

## **How to Avoid the Comparative Fallacy during Data Analysis: A Review of Doughty and Varela (1998) and Mackey (1999)**

**Andrea Révész**

The problem of the comparative fallacy, i.e., assessing the L2 learner against the native speaker in second language studies, was first addressed twenty years ago by Bley-Vroman (1983). In a critical review of the framework proposed by Tarone, Frauenfelder, and Selinker (1976) to study the issue of interlanguage systematicity and variability, Bley-Vroman demonstrated that the practice of employing analytical concepts defined in terms of the target language can seriously hinder the investigation of learner languages. Bley-Vroman showed that, as a result of the comparative fallacy, L2 studies may result in incorrect and unrevealing analyses, and may lead to inadequate description of the nature of interlanguages. Despite the early warning, SLA research has often fallen into the trap of the comparative fallacy over the past two decades. The majority of SLA studies have tended to employ blunt measures of zero-to-target change instead of using more sophisticated interlanguage-sensitive, or developmental, measures. Clearly, measures that set the target language as the sole criterion of successful treatment may result in a failure to acknowledge relevant evidence of language development. Very importantly, this is not to suggest that SLA research should cease to use the native speakers' language as one perspective on the language of L2 learners. Rather, what is proposed is that the native speakers' language should not be used as the only measure of achievement in the L2 (Cook, 1999).

This paper introduces two recent empirical studies that have made promising attempts at circumventing the problem of the comparative fallacy. Doughty and Varela (1998), along with Mackey (1999), strove to analyze data in an interlanguage-sensitive fashion, and, as a result, succeeded in providing a relatively accurate description of the change in the learners' interlanguage system. This paper intends to pinpoint the advantages of the analytical procedures these studies employed, and, in doing so, to create an incentive for future research to use such interlanguage-sensitive measures.

First, let us review the analytical method implemented in Doughty and Varela. Doughty and Varela investigated the impact of corrective recasting on the development of past-tense marking and conditional-sentence forms in the context of an ESL science class. The researchers measured participants' learning by analyzing two types of data: the learners' transcribed oral science reports and their written science reports. The coding of the data involved five steps. First, every context in which a verb was or should have been supplied was identified. Second, it was noted whether or not a verb was supplied in this context. Third, the researchers determined whether past marking was necessary in that particular context. Eventually, only those verb contexts that required past marking, along with those contexts in which there was unnecessary past marking, were kept for further analysis. After the contexts and verbs had been identified, the nature of the past marking was examined. The fourth step was thus to determine whether the past or the conditional form was used; and finally, in cases where past marking was present, it was determined whether it (a) was target-like, (b) was an interlanguage attempt at past reference, or (c) was inappropriate, meaning that some kind of time reference was obvious but inappropriate to the context. Development was determined by looking at four types of evidence: (a) decreases

in the absence of past-tense marking in obligatory contexts, (b) increased attempts at the use of past reference (including interlanguage forms), (c) temporary overproduction of past marking, and (d) increasing accuracy.

By following this procedure, Doughty and Varela avoided two instances of the comparative fallacy that are frequently present in SLA studies. First, by including unnecessary past-tense marking in the analysis, the researchers did not restrict their examination to the suppliance of past forms in obligatory contexts specified by the target language: They were also able to consider past marking provided in contexts defined by the learners' interlanguage. According to Bley-Vroman, one of the main culprits in Tarone et al.'s (1976) misleading analysis is the fact that they failed to acknowledge contexts determined by learner languages and only analyzed obligatory contexts specified by target-language norms. A second advantage of the analysis employed by Doughty and Varela lies in their assessment of the past forms produced. Instead of using a binary measure (target-like or non-target-like), they evaluated the learners' use of past reference in terms of three categories (target-like, interlanguage attempt, and non-target-like). In this way, Doughty and Varela were able to report not only shifts to target-like use and declines in non-target-like use, but also evidence of changes that showed improvement while still not reaching the full target variant. Implicit in this measure is that learning does not necessarily mean reaching the accuracy level of native speakers: Learning is any development that happens relative to a previous state of the learner's interlanguage system.

