Analytical Measures of Learners' Written Interlanguage

by Teresa Navés tnaves@ub.edu

University of Barcelona. English Department. C/ Gran Via, 585 08007

Barcelona. Catalonia. Spain Tel 93 403 5866. Fax 93 317 12 49

ABSTRACT

Esta comunicación se propone reconsiderar algunas medidas analíticas utilizadadas para medir la interlengua, en la escritura, de aprendices jóvenes y con baja competencia en inglés como lengua extranjera a la luz de estudios previos y en preparación dentro del proyecto BAF y en particular reconsiderar si las medidas de la longitud media de las cláusulas y de las oraciones deberían ser contempladas como medidas de fluidez, complejidad sintáctica o bien como pertenecientes a otro consturcto. A la luz de un análisis factorial (Navés, en preparación) y análisis multivariados y correlacionales (Navés et al. 2003, Navés, 2006, Torres et al. 2006) parece evidente que las relaciones entre algunas medidas y los constructos que éstas presuntamente representan varían dependiendo de la poblacion estudiada por lo que respecta a la madurez cognitiva de los participantes (edad) así como de la competencia (número de horas de instrucción recibidas). Por último, las medidas de número de palabras por cláusula y oración parecen que deberían ser consideradas no como representativas ni de la fluidez, ni de la complejidad sintáctica, sino probablemente de otro constructo, todavía por determinar. Más investigación con estudios de regresión de componentes son necesarios antes de poder

determinar y validar los constructos de los componentes de la escritura y su medición

ABSTRACT

This paper aims at reconsidering some analytical measures to best encapsulate the interlanguage, in writing, of young beginner learners of English as a foreign language in the light of previous and work-in-progress research conducted within the BAF project, and in particular, whether clause and sentence length should be best viewed as a fluency or syntactic complexity measusre or as part of a different construct. In the light of a factor analysis (Navés, forthcoming) and multivariate and correlation studies (Navés et al. 2003, Navés, 2006, Torres et al. 2006) it becomes clear that the relationship between different analytical measures is also dependent on learner's cognitive maturity (age) and proficiency (amount of instruction). Finally, clause and sentence length should not be viewed as either a fluency or sytactic complexity measure but as part of a different construct. It is concluded that further research using regression analysis and cluster analysis is neeed in order to identify and validate the constructs of the writing components and their measurements.

There are basically two approaches to assessing writing, using band scales and by means of analytical measures. By far the commonest method of assessing learners' writing is holistic, using band scales, "impression marking" as Gannon (1985:61) would put it.

Several studies have suggested that holistic ratings of written products are not a reliable indicator of language development or change. For Hyland (2002), however, a detractor of using analytical metrics in assessing writing, the problem is that there is little evidence to show that syntactic complexity or grammatical accuracy are either the principal features of writing development or the best measures of good writing.

Larsen-Freeman (1978) defines *index of development* as "an independent yardstick by which we can expediently and reliably gauge proficiency in a second language" (p. 439). Her view of the need for a yardstick to compare learners' production has been very influential.

In re-examining Wolfe-Quintero et al's. (1998) technical report on the analytical measures of accuracy, fluency and complexity used to assess learner's language development, N. Ellis and Larsen-Freeman (2006) concluded that "unfortunately, while some of these proved to be better than others at discriminating different developmental levels for groups, at the level of the individual, the results are less heartening. Clearly, the measures are not always sensitive to individual differences (Larsen-Freeman 1983), with some learners not conforming to the general patterns of development at all." (p. 563)

Following Applebee (2000) and Wolfe-Quintero et al. (1998), most of the measures that have been used in developmental index studies consist of intuitive operationalisations of fluency, accuracy and complexity. The underlying assumption is that these indices develop in tandem, i.e. as learners become more proficient, they write more fluently, more accurately and the texts they produce are more grammatically and lexically complex.

From a different perspective, these three areas are seen to enter into competition with one another for attentional resources (Foster & Skehan, 1996; Skehan & Foster (1997, 1999) have provided some adequate definitions:

- *Accuracy* is concerned with how well language is produced in relation to the rule system of the target language (1997: 22)
- *Fluency*, "the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems" (1999: 96-97).
- *Complexity/range* "the capacity to use more advanced language, with the possibility that such language may not be controlled so effectively. This may also involve a greater willingness to take risks and use fewer controlled language subsystems.

They assume that these three goals are in some degree of mutual tension. Writers cannot give their full attention to each of these goals. Skehan and Foster (1999) hypothesised that, when faced with cognitively demanding production tasks, L2 learners will attend to conveying meaning first and to

accuracy and the linguistic complexity of the output last. According to Skehan and Foster (2001), attentional resources are limited and attending to one aspect of performance, operationalised as accuracy, fluency and complexity, may mean that other dimensions are neglected. They propose that for language development to proceed optimally, a balance needs to be established between these three performance dimensions.

