



## Interlanguage Theory and Emergentism: Reconciliation in Second Language Developmental Index Studies

Wenting Wang

College of Humanities and Social Sciences

National University of Defense Technology

Changsha 410073, China

E-mail: ylnbbb@126.com

### Abstract

The present paper generally reviews the history of second language (L2) researchers' efforts in an attempt to find such an index and the possible reasons for the difficulties in establishing the developmental index from both the theoretical and the empirical viewpoints. Two contradictory views—interlanguage theory and emergentism—can finally reconciled so as to have a comprehensive view of second language acquisition (SLA) process.

**Keywords:** Interlanguage theory, Emergentism, L2 developmental index, Reconciliation

### 1. Introduction

The quest for a L2 developmental index was inspired by the work done in search of the index of children's first language (L1) development. To date, mean length of utterance (MLU) has been shown as "a useful measure of a child's gross language development" (Parker & Brorson, 2005). L2 researchers have acknowledged the need for an index of L2 development (Hakuta, 1975), by which they can "expediently and reliably gauge proficiency in a L2" (Larsen-Freeman, 1978) as "objective, quantitative and verifiable" (Nihalani, 1981) as possible. Such a yardstick will give a numerical value to different points along a L2 developmental continuum—numerical values which "would be correlates of the developmental process and would increase uniformly and linearly as learners proceed towards full acquisition of a target language" (Larsen-Freeman, 1978).

It was Larsen-Freeman (1976) who advocated the construction of an index of L2 development. However, due to the complexity and variation of SLA, it is more difficult to find such a yardstick for the L2 than for the L1. The present paper generally reviews the history of L2 researchers' efforts in an attempt to find such an index and the possible reasons for the difficulties in establishing the developmental index from both the theoretical and the empirical viewpoints.

### 2. Interlanguage Theory

The past fifty years witnessed a fierce debate on the role of L1 in the process of SLA, resulting in the predominance of error analysis over contrastive analysis. Intensive empirical studies indicated that the majority of the errors made by the L2 learners do not come from their L1 (Dulay & Burt, 1973; Dulay & Burt, 1974b; Bailey, Madden & Krashen, 1974; Larsen-Freeman, 1976; Krashen, Butler, Birnbaum & Robertson, 1978).

Where these errors come from received much attention further. If they are not target-like, they must be learner-internal. Inspired by the L1 acquisition theory, Corder (1967), the forerunner of error analysis, suggested that L2 learners make errors in order to test out certain hypotheses about the nature of language they are learning, thus proposing that the making of errors is a strategy, as an evidence of learner-internal processing (Ellis, 1985:47).

Now that SLA has little to do with the L1 and is learner-centered, then, similar to children's L1 acquisition which is regarded as an independent field rather than an approximation of adult language, L2 learners' language development and its overall characteristics are shed light on. This is the origin of *interlanguage*, coined by Selinker (1972) to refer to the language produced by the L2 learners, "both as a system which can be described at any one point in time as resulting from systematic rules, and as series of interlocking systems that characterize learner progression" (Mitchell & Myles, 1998: 31).

Interlanguage theory is based upon a series of assumptions listed as follows (Larsen-Freeman, 2006: 590-591):

- a. SLA is a process of increasing conformity to a uniform target language;
- b. There are discrete stages through which learners traverse along the way;

- c. Progress can best be defined in terms of one dimension of one subsystem, i.e., accuracy in morphosyntax as opposed to the dimensions of fluency or complexity;
- d. Learners move through the process in a fairly consistent manner;
- e. It is possible to adopt a two-phase research agenda, concentrating first on understanding learning and later on accounting for learners' differential success.

It is from the starting point of the interlanguage theory that studies of grammatical development have been carried out. Studies that seek to gauge overall progress by a developmental index as one of the two types of L2 developmental studies mentioned by Bardovi-Harlig and Bofman (1989) are also originated from the theory. Work done in this field is marked by Larsen-Freeman (1978, 1983), Larsen-Freeman and Strom (1977) and Nihalani (1981) as well as Gaies (1976) and Monroe (1975).

