Applied Linguistics 2014: 35/5: 607–614
doi:10.1093/applin/amu047 Advance Access published on 24 August 2014

FORUM

Complexity, Accuracy, and Fluency in Task-based L2 Research: Toward More Developmentally Based Measures of Second Language Acquisition

[1,*]CRAIG LAMBERT and [2]JUDIT KORMOS

[1]Faculty of Foreign Studies, The University of Kitakyushu, Japan and [2]Department of Linguistics and English Language, Lancaster University, UK
*E-mail: lambert@kitakyu-u.ac.jp

This article surveys how complexity, accuracy, and fluency (CAF) have been operationalized in studies of task-based L2 production, pointing out some problems with this approach and the need for more precise information about L2 development during task performance. Research into developing L1 text construction ability is then discussed and some approaches for establishing measures of the relevant constructs in L2 performance are suggested.

INTRODUCTION

Syntactic complexity, grammatical accuracy, and fluency (CAF) have provided the standard of measurement in task-based L2 research for nearly two decades. However, these measures have been criticized in regard to their validity for second language acquisition (SLA) (e.g. Larsen-Freeman 2009; Norris and Ortega 2009; Pallotti 2009; Biber *et al*. 2011). Three main facets of this criticism are (i) measures of CAF are unlikely to have a linear relationship with L2 proficiency (e.g. Larsen-Freeman 2009; Norris and Ortega 2009), (ii) measures of CAF may not reflect the constructs they are purported to measure (e.g. Kormos and Denes 2004; Biber *et al*. 2011), and (iii) while CAF may occasionally overlap with more adequate or more advanced language, they represent separate underlying constructs (Pallotti 2009). This article discusses how CAF have been operationalized in task-based research, points out some serious concerns, and identifies directions for establishing measures of performance which may relate more directly to L2 development.

COMPLEXITY

Standard ways of operationalizing syntactic complexity in L2 research have focused on verbal subordination. The most commonly used measure has been

the ratio of finite and non-finite clauses to a sentential unit of analysis such as the terminal unit (Hunt 1965), the communication unit (e.g. Crookes 1990; Bardovi-Harlig 1992), or the analysis of speech unit (Foster *et al*. 2000). However, this approach potentially obscures developmental processes by (i) not differentiating between types of subordination, (ii) not controlling for item-based use of subordinate structures, and (iii) not considering potential interactions between subordination, discourse genre, and mode of production.

First, measuring subordination as a unitary construct masks three distinct syntactic processes in English: (i) using nominal clauses as objects of superordinate verbs; (ii) using adverbial clauses to modify superordinate verbs; and (iii) relativizing clauses to modify superordinate nouns (Nippold *et al*. 2005a,b; Schmid *et al*. 2011). Evidence shows that these three processes emerge at different points in the developmental process. For example, in two studies of 180 and 120 L1 English participants in three age groups (11 years, 17 years, and 20–29 years), Nippold *et al.* found no effect for age on overall subordination in speech or in writing. In fact, their results confirmed a plateau in subordination at around the age of 11 years (cf. Rubin and Piche 1979; Rubin 1982). However, *post hoc* analyses revealed that relative clause subordination continued developing into adulthood (20–29 years), and nominal subordination, particularly with early-emerging verbs such as *think*, was the predominate form of subordination for the youngest learners. These findings demonstrate how the use of measures that conceptualize subordination as a unitary process can mask, rather than illuminate, developmental variation during task performance.

Similar claims have been made about L2 development. Norris and Ortega (2009), for example, referencing work by Halliday *et al.* (Halliday and Martin 1993; Halliday and Mattiessen 1999), argue that progress involves increased coordination of simple independent clauses at the beginning level, increased verbal subordination at the intermediate level, and nominalization of information at the advanced level. Schmid *et al*. (2011: 50) also claim that '. . . relative clauses develop somewhat later than adverbial and nominal clauses, and at one point seem to compete with them'. Again, these studies show how measures of verbal subordination as a ratio of clauses to sentential units can mask important developmental variation. *Increased* subordination might reflect development at the lower-intermediate level, whereas *decreased* subordination might do so at the upper-intermediate level (Norris and Ortega 2009).