Contrary to Skehan and Foster's Limited Attentional Capacity Model, Robinson (2001) proposes that learners can access multiple and non-competitional attentional pools. Robinson's proposal is consistent with Long (1996) and with Schmidt's (2001) hypothesis on Cognition. The Cognition Hypothesis, integrating information-processing and interactionist explanations of L2 task effects (Long, 1996; Schmidt, 2001), predicts that gradually increasing the cognitive demands of tasks will push learners to greater accuracy and complexity in L2 production.

Wolfe-Quintero et al. (1998) agreed with Skehan and Foster that there are times in which it seems as if one aspect of development may progress at the expense of another. They also concluded that fluency and complexity may be more related to language development than accuracy is.

Wolfe-Quintero et al. (1998) surveyed 39 studies and concluded that the best measures of accuracy are error-free T-unit ratio and errors per T-unit; the best measures of development in fluency are the number of words per T-unit, the number of words per clausem the number of clauses per T-unit and the number of words per error-free t-unit; they chose type/token ratio of word type

variation per total words along with type/token ratio of sophisticated word types (based on frequency lists) per total word types as the best measures of lexical variety and finally the best metrics for grammatical complexity are the number of clauses per T-unit and the dependent clauses per clause

Larsen-Freeman (2006) based on Wolfe-Quintero et al.'s (1989) and Larsen-Freeman and Strom (1977) regard the following meaures as the "indices [which] have been determined to be best measures of second language development in writing": the average number of words per t-unit (fluency) grammatical complexity (average number of clauses per t-unit), accuracy (the proportion of error-free t-units to t-units), and vocabulary complexity (a sophisticated type–token ratio—word types per square root of two times the words—that takes the length of the sample into account to avoid the problem that regulartype–token ratios are affected by length (Ellis and Barkhuizen 2005).

The greatest amount of research in developmental processes seems to revolve around sentence complexity. Hillocks (1986) argues that clause length is more indicative of maturity than sentence length, as immature writers will often string together short, immature clauses. As writers mature, they are more likely to use more adjective clauses, more modifiers, and more complex nominals, as well as making more use of gerunds and infinitives. Thoughts tend to get consolidated into gradually fewer clauses and sentences, as writers increase their abilities to incorporate more ideas into a single clause.

This paper aims at reconsidering some analytical measures to best encapsulate the interlanguage, in writing, of young beginner learners of English as a foreign language in the light of previous and work-in-progress research conducted within the BAF project, and in particular, whether clause and sentence length should be best viewed as a fluency or syntactic complexity measusre or as part of a different construct. In the light of a factor analysis (Navés, forthcoming) and multivariate and correlation studies (Navés et al. 2003, Navés, 2006, Torres et al. 2006) it becomes clear that the relationship between different analytical measures is also dependent on learner's cognitive maturity (age) and proficiency (amount of instruction). Finally, clause and sentence length should not be viewed as either a fluency or sytactic complexity measure but as part of a different construct. It is concluded that further research using regression analysis and cluster analysis is neeed in order to identify and validate the constructs of the writing components and their measurements.

Navés (2006), Navés, Torras and Celaya (2003), Torras, Navés Celaya and Pérez-Vidal (2006) analysed the long term effects of an early start on EFL writing by bilingual Catalan-Spanish learners in an instructed foreign language context. The authors used a battery of analytical measures following Wolfe-Quintero et al. (1998) as indices of accuracy, fluency, lexical and syntactic complexity. The participants with two different onset ages (8 and 11 respectively) had received exactly the same amount of exposure when compared. Participants were at grades 11 and 12 when tested before entering university. These three studies found that the older learners, the late starters who had started learning EFL at the age of 11, significantly outperformed their

younger peers, the younger learner groups, early starters who, conversely, had started learning EFL at the age of 8 in all the four written domains studied: accuracy, fluency, lexical and syntactic complexity. However, as discussed in their studies, some metrics, sentence and clause length in particular were controversial. Contrary to Wolfe-Quintero et al., Ortega (2003) and Sotillo (2000) had already concluded that words per clause and words per sentence should be viewed as indices of syntactic complexity rather than as fluency measures. The authors from the BAF project (Muñoz, 2006) suggested that these length production measures behave slightly differently from either fluency or syntactic complexity indices.

In a first attempt to clarify these constructs, Navés, Torras and Celaya (2003) in their second study examined the two patterns of writing development which emerged from the differences found between the six groups of learners studied. They observed that for the three groups of young beginner learners aged below 13 who had received less up to 416 hours of instruction there was no syntactic or lexical development while accuracy and fluency seemed to increase in parallel. The authors concluded that the reason why fluency and accuracy seemed to go together in those beginner young learner groups was basically because they could only write what they knew-- which was very little and simple. However, when the written production of older and more instructed groups of learners was analysed, the predicted trade offs of accuracy and syntactic complexity were found. The development of these groups of learners' interlanguage seemed to take place not in the area of fluency and

accuracy but in the area of syntactic complexity and lexical diversity. The steady development of accuracy and fluency suddenly stopped as learners become older and more competent. It is when learners become older and more instructed that syntactic and lexical complexity steadily increase, and these two areas do so at the expense of fluency and accuracy's growth ceasing (See Larsen-Freeman, 2006 in this respect). It could be interpreted that for different threshold levels first a development of fluency and accuracy might be expected. However, when learners are old enough and become more sophisticated, they start neglecting accuracy and fluency and seem to concentrate on syntactic complexity and lexical variety. In their third study, Navés, Torras and Celaya (2003) conducted a correlation analysis of the measures based on the two patterns emerged in the previous study which confirmed that accuracy and fluency were highly correlated for the younger group of low proficiency learners while adverb sophistication and syntactic complexity measurements were highly correlated only in the older and more instructed groups of learners.

