

## **The Role of Theory in L2 Empirical Research on Language Complexity, Accuracy and Fluency**

**Yanina Tsupa**

*Teachers College, Columbia University*

### **INTRODUCTION**

Language proficiency is multi-componential in nature and, according to many SLA researchers and L2 practitioners, is best captured by the concepts of complexity, fluency, and accuracy (CAF). In the 1970s, the CAF triad appeared and has since been used as metrics of L2 development. Gaining more prominence in the 1990s and the recent years, CAF have started to figure as central foci of investigation on proficiency (Housen et al., 2012, p.1). The research on language complexity, accuracy, and fluency resides in task-based studies that investigate how task complexity impacts the development of CAF and their interdependence. Two of the three studies reported in this paper adopted the Cognition Hypothesis (Robinson, 2003) as the means to address the question of whether increasing task complexity positively impacts learning and performance. The Cognition Hypothesis claims that increasing cognitive task complexity along specific dimensions has the potential to promote “greater analysis, modifications, and restructuring of the IL (interlanguage) with consequent performance effects” (Robinson, 2001a, p. 302). The studies that invoke the Cognition Hypothesis either refer to Robinson’s Triadic Framework (Robinson, 2001a) and the Cognition Hypothesis to structure their research and explain the findings, or they take a more radical stand to argue that the premises underlying the Cognition Hypothesis and the claims derived from it are problematic.

This forum attempts to uncover the role played by theory in three empirical studies. To that end, in the sections that follow, I will begin with the review of literature on language complexity, accuracy, and fluency. Next, I will report on metric indices employed in the studies to measure the phenomenon of linguistic complexity and the results pertaining to grammatical and lexical complexity. Then I proceed with describing the status of theory in empirical research and the challenges research on CAF faces. The closing presents the discussion on how theory impacted the design of the empirical studies sampled for this paper.

### **LITERATURE REVIEW**

The studies on CAF which invoke the concepts of complexity, accuracy and fluency can be broadly divided into two categories. The first category of studies is that of exploratory nature. Even though the research in this category yields abundant insights and findings on CAF, such studies do not adopt any particular view on the theory they cite in their attempt to structure the research and explain the findings (see, e.g., Ferrari, 2012; Myles, 2012; Gunnarsson, 2012; Tonkyn, 2012). The other category includes the studies by Kuiken and Vedder (2012) and Levkina and Gilabert (2012) which took the Cognition Hypothesis as a starting point to develop their argument.

Following the suit of the exploratory studies, Tonkyn’s (2012) study lacked a particular theory to motivate its design and research questions. Yet, the value of his investigation lies in the

attempt to employ a wide range of indices of grammatical and lexical complexity to evaluate the sensitivity of measures to be able to capture short-term gains in CAF. To report on the nature and the extent of progress in speaking skills of intermediate instructed learners and the assessors' perception of the progress, Tonkyn adopted a range of measures for fluency, and accuracy. With regard to complexity, eleven indices of grammatical complexity were used, including global complexity indices (e.g., total number of tokens, syntactic complexity count, etc.), more specific complexity measures that capture the elaboration of noun and verb phrases (e.g., modal auxiliaries, noun phrase premodification, etc.), and seven lexical complexity measures that included ones for lexical variety and sophistication. In short, the results of the study suggested that more general complexity metrics seemed to be more sensitive to measure the short-term gains. Most remarkably, the study uncovered issues with the subjective assessment of the learners' short-term gains by raters: (a) grammatically complex language was not perceived as such if it involved repetition and was irrelevant to the discussion, and (b) complex language was not recognized as such if it was laboriously produced and non-fluent.

Kuiken and Vedder (2012) looked into the impact of task complexity on linguistic performance in oral and written modes and the impact of L2 proficiency on linguistic performance in different tasks. The researchers reported on three studies and argued that the premises underlying the Cognition Hypothesis (Robinson, 2003) and the claims derived from it are problematic. The first study reported no significant difference in syntactic complexity and lexical variation between more complex and less complex tasks, which showed no evidence for Robinson's Cognition Hypothesis. The second study which zeroed in on lexical variation as impacted by task complexity reported different results for French and Italian students. Specifically, Italian students used significantly more words with high frequency in more complex tasks, whereas French students used more infrequent words in complex tasks. Finally, the last study demonstrated that the influence of task complexity on linguistic performance was not constrained by the mode.

With the Cognition Hypothesis as the focus of investigation, the research by Levkina and Gilabert (2012) investigated the effects of the two dimensions (pre-task planning time and the number of elements) of a cognitive task complexity on L2 oral production. The research tested the Cognition Hypothesis by hypothesizing that, first, reducing cognitive task complexity by providing pre-task planning time will increase fluency, lexical and syntactic complexity, but will not affect accuracy. Second, increasing cognitive complexity along the number of elements will increase lexical complexity, syntactic complexity and accuracy, but will decrease fluency. Similar to the study of Kuiken and Vedder (2012), Levkina and Gilabert's study yielded mixed results that neither fully supported nor rejected the Cognition Hypothesis. Specifically, in relation to the performance effects of pre-task planning time, Levkina and Gilabert (2012) reported no significant effects on fluency, syntactic complexity, and accuracy, but positive effects on lexical complexity. Increasing the number of elements in the task yielded negative effects for fluency and insignificant changes in accuracy. In sum, the combination of pre-task planning and the task having more elements had the strongest positive effect on lexical complexity. Syntactic complexity was not affected by changes in planning time or the number of elements – the outcome contradictory to many other studies on syntactic complexity.

