurement of complexity be fully automated? What is the role of computers and humans in complexity research, particularly with respect to grammatical taggers and computational tools?
7. How should accuracy be accounted for in studies of grammatical complexity within the context of language development?
8. How is complexity related to language development in speech? Should we measure complexity differently for speech and writing? What about for more fine-grained register distinctions (e.g., different types of writing)?
9. Throughout the workshop we have presented results from several different studies. What are the implications of these studies for ESL learners? For language testing?

During the roundtable discussion, we also invite participants to introduce their own research projects on complexity, as a venue for asking questions of other researchers working in this area, problem-solving concrete issues related to complexity research, and sharing their approach/preliminary findings. Participants wishing to discuss their complexity research during the roundtable discussion should send a 1-page summary of their research to the workshop organizers 1 week prior to the workshop. On the 1-page summary, please include the study's research question(s), a description of the context and data/corpus, a listing of the complexity variables used (and definitions if needed), and an indication of what the researcher would like to discuss with the group (e.g., specific questions or problems that the researcher has, results the researcher would like to discuss, etc.).

References

Anthony, L. (2014). AntConc (Version 3.4.4) [Computer software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

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

Biber, D. (1992). On the complexity of discourse complexity: A multidimensional analysis. *Discourse Processes, 15,* 133-163.

Biber, D. & Gray, B. 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 characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45(1),* 5-35.

Biber, D., Gray, B., & Poonpon, K. (2013). Pay attention to the phrasal structures: Going beyond T-units—A response to WeiWei Yang. *TESOL Quarterly, 47(1),* 192-201.

Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics, 37(5),* 639-669.

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

Crossley, S., & McNamara, D. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26,* 66-79.

Hunt, K. (1966). Recent measures in syntactic development. *Elementary English, 43(7),* 732-739.

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

Nesi, H., Gardner, S., Thompson, P., & Wickens, P. (2008-2010). *The British Academic Written English (BAWE) corpus.*

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

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24(4),* 492-518.

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.

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication, 33(2),* 149-183.

Staples, S. & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing, 32*, 17-35.

Wolfe-Quintero, K., Inagaki, S., & Kim, H-Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity.* University of Hawai'i at Manoa.

Yan, X., & Staples, S. (in press). Investigating lexico-grammatical complexity as construct validity evidence for the ECPE writing tasks: A multidimensional analysis. *CaMLA Working Papers*.