The best indices for the L1 development were undoubtedly listed in the potential candidates for L2 development. Mean length of utterance was excluded first due to the fact that L2 learners are "more cognitively mature" and therefore "capable of producing utterances which are more than a few morphemes in length shortly after they have had initial contact with the target language" (Larsen-Freeman & Strom, 1977) Therefore, L2 researchers resorted to T-unit-related measures which is more frequently used by the L1 researchers to gauge schoolchildren's language development.

As far as mean length of T-unit (MLTU) is concerned, Thornhill (1969) studied the development over a 9-week period of syntactic fluency of four adult Spanish-speakers learning ESL and found MLTU "a usable measure of development toward maturity in L2 production" (1969:37, cited from Larsen-Freeman, 1978). Similarly, Monroe (1975) for French, Gaies (1976) for ESL and Cooper (1976) for German did find that MLTU discriminated among learners of L2 at different levels of proficiency. However, Pike (1973) did not find very high correlations between MLTU written by subjects who were administered the Hunt (1970) "aluminum re-write passage" and two of his own subjective indicators of their writing ability in essays.

T-unit was first adapted in the form of error-free T-unit by Scott and Tucker (1974) in their error analysis of 22 Arabic-speaking ESL students, when they pursued an index of measurement which reflected error frequency as well as syntactic complexity. They counted the number of error-free T-units in both written and oral production and found a significant growth during the 12 weeks of intensive English training.

Based on the above-mentioned studies, Larsen-Freeman and her colleagues reported a series of progress on their construction of a L2 developmental index. This includes Larsen-Freeman and Strom (1977)'s pilot study of 48 compositions, which discovered that "the measures of length and error-free T-units seem to be much more viable contenders on which to base an index of development since they are easily quantified, linear progression does seem possible and they would seem to be impervious to differences in language backgrounds" (ibid.), Larsen-Freeman (1978)'s project on 212 compositions, in which two error-related measures (the percentage of error-free T-units and the average length of error-free T-units) were proved to be the best discriminators among the five levels of ESL proficiency represented in the population, followed by MLTU, and Larsen-Freeman (1983)'s comprehensive experiments on both cross-sectional and longitudinal data. In the last study, Larsen-Freeman (1983) discovered that "sometimes a performance variable, like the percentage of error-free T-units, worked well in one study but not in another; sometimes a performance variable discriminated significantly between adjacent proficiency levels and sometimes it did not; sometimes a performance variable did not distinguish between levels with small n's and sometimes it did".

With increasing number of studies examining the validity of these T-unit measures, more controversies among the results are given rise to. Some researchers supported (O'Donnell, 1976; Farhady, 1978; Gaies, 1980) while others criticized and proposed to utilize sentence, clause and other indices (Bardovi-Harlig, 1992; Ney, 1966; Moffett, 1968).

Now that T-unit measures can not be assured as the index of L2 development, other objective measures were proposed. Some investigators preferred generalized measures such as clause-based and sentence-based measures while others relied upon more specific ones (Skehan, 1998: 275). For example, Bardovi-Harlig (1992) suggested reconsidering sentence as the basic unit to analyze. She proposed a coordination index in order to examine the coordination ability of the L2 learners. Ishikawa (1995) compared the T-unit-based and clause-related objective measures and concluded that clause is a better unit than T-unit when learners with relatively low proficiency are the subjects. In the data of these subjects, grammatical and lexical errors are so frequent and of such a nature that they tend to interfere not only with the reader's understanding, but also with the researcher's ability to tabulate T-units (Larsen-Freeman & Strom, 1977; Vann, 1978). Other researchers reclassified errors into three degrees (Homburg, 1984), recategorized clause into adjectival, adverbial or nominal ones (Kameen, 1979), or reexamined the errors in term of pronouns, articles and connectors (Evola et al., 1980), and etc. These objective measures, not as prevalent as the sentence-based, T-unit-based and clause-based ones, are more idiosyncratic to specific subjects, users and contexts, but may be more helpful to detecting the L2 developmental sequence or the natural route of L2 development.