Now let us turn to Mackey (1999), another study that succeeded in conducting an interlanguage-sensitive data analysis. The study set out to explore the relationship between different types of conversational interaction and second language acquisition. It employed a pretest, treatment, posttest design, and the data collected were transcriptions of task-based interactions that were designed to elicit ESL questions. Development was measured by changes in the learners' use of question forms from the pretest to the posttest. Both the pre- and the posttests were coded for two measures of development: (a) developmental stages of participants with respect to ESL question forms (Pienemann & Johnston, 1987), and (b) different stages of questions produced. Development was operationalized as movement from an earlier stage of question formation to a more advanced stage. Unlike Pienemann and Johnston (1987), who considered two different usages of two different structures sufficient evidence for the acquisition of a certain stage, Mackey set a more stringent criterion for sustained development by requiring the presence of at least two examples of structures in *two* different posttests.

Mackey's study provides another example of how the comparative fallacy may be avoided during data analysis. She used a developmental sequence as a criterion for L2 learning, and clearly, the notion of developmental sequence itself entails that learning means gradual changes in linguistic behavior, and it does not necessarily mean reaching target-like accuracy with respect to an L2 feature. In a developmental sequence-based data analysis, any advancement from a certain developmental stage to the next is considered learning regardless of whether the accuracy of the interlanguage forms still falls short of or reaches the target language norm regarding the L2 feature.

As to the accuracy of the data analysis, operationalizing learning in terms of a developmental sequence offers clear advantages, even compared to interlanguage-sensitive, but non-developmental (i.e., measures not defined in terms of a developmental sequence), measures. First, this procedure enables the researcher to assess interlanguage forms with reference to a natural learning sequence. In contrast, interlanguage-sensitive but non-developmental measures,

for example, the category “interlanguage attempt” in Doughty and Varela, are arbitrary and thus may lead to difficult coding decisions. For example, distinguishing between non-target-like forms and “interlanguage attempts” may prove extremely challenging at times. Another advantage of using developmental sequences as reference through coding lies in the fact that they enable the researcher to make a refined decision as to the accuracy level of a particular interlanguage form. For instance, since the developmental sequence for question formation includes six stages, Mackey was able to analyze her data in terms of six categories. This is not to suggest, of course, that the more coding categories a study employs, the more refined its measure of development is going to be. The creation of additional coding categories in the absence of a developmental sequence for a particular structure may be arbitrary and easily result in incorrect and unrevealing analyses.

In summary, this paper has discussed two recent SLA studies that, by employing interlanguage-sensitive measures of development, succeeded in avoiding the trap of the comparative fallacy. Doughty and Varela, departing from common practice, did not only analyze learners’ production in target-language obligatory contexts, but also examined learners’ suppliance of target forms in contexts considered non-obligatory according to target-language norms. In this way, the researchers were able to observe interlanguage phenomena, for example, overproduction, that otherwise might have been easily overlooked. Furthermore, as a result of including an intermediate category for interlanguage attempts, neither completely target-like nor utterly non-target-like, their interlanguage analysis enhanced the chance of reporting partial learning of the target structures. Mackey’s data analysis arguably led to an even more accurate measurement, since she defined her coding scheme in terms of a developmental sequence. Analyzing data with reference to a developmental sequence seems a more fine-grained method of detecting partial learning than other interlanguage-sensitive, but non-developmental, measures. Unfortunately, however, only a few developmental sequences have been identified to date, and, therefore, the application of this analytic procedure is very limited. As a consequence, for most linguistic structures, until a corresponding developmental sequence is explored, researchers need to rely on alternative analytical procedures to escape the trap of the comparative fallacy. Doughty and Varela provide an excellent example of one such method.

## REFERENCES

- Bley-Vroman, R. (1982). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning*, 33, 1-17.
- Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, 33, 185-209.
- Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). New York: Cambridge University Press.
- Mackey, A. (1999). Input, interaction, and second language development. *Studies in Second Language Acquisition*, 21, 557-587.
- Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In D. Nunan (Ed.), *Applying second language acquisition research* (pp. 45-141). Adelaide, Australia: National Curriculum Research Centre, AMEP.

Tarone, E., Fraunfelder, U., & Selinker, L. (1976). Systematicity/variability and stability/instability in interlanguage systems. *Language Learning*, 4, 93-134.

Andrea Révész is a doctoral student in TESOL at Teachers College, Columbia University. Her research interests are task-based language learning and the role of corrective feedback in second language acquisition. She is currently working on her pilot study investigating the impact of task effects on the efficacy of implicit corrective feedback.