A second problem with current measures of syntactic complexity is item-based usage. Diessel and Tomasello (2001) provide evidence that some early-emerging verbs (e.g. *think*, *see*) that appear to be accompanied by subordinate clauses are not actually cases of subordination because the process cannot be extended to other contexts. If this is the case, estimates of subordination will be distorted by item-based production. This would be particularly problematic with L2 learners at the beginning and lower-intermediate levels

where subordinate structures with verbs such as *think* and *see* may predominate.

Finally, verbal subordination may not be equally relevant across discourse genres and modes. Yule (1997) argues that the discourse demands of tasks (description, instruction, narration, opinion) impose distinct, developmentally relevant linguistic demands on L2 learners. Support for this position is provided by Nippold *et al.* (e.g. Nippold 2004) and Berman *et al.* (e.g. Berman and Nir-Sagiv 2007; Berman 2008). Furthermore, Biber *et al.* (2011) and Halliday and Martin (1993) have shown that different linguistic features characterize spoken and written production. Speech relies less on phrasal embedding and complex sentences, and writing is typically characterized by complex nominalization and the use of abstract and compound nouns (Fang *et al.* 2006).

ACCURACY

Measures of grammatical accuracy in task-based research have typically estimated the proportion of errors produced in different task conditions as indices of learners' attention to form. Two common approaches have been (i) calculating the ratio of errors in a text to some unit of production (e.g. words, clauses, sentential units) and (ii) calculating the proportion of these units that are error free. However, there are validity and reliability problems with both approaches. Pallotti (2009), for example, points out that errors cannot provide valid measures of L2 development as a given piece of discourse might contain perfectly accurate use of early-emerging structures, whereas another piece of discourse might contain many errors in the use of late-emerging structures. The former discourse would be more accurate but less advanced, whereas the latter would be less accurate but more advanced. Furthermore, Thewissen (2013) has shown that above the intermediate level the frequency of errors does not accurately differentiate students at different levels of proficiency.

Identifying errors is also problematic for reliability. Dichotomous decisions between right and wrong language use mask a considerable amount of variation. Bard *et al.* (1996), for example, make a distinction between levels of grammaticality and levels of acceptability. The former relate to the grammatical conventions of the language itself which will allow a certain degree of variation based on social and regional differences, and the latter relate to the acceptability of a sentence in a given situation. They argue that in making grammaticality judgments, raters do not only respond to the grammaticality of sentences, but to other factors which include the estimated frequency with which the structure has been heard, the degree to which an utterance conforms to a prescriptive norm, and the degree to which the structure makes sense to the rater semantically or pragmatically. Such acceptability factors are difficult to separate from grammaticality even for experienced raters.

FLUENCY

In task-based L2 research, fluency is often conceptualized as (i) *break-down fluency*, which relates to pausing behavior, (ii) *repair fluency*, which relates to the frequency of repetitions and self-corrections, and (iii) *speed fluency*, which relates to rate of delivery (for a recent discussion see Bosker *et al*. 2013). One of the most frequently used measures of fluency is *speech rate* which is usually calculated as a ratio of syllables produced to time taken to produce them. Another is *dysfluency* which is typically the ratio of dysfluency markers (e.g. filled/unfilled pauses, hesitations, false-starts, verbatim repetitions, self-repairs) to some discourse unit (e.g. words, clauses, or sentential units).