Some work in progress studies such as Navés (forthcoming) using factor analysis seem to suggest that some metrics such as sentence and clause length cluster together with more traditional syntactic complexity indices while they constitute a different factor of its own.

It can thus be concluded that further research is needed in order to investigate and validate the constructs of accuracy, fluency, lexical and syntactic complexity using analytical measurements. Recent research seems to

suggest (Ishikawa, 1995, Ortega, 2003, 2005, Navés 2006, Torras, Navés, Celaya et al. 2006) that careful attention needs to be paid to other factors such as cognitive matrurity to explain the results provided by the use of certain analytical measures in second and foreign language written production. The studies reviewed revealed that we are far from having found a yardstick to measure learner's writing performance since:

1) As Hillocks (1989), Ishikawa (1995) and Navés et al. (2003), Torres et al. (2006) pointed out there are measures which seem to be better to gauge young and low-proficient learners. Thus making the comparison with other more experienced learners more difficult.

2) The relationship between analytical meausures and the components they are believed to represent vary depending on the proficiency and cognitve maturity of the learners (Navés et al. 2003)

3) Preliminary factor analyses (Navés, forthcoming) with analytical measures suggest that instead of just four components (accuracy, fluency, lexical and syntactic complexity), clause and sentence length in particular may constitute a different factor.

Further research using regression and cluster analysis is needed to identify and validate the constructs of writing components and their measurement.

Applebee, A. N. (2000). Alternative models of writing development. In R. Indrisano & J. R. Squire (Eds.), *Perspectives on writing: Research, theory and practice* (pp. 90-111.Earlier version available at http://cela.albany.edu/publication/article/writing.htm). Newark: International Reading Association.

Connor, U. (2001). Research frontiers in writing analysis. In T. Silva & P. K. Matsuda (Eds.), *Lanmark essays on ESL writing* (pp. 75-90).

Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics introduction to the special issue. *Applied Linguistics, 27*(4), 558-589.

Foster, D. (1999). The challenge of contingency: Process and the turn to the social in composition. In Kent-Thomas Post-Process Theory: Beyond the Writing-Process Paradigm. Carbondale, IL: Southern Illinois UP.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299-233.

Gannon, P. (1985). *Assessing writing: Principles and practice of marking written English*. London: Arnold.

Hillocks, G. (1986). *Research on written composition. New directions for teaching*. Urbana: ERIC Clearinghouse on Reading and Communication Skills

Hyland, K. (2002). A conceptual overview of writing. In *Teaching and researching writing* (pp. 5-48). London: Longman.

Hyland, K. (2002). *Teaching and researching writing*. London: Longman.

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4*(1), 51-69.

Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly, 12*(4), 439-448.

Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H. W. Selinger & M. H. Long (Eds.), *Classroom oriented research in*

second language acquisition (pp. 287-305). Cambridge: Newbury House.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619.

Larsen-Freeman, D., & Strom, V. (1977). The construction of a second language acquisition index of development. *Language Learning, 27*(1), 123-134.

Long, M. H. (1998). SLA: Breaking the siege. Plenary address to the third pacific second language research forum (PACSLRF).Tokyo: Aoyama Gakuin University. National Institute of Education.

Muñoz, C. (2006). The effects of age on foreign language learning: The BAF Project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1-40). Clevedon: Multilingual Matters.

Navés, T. (2006) The Long-term Effects of an Early Start on EFL Writing. Unpublished PhD dissertation. Universitat de Barcelona. English Department. September 2006. Directed by Carmen Muñoz.

Navés, T., Torras, M. R., & Celaya, M. L. (2003). Long-term effects of an earlier start. An analysis of EFL written production. In S. Foster-Cohen & S. Pekarek (Eds.), *Eurosla-yearbook. Annual conference of the European second language association* (Vol. 3, pp. 103-130). Amsterdam: John Benjamin.

Ortega, L. (2003). Syntactic complexity measures and their relationship to l2 proficiency: A research synthesis of college-level l2 writing. *Applied Linguistics, 24*(4), 92-518.

Robinson, D. W. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research, 17*(4), 368–392.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 375-385.

Skehan, P., & Foster, D. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 1-27.

Skehan, P., & Foster, D. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*(1), 93-120.

Sotillo, S. M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology,* *4*(1), 82-119 http://llt.msu.edu/vol114num111/sotillo/default.html.

Torras, M. R., Navés, T., Celaya, M. L., & Pérez-Vidal, C. (2006). Age and IL development in writing. In C. Munoz (Ed.), *Age and rate of foreign language learning* (pp. 156-182). Clevedon: Multilingual Matters.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity. Technical report 17*.Manoa, Hawai'i US: University of Hawai'i Press.