## **DISCUSSION**

The first cohort of studies on CAF which I refer to as having an exploratory nature usually takes an exploratory path in investigating the relationships between task conditions and linguistic performance. The goal of such studies appears to either shed new light on the concepts of complexity, fluency, and accuracy as they interact in various conditions (Gunnarsson, 2012; Myles, 2012) or to fill a gap in the research on CAF by investigating one particular aspect of linguistic proficiency (Ferrary, 2012; Tonkyn, 2012). In contrast, the research questions in the studies by Kuiken and Vedder (2012) and Levkina and Gilabert (2012) were motivated by Robinson's Cognition Hypothesis. Similar to a larger body of research that examined the relationship between task conditions and linguistic performance, the second cohort of the studies reviewed in the paper explored the CAF concepts from a psycholinguistic point of view. Regardless of the paths taken by the studies (exploratory vs. hypothesis testing), many studies on CAF generally report conflicting findings following the suit of many empirical studies on CAF resulting in contrasting predictions about the distribution of the attentional resources.

Even though none of the studies selected for this paper focused specifically on the phenomenon of complexity, all three studies made interesting observations pertaining to linguistic complexity, which was measured through a variety of tools. The studies by Kuiken and Vedder (2012) and Levkina and Gilabert (2012), for example, employed global measures of syntactic complexity – number of clauses per T-unit and AS-unit. Dramatically different in the scope of measures used, Tonkyn's research employed 11 indices of grammatical complexity and 7 lexical complexity measures, which is the most extensive collection of measures to capture the extent of progress in speaking skills made by a group of intermediate learners. The contrasting findings reported by Kuiken and Vedder in the three studies they reported pointed to the inherent flaw of the Triadic Framework and the predictions of the Cognition Hypothesis. Alongside the issue of the definition of the cognitive task complexity by Robinson, the study by Kuiken and Vedder yielded results that contradict the Cognition Hypothesis in the area of syntactic complexity. This means that the construct of syntactic complexity might be less prone to changes even when the task complexity grows or that we might not understand fully how and under what conditions complexification occurs. In contrast, Levkina and Gilabert's study confirmed the predictions of the Cognition Hypothesis in the areas of lexical complexity; however, the results did not show any effect of the more complex tasks (with a higher number of elements) on syntactic complexity. Finally, embracing a larger number of metrics to measure complexity, Tonkyn suggested that more general complexity metrics seemed to be more sensitive to measuring the short-term gains as reported in his study. In addition, the researcher made interesting observations about the perceptual difficulty in capturing complexity of responses by raters.

From the review of additional literature and the analysis of the three empirical studies, I conclude that the research on CAF faces significant challenges that speak to the plethora of inconsistent findings the CAF literature has produced. Among the challenges for the CAF research, Housen et al. (2012) identified at least 5 areas: (1) conceptualization and definitions of CAF as constructs, (2) cognitive, linguistic and psycholinguistic underpinnings of CAF, (3) interconnection of CAF components, (4) operationalization and measurement of CAF, and (5) factors impacting the manifestation and development of CAF in L2 use and learning.

As far as the first challenge is concerned, the research on CAF in general has shown to implement a variety of definitions of CAF, which led to the limited interpretation and comparability of CAF findings. The second challenge pertains to the difficulty of capturing the cognitive, linguistic and psycholinguistic mechanisms that are at play during the synchronic

manifestation of CAF in task conditions as well as CAF diachronic development in the process of L2 acquisition. The third challenge of the CAF research is associated with the interconnection of the CAF components, which, as pointed out by Housen et al. (2012), should be viewed as multivariate and dynamic. Operationalization and measurement of CAF is as problematic as their definitions. Studies that conducted survey of measurement practices in CAF research (Halleck, 1995; Norris & Ortega, 2009; Ortega 2003; Polio 1997, 2001; Wolfe-Quintero et al., 1998) highlighted the lack of consensus on CAF metrics, which leads to problems with comparability, reliability and validity of the studies. Lastly, as Housen et al. (2012) argued, the challenge of the CAF research is in identifying the factors which impact CAF manifestation in tasks as well as CFA development over time.

Equally worrisome is the situation with how complexity is measured in research. Having reviewed the representative sample of 40 empirical studies on task-based language learning published between 1995 and 2008, Bulté and Housen (2012) succinctly summed up that the empirical CAF research has taken a reductionist view on what constitutes L2 complexity by misusing the measures of complexity, over-amplifying some constructs of linguistic complexity and underrepresenting grammatical complexity and its constructs.