Though abundant the objective measures are, no one has been approved to be the index. Wolfe-Quintero, Inagaki and

Kim (1998) did a thorough review of all the developmental measures used in thirty-nine related L2 research studies in terms of written communication and discussed the variance of the results. They attributed gaps in these studies to “the different sets of developmental measures and the variety of proficiency measures that were used as the basis for comparison” (1998: 122). As a matter of fact, all the studies utilized different sets of objective measures with their own preferences and subjects. They defined proficiency in different ways, i.e., they employed varied criteria to determine proficiency levels. The variety of ways that proficiency has been conceptualized including “rating scales, standardized tests, program levels, school levels, classroom grades, short-term change in intact classes, and comparisons with native speakers” (1998: 117). These are only two elements, and there is more that can factor into the indetermination of the L2 developmental index, e.g., no agreement on the definitions of the measures and error-free, differences in genre, time allowed and statistical treatments used, etc. in Ishikawa (1995). Consequently, L2 researchers seem to be perplexed about a global index of L2 development which can be tailored into any situation and turn to the other polarity against it.

### 3. Emergentism

Failure in finding a global index of L2 development provides an evidence for emergentism, which attaches more importance to individual variability. Smith and Thelen (1993: 155) discussed the controversies between theories underlying global development and individual differences as follows:

Within the framework of structural theory, developmental studies have the following typical form: older and younger children are tested in a task and the mean performances at the two age levels are calculated. The typical finding is that younger children perform less well than older children. These mean differences in performance are considered to be the developmental facts to be explained. There are other kinds of data that could be the principal data of developmental psychology—the trajectories of change of individual children...or the magnitude of between-subject variability and changes in that variability with task and age...However, these sorts of data are not the relevant data for structural theorists to find the global order—the common structure—that transcends individual uses of presumed knowledge structure.

Therefore, the emergentists view language learning as a complex and dynamic process in which various components emerge at various levels, to various degrees, and at various times. “Individual differences are a natural consequence of learning within such a framework because of dynamic and multi-faceted nature of the emergent system. Slight differences in the relative rate, strength, or timing of the component achievements can result in relatively significant differences between individuals in behavioral outcomes” (Marchman & Thal, 2005: 150).

Corresponding to the assumptions the interlanguage theory is based upon, there are also assumptions underlying the emergentist viewpoint (Larsen-Freeman, 2006: 592-594):

- a. Although progress in SLA may be viewed as the degree to which a language learner’s interlanguage aligns with the target language, there will never be complete convergence between the two systems;
- b. There are no discrete stages in which learners’ performance is invariant, although there are periods where certain forms are dominant, periods that have been referred to as stages in the acquisition of certain grammatical structures;
- c. There are many dimensions to language proficiency—fluency, complexity and accuracy being three that are theorized to have independent status in L2 performance in that learners can have different goals at different times when performing in an L2;
- d. Learners do not progress through stages of development in a consistent manner. There is a great deal of variation at one time in learners’ performances and clear instability over time;
- e. Individual developmental path, each with all its variation, may be quite different one from another, even though in a “grand sweep” view, these developmental paths appear quite similar.

As emergentism emphasizes on learners variation, related experimental studies mainly focus on two aspects: firstly, some describe the different developmental paths of individual L2 learners with those objective measures and others just apply these objective measures to the examination of the elements that may factor into SLA and L2 writing assessment in terms of fluency, complexity and accuracy.

