The Complexities of Complexity: Ways and Reasons to Measure Complexity in L2 Writing

Laurence Craven, American University of Sharjah, United Arab Emirates

Abstract

There are many ways to measure students' written second language (L2) performance and one way to measure it is using complexity, accuracy and fluency (CAF). Teachers rarely use these measures; however, these measures are often used in L2 acquisition research. Research on CAF measures began in the 1980s when researchers started to point out a difference between fluent and accurate language use and complexity was later to appear, in the 90s. Out of these three proficiency measures, complexity is usually is the least straightforward, complex and can be measured and defined in different ways. Choosing which definition and measurement of complexity to use is an issue for researchers, since various ones coexist. This article will examine the way researchers have measured complexity and point out the advantages, disadvantages and controversies relating to these measures. It will attempt to propose the most up to date way and useful ways of measuring complexity that would be the most appropriate for those researching L2 writing performance. Finally, it will present some reasons why measuring complexity is important for researchers and also teachers.

Keywords: Complexity, CAF Measures, Second Language Acquisition

Introduction

There are many ways to measure students' written second language (L2) performance. One way to measure L2 performance is using complexity, accuracy and fluency measures (CAF). Practitioners rarely use these measures, but these measures are often used in L2 acquisition research. Research on CAF measures began when researchers started to point out a difference between fluent and accurate language use, but complexity was later to appear in CAF literature (Housen & Kuiken, 2009). Complexity is usually is the least straightforward of the CAF measures and can be measured and defined in different ways. Choosing which definition and measurement of complexity to use is an issue for researchers, since various ones coexist. This article will examine the way researchers have measured complexity and point out the advantages, disadvantages and controversies relating to these measures. It will then attempt to propose the most up to date ways of measuring complexity that would be the most appropriate for those researching L2 writing performance and lastly, it will present some reasons why measuring complexity is important for researchers and also teachers.

CAF Measures

When researchers look at students' writing, they sometimes observe the variation in complexity, accuracy, and fluency of student output. These CAF measures are useful for research purposes, but can also be used by English as a second language teachers because they can utilise the research findings to improve their practice and their students' language performance. Complexity is usually said to be the most controversial out of the three CAF measures and has been defined in various ways. Because many definitions of complexity exist, the problem for researchers is choosing which one to use. Wolfe-Quintero et al. (1998, p.69) define complexity as "a wide variety of both basic and sophisticated structures and words that are available to the learner", but Ellis and Barkhuizen (2005, p.139) define complexity "as the extent to which learners produce elaborated language." It seems that these definitions appear to be worded differently, but they are similar in meaning.

As well as issues with defining complexity, there are also issues and controversies when measuring it. Norris and Ortega (2009), argue that three grammatical complexity measures as a minimum (global complexity, phrasal complexity, and complexity by subordination) should be measured since language can be elaborated at three different levels**.** Morphological complexity is a relatively new measure of complexity used in SLA studies (Pallotti, 2015) (Brezina & Pallotti, 2016). Brezina and Pallotti (2019) define morphological complexity as "the diversity of inflectional types of a given word class" (p.100), but in SLA research Bulté and Housen, point out, most researchers mostly focus on syntactic or grammatical complexity (2012).

Ways to Measure Complexity

Traditionally, syntactic complexity has been measured in different ways. The number of dependent clauses per total clauses, clauses per T-unit; the shortest grammatically allowable sentences into which writing can be split, or number of dependent clauses per T-unit are common methods. Type-token ratio (TTR), which is the number of word types divided by all word tokens is another method to measure complexity by looking at lexical diversity, but TTR, has been called flawed by some because long works depress TTR (Ellis & Barkhuizen, 2005). The diversity index D (Malvern et al., 2004), which is a mathematical transformation of the standard TTR reduces the effects of text length and also provides the degree of word

repetition in a text is a newer way of measuring lexical diversity. *voc-d* which is a method for measuring the diversity of text units, and takes a number of subsamples of tokens at random from the data (McKee et al., 2000), is the way D can be computed. MTLD (McCarthy & Jarvis, 2010) is another newer way to measure complexity. MTLD is similar to D, but D tends to be based on TTR using random selection and curve fitting to reduce the impact of text length. MTLD however, uses TTR as a cut-off point to examine the text length, for which a writer can maintain a certain level of lexical diversity (McCarthy, 2005). HD-D (hypergeometric diversity of D) is another recent complexity measure based on D and similar to MTLD. It is slightly more accurate and slightly more stable than VOC-D (McCarthy & Jarvis, 2010).

Although complexity is an important measure of second language (L2) writing proficiency (Lu, 2011), and large-grained indices such as the mean length of T-unit (MLTU) are the most prevalent in SLA studies (Ortega, 2003), they have recently have come under criticism. Some of the criticisms include their focus on clausal subordination (Biber, Gray, & Poonpon, 2011), and that the types of structures found within a T-unit are not captured by large-grained indices. Kyle & Crossley, (2017) have argued that indices related to usage-based characteristics of verb argument construction (VAC) use could be better indicators of writing development than the traditional measures of syntactic complexity. However, most researchers presently use traditional CAF measures.