## CONCLUSION

The research on complexity, accuracy, and fluency and their interaction in different task conditions seems to occupy a particular position in the debate on how different task environments impact CAF. As the selected studies in this paper exemplify, often findings on CAF under different task conditions are contradictory and point to a plethora of issues. With regard to how theory impacted the design of the empirical studies sampled, it can be concluded that regardless of whether the study took an exploratory path or tested the Cognition Hypothesis, the designs of the studies were similar in a way they explored a specific aspect of CAF triad and the interaction of its elements in task conditions. Implications of the studies, however, were different. Kuiken and Vedder's (2012) study, for example, shed light on the problem with the cognitive task complexity definition by Robinson. In addition, they raised a question of whether writing task proficiency is best assessed by general or specific measures of performance, succinctly suggesting that the two types of measures can complement each other. The implications of Levkina and Gilabert's (2012) research are more classroom-oriented and explain how better task design and sequencing may affect learning over time. Finally, the implications of the exploratory studies have to do with a specific component of the CAF triad. Tonkyn (2012), for instance, argued that measures used to capture CAF development and complexity, may not be sensitive enough, so the evaluators need to engage in multiple repeated assessments.

## REFERENCES

- Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, I. Vedder, & F. Kuiken (Eds.), *Dimensions of L2 performance and proficiency investigating complexity, accuracy, and fluency in SLA* (pp. 171–198). Amsterdam: John Benjamins.
- Ferrari, S. (2012). A longitudinal study of complexity, accuracy and fluency variation in second language development. In A. Housen, I. Vedder, & F. Kuiken (Eds.), *Dimensions of L2*

- performance and proficiency investigating complexity, accuracy, and fluency in SLA* (pp. 277–298). Amsterdam: John Benjamins.
- Gunnarsson, C. (2012). The development of complexity, accuracy and fluency in the written production of L2 French. In A. Housen, I. Vedder, & F. Kuiken (Eds.), *Dimensions of L2 performance and proficiency investigating complexity, accuracy, and fluency in SLA* (pp. 247–276). Amsterdam: John Benjamins.
- Halleck, G.B. (1995). Assessing oral proficiency: A comparison of holistic and objective measures. *Modern Language Journal*, 79(2), 223–234.
- Housen, A., Kuiken, F., & Vedder I. (2012). Complexity, accuracy and fluency. Definitions, measurement and research. In A. Housen, I. Vedder, & F. Kuiken (Eds.), *Dimensions of L2 performance and proficiency investigating complexity, accuracy, and fluency in SLA* (pp. 1–20). Amsterdam: John Benjamins.
- 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, I. Vedder, & F. Kuiken (Eds.), *Dimensions of L2 performance and proficiency investigating complexity, accuracy, and fluency in SLA* (pp. 143–167). Amsterdam: John Benjamins.
- Levkina, M., & Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. In A. Housen, I. Vedder, & F. Kuiken (Eds.), *Dimensions of L2 performance and proficiency investigating complexity, accuracy, and fluency in SLA* (pp. 171–198). Amsterdam: John Benjamins.
- Myles, F. (2012). Complexity, accuracy and fluency. The role played by formulaic sequences in early interlanguage development. In A. Housen, I. Vedder, & F. Kuiken (Eds.), *Dimensions of L2 performance and proficiency investigating complexity, accuracy, and fluency in SLA* (pp. 71–93). Amsterdam: John Benjamins.
- Norris, J.M., & Ortega, L. (2009). Measurement of understanding: An organic approach to investigating complexity, accuracy, and fluency in SLA. *Applied Linguistics*, 30(4), 555–578.
- Ortega, L. (2003). Syntactic complexity measures and their relationship in L2 proficiency: A research synthesis of college-level writing. *Applied Linguistics*, 24, 492–518.
- Polio, C. (1997). Measure of linguistic accuracy in second language writing research. *Language Learning*, 47, 101–143.
- Polio, C. (2001). Research methodology in second language writing: The case of text-based studies. In T. Silva, & P. Matsuda (Eds.), *On second language writing* (pp. 91–116). Mahwah, NJ: Lawrence Erlbaum Associates.
- Robinson, P. (2001a). Task complexity, cognition resources, and syllabus design. A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). Cambridge: CUP.
- Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, 27–57.
- Robinson, P. (2003). The Cognition Hypothesis, task design, and adult task-based language learning. *Second Language Studies*, 21, 45–105.
- Tonkyn, A. (2012). Measuring and perceiving changes on oral complexity, accuracy and fluency. Examining instructed learners' short-term gains. In Housen, A., Vedder, I., & Kuiken, F. (Eds.), *Dimensions of L2 performance and proficiency investigating complexity, accuracy, and fluency in SLA* (pp. 277–298). Amsterdam: John Benjamins.

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

Yanina Tsupa an MEd student in Applied Linguistics at Teachers College, Columbia University. Specializing in Second Language Acquisition (SLA), her research interests include interlanguage variation and syntactic complexity. Correspondence should be sent to E-mail: [yt2688@tc.columbia.edu](mailto:yt2688@tc.columbia.edu)