The first kind of study is best exemplified in Casanave (1994) and Larsen-Freeman (2006). Casanave (1994) examined changes in the writing of a small group of intermediate English students over three semesters of their intensive language program in Japan with MLTU, C/TU, percentage of complex T-units, EFT/T and EFT. Analysis demonstrated that the writing of all the students changes over time, but in a variety of ways not necessarily predicted by the T-unit research. Finally, he concluded that improvement can not be measured only quantitatively through group averages, but that it must be identified in a variety of ways that differ for individual writers. Similarly, Larsen-Freeman (2006) both quantitatively and qualitatively examined the oral and written production of five Chinese learners of English over a six-month period with C/TU, EFT/T and MLTU. The result clearly showed the emergence of fluency, complexity and accuracy, “not as the unfolding of some prearranged plan, but rather as the system adapting to a changing context, in

which the language resources of each individual are uniquely transformed through use” (2006: 590).

As far as the second type of research is concerned, as there are three stages engaged in the information processing procedure of SLA—input, central processing and output (Skehan, 1998: 6), these studies also mainly focus on the three aspects.

Input-related studies usually investigate into the effect of a particular type of instruction on L2 development, e.g., grammar instruction (Frantzen, 1995), bilingualism (Carlisle, 1989) and corrective feedback (Kepner, 1991; Robb, Ross & Shortreed, 1986). They picked out some objective measures to gauge the differences between the control group without particular instruction and the experimental group with the targeted instruction, or the pretest-posttest change due to the instruction. For example, Robb, Ross and Shortreed (1986) reduced an initial set of 19 measures of writing skill to a subset of 7, total words written and total clauses measuring fluency, the ratio of additional clauses to total words written and the total number of additional clauses measuring complexity, and EFT/T, EFT/C and W/EFT measuring accuracy. With these measures they evaluated the effects of four different types of feedback on error (correcting all categories of errors, coded feedback, uncoded feedback and marginal feedback, from the most comprehensive and salient feedback to the least) in the written work of L2 writers and found no significant difference.

Processing-related studies mainly focus on the cognitive aspects of learners and explored what factors may influence their cognitive ability, and thus in turn affect their processing of TL. Planning, viewed as one of the several processes involved in the production of both written texts and utterances, is most often discussed topic here in terms of fluency, complexity and accuracy (Foster & Skehan, 1996; Skehan & Foster, 1997; Ortega, 1995; Ellis & Yuan, 2004). Among the studies, Ellis and Yuan (2004) were the first to carry out a study of the effects of planning on L2 learners’ written narratives. Using syllables per minute and number of dysfluencies as the measures of fluency, C/T, number of different grammatical verb forms and mean segmental type-token ration as those of complexity, and error-free clauses and correct verb forms as those of accuracy, they found different planning tasks may lead to different results. While the effect sizes reflected by syllables per minute, number of dysfluencies and number of different grammatical verb forms were significantly different among the three planning tasks—pretask planning, unpressured on-line planning and no planning, other measures showed no significant differences. Therefore, they concluded that whereas pretask planning resulted in greater fluency and greater syntactic variety, the opportunity to engage in unpressured on-line planning assisted greater accuracy.

Output-related studies are more closely related to assessment conditions, e.g., time restraints (Kroll, 1990), audience (Hirano, 1991), task type (Larsen-Freeman, 1983; Lim, 1982) and topic chosen (Tedick, 1990; Reid, 1992; Tapia, 1993). These factors can directly affect test-takers’ performance in terms of fluency, complexity and accuracy. For instance, Lim (1982) analyzed syntactic features of texts produced in two tasks by 120 L2 writers at varying levels of proficiency. The objective measures used were MLC, C/T, MLTU, T/S, MLS, W/EFT and EFT/S. A comparison of compositions and rewritings indicated that the rewriting task restricted writers’ choice of sentence structure somewhat. Therefore, free compositions were proved more useful than rewritings in discriminating between proficiency levels.