Kormos and Denes (2004) correlated 13 typical measures of L2 fluency with the holistic fluency judgments of experienced native and non-native teachers on the oral performances of 16 L2 learners of English. They found that the measures which correlated highly with expert norms of proficiency were speech rate, mean length of run, phonation time ratio, and number of stressed words per minute. Accuracy and lexical diversity also correlated with some raters' fluency judgments. On the other hand, dysfluency measures (i.e. silent pauses, filled pauses, and total pause time) did not show strong relationships with raters' judgments. De Jong *et al*. (2012) provide further support for the use of speech rate over pause phenomena in measuring L2 fluency.

Although speed and pausing measures might provide an indication of automaticity and efficiency in the speech production process with respect to specific forms, their fluctuation is subject to too many variables to reflect development directly. Towell and Dewaele (2005), for example, found no concomitant increase in fluency with grammatical development. Furthermore, a plateau in speech rate has been observed in line with comprehensibility (Munro and Derwing 2001), and fluency may be at least partly a trait-like characteristic dependent on working memory resources. Learners' L1 fluency might thus affect L2 fluency development (Towell and Dewaele 2005). In short, even if fluency measures correlated positively with grammatical development in specific cases, a daunting range of individual and situational factors would need to be controlled before this data could be interpreted developmentally.

WHAT DO CAF MEASURE?

The question of what CAF might accurately measure is crucial to task-based L2 research. On the one hand, fluency and accuracy seem to correlate with the proficiency norms of certain speech communities as measured by L2 teachers' fluency judgments (Kormos and Denes 2004). While ultimately relative, such norms are routinely used in high-stakes decision-making. CAF will thus remain important in theory and research related to L2 instruction and evaluation, and work to further define and operationalize them continues (e.g. Housen *et al*. 2012). On the other hand, CAF has been argued to provide insight into learners' allocation of attention during L2 use and to reflect the

cognitive processes that they engage in with respect to language (e.g. Robinson 2011; Skehan 2014). As we have seen, however, too many other factors are involved in fluctuations in CAF during L2 task performance for them to provide clear insight into development.

TOWARD MORE DEVELOPMENTALLY BASED MEASURES OF L2 TASK PERFORMANCE

One approach to identifying developmentally sensitive measures of task-based L2 performance is cross-sectional. An example is provided by Ravid and Berman (2010) who propose a multi-faceted model of noun phrase complexity, which they show to be related to the developmental level of L1 participants in four age groups (9–10 years, 12–13 years, 16–17 years, and 20–30 years) across task types (narrative and expository), language modes (spoken and written), and typologically distinct languages (English and Hebrew). Based on their results, the authors conclude that the noun phrase is a late-developing aspect of the language system and that their model is sensitive to more or less advanced use of this system. While this approach adds a degree of developmental depth to linguistic measures of performance, the result is still a set of context-free quantitative measures that are subject to the same criticisms as the syntactic complexity measures discussed above. Berman (2008), however, proposes a more comprehensive model of developing text construction ability that balances local linguistic measures (lexical and syntactic) with measures of discourse strategy, information density, connectivity, and perspective. Based on nearly two decades of cross-sectional empirical research, Berman *et al.* (see Berman 2008 for an overview) provide evidence that these five aspects of text construction are both sensitive to developmental differences and compete with one another during the developmental process.

A second alternative is researching how L2 development on tasks might be more accurately conceptualized and measured is a dynamic systems approach (Verspoor *et al.* 2011). Bassano and van Geert (2007), for example, demonstrate how quantitative longitudinal data can be used to understand how new forms emerge and how change is shaped. In modelling the discontinuous growth patterns of L1 French utterances, the authors argue that the emergence of a given form involves interactions between (i) individual differences and input stimuli, and (ii) target forms and other developing language skills. They also demonstrate how relationships between variables in the developmental process might be neutral, supportive, conditional, and competitive. The insights this approach provides are essential for understanding relationships between task performance and L2 development. Their approach might be used, for example, to investigate how the five areas of text construction ability identified by Berman (2008) interact at different points in the developmental process, and which aspects of performance might be used to accurately measure development at different proficiency levels and in different discourse

contexts. In particular, developmental measures in both L1 and L2 performance are likely to depend to some extent on discourse genre (e.g. Yule 1997; Nippold 2004; Berman and Nir-Sagiv 2007; Berman 2008) and mode of production (e.g. Halliday and Martin 1993; Fang *et al*. 2006; Byrnes *et al*. 2010; Biber *et al*. 2011; Norris and Manchon 2012).