There are many online tools that are available to the general public to be able to measure writing complexity. The L2 Syntactic Complexity Analyzer (L2SCA) (Lu, 2010) is a computational tool that can produce syntactic complexity indices. It is often used in research studies that focus on L2 writing performance and can calculate 14 syntactic complexity indices. The Lexical Complexity Analyzer (LCA), (Lu, 2010), also developed by Professor Xiaofei Lu at The Pennsylvania State University, likewise analyses lexical complexity of written English using 25 different measures of lexical density, variation and sophistication and is also an online tool that is easy to use. Another online computational tool that is equally as prevalent in use in L2 writing performance studies is Coh-Metrix. Coh-Metrix can analyse the cohesion and coherence of a text (Graesser, McNamara & Louwerse, 2003) and can provide the indices for measuring complexity especially voc-D and measures of lexical diversity MTLD. Most recently, an online tool that was not specifically developed to measure complexity, but can do so is the AI chatbot ChatGPT. ChatGPT, developed by OpenAI and released last year is a type of large language model, but its most notable disadvantage, however, is a tendency to provide inaccurate information and thus when using it to calculate syntactic and lexical diversity measures, caution must be taken. Open AI acknowledges that ChatGPT can write incorrect answers which is common behavior for large language models. Another tool to analyse written complexity is CLAN (Computerized Language Analysis) a cross-platform program. It was created for analyzing transcripts in the Child Language Exchange System (CHILDES) database. CLAN has been used to create and analyze a wide variety of corpora, but unlike the above mentioned programs, the data that is inputted must be formatted in the CHAT transcription format. This format needs programming knowledge and unlike the other programs mentioned above, that just need text to be cut and pasted to run the measures, it is thus much more complicated and time consuming.

There are a variety of reasons to use the above tools. Researchers examining writing proficiency can use complexity measures to see if certain conditions are affecting the students' written performance. These can include task type, task factors such as genre of the tasks, task condition, task structure, the familiarity regarding the topic, planning time, as well

as the level of cognitive complexity of the tasks (Rahimpour, 1999, 2008). The type of pedagogical intervention is also an external factor that researchers look at to see if it may affect complexity, such as different types of written corrective feedback. A study by Van Beuningen et al. (2012), for example looked if written corrective feedback would improve complexity in text revisions and found that the group that did not receive feedback, but wrote a new text instead displayed less complex writing than the writing of students who received direct or indirect feedback. Individual difference variables, for example learners' proficiency level, anxiety of the L2 learners, motivation, or aptitude are other factors that can influence complexity (Rahimpour, 1999, 2008).

Conclusion and Discussion

As well as being useful for researchers, complexity measures are very seldom used by teachers of writing or academic writing, but could be incorporated in classroom activities or as one of the goals of a course. Since syntactic and lexical diversity measures are easy to calculate using on-line tools, incorporating the on-line tools as a classroom activity can be an enjoyable change for both the students and the teacher. An academic essay could be posted into the on-line tools mentioned above and certain complexity measures could be selected. The students, with the guidance of the teacher, could then work towards improving the complexity of the essay and then looking at the result to see, if in fact the essay has improved in complexity according to one the measures selected. It is important to note, however, that examining the assessment of complexity in second language writing for communicative purposes, rather than just for research, is not fully practical without taking into account its functional dimension. Thus, thinking about the concept of functional adequacy when looking at complexity could also be considered (Pallotti, 2009; Kuiken, Vedder & Gilabert, 2010) and not only relying on complexity measure such as MTLD and voc-d.

References

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. https://doi.org/10.5054/tq.2011.244483

Brezina, V. & Pallotti, G. (2016). Morphological complexity in written ESL texts. *Second Language Research, 35(1), 99-109.* https://doi.org/10.1177/0267658316643125

Bulté, B., & Housen, A. (2012). Defining and Operationalising L2 Complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds), *Dimensions of L2 Performance and Proficiency - Investigating Complexity, Accuracy and Fluency in SLA* (pp. 21-46). Amsterdam: John Benjamins.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing,26*, 42–65.https://doi.org/10.1016/j.jslw.2014.09.005

Ellis, R., & Barkhuizen, G. (2005). *Analysing Learner Language.* New York: Oxford University Press.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquistion. *Applied Linguistics 30*(4), 461-473.https://doi.org/10.1093/applin/amp048 http://www.ehow.com/way_5766640_studies-attitudes-towards-learning-english.html

Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing, 34*(3), 321–336. https://doi.org/10.1177/0265532216663991

Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing, 34*(4), 513–535. https://doi.org/10.1177/0265532217712554

Lu, X. (2010). Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics, 14*(1), 3–28. https://doi.org/10.1075/ijcl.14.1.02lu

Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development: quantification and assessment.* New York: Palgrave Macmillan.

McCarthy, P.M. (2005). An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). *Dissertation Abstracts International, 66,* 12.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

McKee, G., Malvern D. D., & Richards, B. J. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing,15*(3), 323-338.https://doi.org/10.1093/llc/15.3.323

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics 30*(4), 555-578.https://doi.org/10.1093/applin/amp044

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.https://doi.org/10.1093/applin/24.4.492

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590-601.https://doi.org/10.1093/applin/amp045

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

Rahimpour, M. (1999). Task complexity and variation in interlanguage. In N. O. Jungheim & P. Robinson (Eds.), *Pragmatic and pedagogy: proceeding of the3rd pacific Second Language Research Forum* (pp.115-134). Tokyo, Japan: Pac LRF.

Rahimpour, M. (2008). Implementation of task-based approaches to language teaching. *Pazhuhesh-e-Zabanha-ye Khareji Journal*, 41, 45-61.

Van Beuningen, C., de Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in Dutch multilingual classrooms. *Language Learning, 62*(1),1-41.https://doi.org/10.1111/j.1467-9922.2011.00674.x

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