#### 4. Conclusion: A Compromise

Either-or solutions are seldom seen in the field of SLA and the above-mentioned two views of L2 development are just two extremes of a continuum. Therefore, a reconcile solution is proposed here: both general rules and individual differences exist in SLA. It may be difficult to find a global index of L2 development, but we can consider about it in a less ambitious way, i.e., some objective measures that can reflect learners’ development in fluency, complexity and accuracy as well as discriminate a specific population will be regarded as valid to depict the status quo of this group. When such a set of objective measures has been found, individual performances can be gauged in terms of fluency, complexity and accuracy so as to see in which dimension he or she is weak or strong. Only in this way can L2 learners’ development be caught both holistically and analytically, and more studies need to be done by combining these two perspectives.

#### References

- Arnaud, P. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests. In Arnaud, P. & Bejoint, H. (eds.), *Vocabulary and Applied Linguistics* (pp. 133-45). London: Macmillan.
- Arthur, B. (1979). Short-term changes in EFL composition skills. In Yorio, C., Perkins, K. & Schachter, J. (eds.), *On TESOL '79: The Learner in Focus* (pp. 330-342). Washington, D.C.: TESOL.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, 26, 390-395.
- Casanave, C. (1994). Language development in students’ journals. *Journal of Second Language Writing*, 3, 179-201.

- Cooper, T. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research*, 69, 176-183.
- Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, 39, 81-141.
- Cumming, A., & Mellow, D. (1996). An investigation into the validity of written indicators of second language proficiency. In Cumming, A. & Berwick, R. (eds.) *Validation in Language Testing* (pp. 72-93). Clevedon, England: Multilingual Matters.
- Ellis, R. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Fillmore, C. (1979). On fluency. In Riggenbach, H. (ed.) *Perspectives on Fluency* (pp. 43-60). University of Michigan.
- Gaies, S. (1980). T-unit analysis in second language research: Applications, problems and limitations. *TESOL Quarterly*, 14, 53-60.
- Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *Modern Language Journal*, 80, 309-326.
- Homburg, T. (1984). Holistic evaluation of EFL compositions: Can it be validated objectively? *TESOL Quarterly*, 18, 87-107.
- Ho-Peng, L. (1983). Using T-unit measures to assess writing proficiency of university ESL students. *RELC Journal*, 14, 35-43.
- Hunt, K. (1970). Syntactic maturity in school children and adults. *Monographs of the Society for Research in Child Development*, 35, No. 134.
- Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, 4, 51-70.
- Kameen, P. (1979). Syntactic skill and ESL writing quality. In Yorio, C., Perkins, K. & Schachter, J. (eds.), *On TESOL '79: The Learner in Focus* (pp. 343-346). Washington, D.C.: TESOL.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12, 439-448.
- Larsen-Freeman, D. (1983). Assessing global second language proficiency. In Seliger, H. & Long, M. (eds.) *Classroom Oriented Research in Second Language Acquisition* (pp. 287-304). Rowley, MA: Newbury House.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners. *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, 123-34.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417.
- Loban, W. (1976). *Language development: kindergarten through grade level*. NCTE Research Report No. 18. Urbana, 111: National Council of Teachers of English.
- McKay, D. (1982). The problems of flexibility, fluency, and speed-accuracy tradeoff in skilled behavior. *Psychological Review*, 89, 483-506.
- McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11, 113-128.
- Mitchell, R., & Myles, F. (1998). *Second language learning theories*. London: Edward Arnold.
- Monroe, J. (1975). Measuring and enhancing syntactic fluency in French. *The French Review*, 48, 1023-1031.
- Ortega, L. (1995). *Planning and second language oral performance*. Unpublished MA thesis, University of Hawaii at Manoa.
- Parker, M. & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language Acquisition*, 25 (3), 1-12.
- Perkins, K. (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. *TESOL Quarterly*, 14, 61-69.
- Perkins, K. (1983). On the use of composition scoring techniques, objective measures, and objective tests to evaluate ESL writing ability. *TESOL Quarterly*, 17, 651-671.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.
- Scott, M. & Tucker, G. (1974). Error analysis and English-language strategies of Arab students. *Language Learning*, 24, 69-97.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistic*, 18, 268-286.

Wolfe-Quintero, K., Inagaki, S. & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Hawaii University Press.