However, it is also possible that fundamental assumptions regarding the relationship between complexity and proficiency need to be reconsidered. It may be, for example, that learners at the intermediate level use advanced forms more often than learners at the advanced level due to the nature of the learning process itself. After mastering new forms for one's purposes, ongoing development may involve optimizing efficiency in their use. White, for example, has argued that proficiency may work against linguistic complexity due to higher proficiency speakers having the experience base necessary to accomplish tasks using the minimum necessary linguistic resources (White and Robinson 1995). It is frequently the case, for example, that expert speakers and writers express complex ideas more simply than novices. This is not due to the availability of linguistic resources but rather to practiced mastery in efficient and effective message formation.

The present article has outlined serious concerns with current approaches to measuring L2 performance in task-based research and proposed some initial directions for identifying more developmentally oriented measures. Both cross-sectional and longitudinal research on L2 development across proficiency levels, discourse genres, production modes, and target languages is important in accurately conceptualizing and measuring the effects of task performance on L2 development. Due to the complex and dynamic nature of the variables involved in the developmental process, however, local fluctuations in accuracy, fluency, and syntactic complexity will not provide adequate insight into task-based SLA. Without theoretical modelling and empirical support linking performance measures to the use of developmentally more advanced language, task-based research is likely to result in mixed findings that are of limited value for SLA.

REFERENCES

Bard, E., D. Robertson, and **A. Sorace.** 1996. 'Magnitude estimation of linguistic acceptability,' *Language* 72: 32–68.

Bassano, D. and **P. van Geert.** 2007. 'Modeling continuity and discontinuity in utterance length: A quantitative approach to changes, transitions and intra-individual variability in early grammatical development,' *Developmental Science* 10: 588–612.

Bardovi-Harlig, K. 1992. 'A second look at t-unit analysis: Reconsidering the sentence,' *TESOL Quarterly* 26: 390–95.

Berman, R. 2008. 'The psycholinguistics of developing text construction,' *Journal of Child Language* 35: 735–71.

Berman, R. and **B. Nir-Sagiv.** 2007. 'Comparing narrative and expository text construction across adolescence: A developmental paradox,' *Discourse Processes* 43: 79–120.

Biber, D., B. Gray, and **K. Poonpon.** 2011. 'Should we use characteristics of conversation to measure grammatical complexity in L2 writing development'?,' *TESOL Quarterly* 45: 5–35.

Bosker, H., A. Pinget, H. Quené, T. Sanders, and **N. De Jong.** 2013. 'What makes speech sound fluent? The contributions of pauses, speed and repairs,' *Language Testing* 30: 159–75.

Byrnes, H., H. Maxim, and **J. Norris.** 2010. 'Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment,' *The Modern Language Journal* 94. (Suppl. 1).

Crookes, G. 1990. 'The utterance and other basic units for second language discourse analysis,' *Applied Linguistics* 11: 183–99.

De Jong, N., M. Steinel, A. Florijn, R. Schoonen, and **J. Hulstijn.** 2012. 'Linguistic skills and speaking fluency in a second language,' *Applied Psycholinguistics* 34: 893–916.

Diessel, H. and **M. Tomasello.** 2001. 'The acquisition of finite complement clauses in English: A corpus-based analysis,' *Cognitive Linguistics* 12: 97–141.

Fang, Z., M. Schleppegrell, and **B. Cox.** 2006. 'Understanding the language demands of schooling: Nouns in academic registers,' *Journal of Literacy Research* 38: 247–73.

Foster, P., A. Tonkyn, and **G. Wigglesworth.** 2000. 'Measuring spoken discourse: A unit for all reasons,' *Applied Linguistics* 21: 354–75.

Halliday, M. and **J. Martin.** 1993. *Writing Science: Literacy and Discursive Power*. Falmer.

Halliday, M. and **C. Matthiessen.** 1999. *Construing Experience through Meaning: A Language-Based Approach to Cognition*. Continuum.

Housen, A., F. Kuiken, and **I. Vedder.** 2012. *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. John Benjamins.

Hunt, K. 1965. *Grammatical Structures Written at Three Grade Levels*. National Council of Teachers of English.

Kormos, J. and **M. Denes.** 2004. 'Exploring measures and perceptions of fluency in the speech of second language learners,' *System* 32: 145–64.

Larsen-Freeman, D. 2009. 'Adjusting expectations: Complexity, accuracy, and fluency in second language acquisition,' *Applied Linguistics* 30: 579–89.

Munro, M. and **T. Derwing.** 2001. 'Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate,' *Studies in Second Language Acquisition* 23: 451–68.

Nippold, M. 2004. 'Research on later language development: the school-age and adolescent years' in R. Berman (ed.): *Language Development Across Childhood and Adolescence*. John Benjamins, pp. 1–8.

Nippold, M., L. Hesketh, J. Duthie, and **T. Mansfield.** 2005a. 'Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults,' *Journal of Speech, Language, and Hearing Research* 48: 1048–64.

Nippold, M., J. Ward-Lonergan, and **J. Fanning.** 2005b. 'Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic and pragmatic development,' *Language, Speech and Hearing Services in Schools* 36: 125–38.

Norris, J. and **R. Manchon.** 2012. 'Investigating L2 writing development from multiple perspectives: issues in theory and research' in R. Machon (ed.): *L2 Writing: Multiple Perspectives*. Walter de Gruyter, pp. 221–44.

Norris, J. and **L. Ortega.** 2009. 'Towards an organic approach to investigating CAF in instructed SLA: The case of complexity,' *Applied Linguistics* 30: 555–78.

Pallotti, G. 2009. 'CAF: Defining, refining and differentiating constructs,' *Applied Linguistics* 30: 590–610.

Ravid, D. and **R. Berman.** 2010. 'Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis,' *First Language* 30: 3–26.

Robinson, P. 2011. *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*. John Benjamins.

Rubin, D. 1982. 'Adapting syntax in writing to varying audiences as a function of age and social cognitive ability,' *Journal of Child Language* 9: 497–510.

Rubin, D. and **G. Piche.** 1979. 'Development in syntactic and strategic aspects of audience adaptation skills in written persuasive communication,' *Research in the Teaching of English* 13: 293–316.

Schmid, M., M. Verspoor, and **B. MacWhinney.** 2011. 'Coding and extracting data' in M. Verspoor, K. de Bot, and W. Lowie (eds): *A Dynamic Approach to Second*

Language Development: Methods and Techniques. John Benjamins, pp. 39–54.

Skehan, P. 2014. *Processing Perspectives on Task Performance*. John Benjamins.

Thewissen, J. 2013. 'Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus,' *Modern Language Journal* 97: 77–101.

Towell, R. and **J. M. Dewaele.** 2005. 'The role of psycholinguistic factors in the development of fluency' in J. M. Dewaele (ed.): *Focus on French as a Foreign Language*. Multilingual Matters, pp. 210–39.

Verspoor, M., K. de Bot, and **W. Lowie.** 2011. *A Dynamic Approach to Second Language Development: Methods and Techniques*. John Benjamins.

White, R. and **P. Robinson.** 1995. 'Current approaches to syllabus design: a discussion with Ron White,' *Guidelines* 17: 93–101.

Yule, G. 1997. *Referential Communication Tasks*. Lawrence Erlbaum.