

ON THE RELIABILITY OF ERROR ANALYSIS

FERNANDO JOSÉ DA ROCHA

Pontifícia Universidade Católica, Porto Alegre

1. *Introduction*

The growing interest in Error Analysis (EA) indicates that its relevance for language learning and teaching has been widely accepted. But, on the other hand, one cannot be equally enthusiastic about its reliability.

One may say that the results obtained in EA studies are reliable to the extent that the methodological procedures used are objective. Nevertheless, in its present state, EA has not yet a standard methodology, and it is known that the general guidelines available are subject to criticism. It follows that a certain degree of subjectivity and arbitrariness is found in each and every study. The problem, then, lies in verifying whether the results obtained in specific studies are sufficiently good for their proposed applications that range from the simple tabulation of errors and the determination of their relative frequencies to the complete description of the nature of the inter-language used by foreign language students. It happens that each EA project requires a set of procedures and a degree of accuracy which are consonant with its goals. We have elected to concentrate our study upon the reliability of those studies which aim at being useful for the improvement of language learning and teaching.

Some works are useful because they provide the kind of data language learning and teaching specialists expect from EA, that is, a list of errors found to be common in a defined stage of language learning, their classification according to their linguistic description and causes (linguistic and psycholinguistic), and their quantification with an eye to establishing their order of importance.

Other studies are important because they treat certain theoretical aspects which are relevant to EA, or serve as partial models for further research.

Arabski (1968) and Duskova (1969) provide us with good examples of this kind of work. Arabski conducted an introductory qualitative study of lexical and grammatical written errors made by Polish students of English. The only analysis he made was that of the errors which could be readily traced to a linguistic cause. He did not attempt to supply any other information (not even numerical) as he was conscious that his corpus consisting of 200 entrance examinations was insufficient for a valid statistical treatment. Nevertheless, his work is valuable for the interesting criterion for the classification of errors he introduces, and the theoretical claim he makes that the pedagogical component of the students' background does not influence the errors they make.

Duskova's "On sources of error in foreign language learning" has a place of importance in European EA, thanks to the answers she found to questions related to the status of Contrastive and Error Analyses, to the justification of Corder's distinction between mistakes and errors, and to the difference between production and reception errors in foreign language learning. Her study is more interesting, we think, when she analyses the role played by interference factors, besides the native tongue. On the other hand, the corpus she used was limited to 50 compositions, which assigns the character of a probe to the remaining aspects of her research.

2. *EA for language learning and teaching specialists*

The kind of data the language teacher, the text book writer and the psycholinguist expect from EA demands a full and accurate account of errors and their genesis. It is the degree of completeness and accuracy, in other words, the degree of reliability found in the current EA studies that constitutes the focus of interest of the present paper. For the sake of exposition, we found it advisable to treat each phase of Error Analysis separately, as if it constituted a whole in itself.

2.1. *Corpus composition*

Due to their nature, the corpora for written error analyses have been composed in a rather classical way. However, a whole series of techniques have been employed by error analysts to collect their data. Multiple choice tests, translations, free compositions, subject bound compositions, essays, personal correspondence, and examination papers rate among the most popular ones. Some of the more recent studies show an interesting tendency. They combine two or more techniques, such as multiple choice tests and free composition (Buteau 1970), or Cloze procedure tests and multiple choice tests (Jackson 1971). This interest reflects a due concern about the representativity of the corpora used in error studies.

It should be pointed out, however, that the quality of the information supplied by each of the above listed techniques is not homogeneous. Not all of them test the same aspect of the student's performance. Composition, for instance, deals with the learner's performance at the production level, while translation reflects performance at the reception level. As errors in production and reception vary in quantity and quality, studies which focus on different aspects of the students' usage of the language cannot produce similar results. The studies conducted by Duskova (1969) and Routhledge (1972) supply empirical evidence for this. Buteau in her 1970 study reports that the results obtained from compositions and objective tests diverge. She found a qualitative difference of 15% between the results provided by these two techniques. It is unfortunate that she did not measure the degree of quantitative divergency.

As the above examples show us, the performance of the students is, to a greater or lesser extent, influenced by the error collecting techniques. The shortcomings of these procedures were set forth by Chaplen (1973), Nickel (1971:192) and Corder (1973:268) among others.

It is commonly agreed among the practitioners of the art that EA should be performed on language material which is spontaneously produced by the learner with the intention of communicating. Essays and compositions are often considered as the best sources. On the other hand, there has never been any sort of agreement as to how much material a relevant corpus should contain. No frequency count has ever been made to indicate the number of sentences, or errors, necessary to compose a general corpus. One educated guess would be that the larger the students' syllabus, the larger the corpus should be. However, practice shows that there is no proportion between the level of proficiency of the students and the volume of the corpus. In fact, corpora have been built according to individual and subjective principles. Some researchers were content to make a linguistic and statistical analysis of a corpus which had no more than 40 compositions. Fortunately other scholars, Castelo (1962) and Aguas (1970) for instance, were more aware of the problems posed by such studies. A good example was set by Ruiz (1963), who made a longitudinal analysis of the errors found in the 20,124 verb occurrences registered in his corpus of 1100 compositions.

However, it is not only the *number* of compositions which accounts for the adequacy of a corpus. Composition, due to its own nature, poses difficult problems too. Recent studies (e.g. Chaplen 1973) have shown that compositions are not capable of providing measurements of learner control of structure, lexis and usage which are sufficiently reliable: they supply no evidence of the sample. In fact, a composition may represent either a good or a poor, a large or a small sample of the student's knowledge.

The kind of composition requested, together with the student's individu-

ality, may also produce wide variation in the number and type of errors. Some students may successfully avoid using all those aspects of the language about which they do not feel confident. The result is that they produce error free compositions, but of low structural quality. Others concentrate on the content of their compositions, and thus make a less strict selection of the items which they do not know well. Consequently, they are likely to commit more errors than the students in the first case. The tendency to avoid structures which one rightly or wrongly believes one has not learned well is present in every student. Since each student reacts to this tendency in his own way, the researcher may be misled when trying to evaluate the student's performance.

Since chance correction is more likely to occur in a small corpus than in a large one, the size of the corpus may be related to the tendency to avoid errors.

One last point to be made in this section pertains to the material of which the corpus is composed. In a single composition nobody will display all he knows and all he doesn't know about a given language. The fact that a student has mastered the operation of a system within a defined range of contexts does not necessarily imply that he can use it correctly in a wide range of contexts.

2.2. *Detection of errors*

Relevance and reliability don't always go together. When the norms for correctness were based on prescribed rules and the logic of grammar, there was a common criterion of judgement. It was objective, and thus reliable, but not relevant according to the principles set forth by the usage movement in the late twenties.

Since it has been accepted that the basic criterion for correct usage should be the consensus of educated people, the reliability of error detection has decreased; the very concept of error has become vague. This is due to the influence of the subjectivity and the degree of prescriptiveness of the correctors. We may say after Strevens (1969 : 5—6) that it is possible for two native speakers to differ, in a surprisingly large proportion of cases, as to whether items are acceptable or unacceptable, and hence as to whether they should be counted as errors. Some of the more elaborated EA studies have largely neutralized this effect by submitting their material to a sufficiently varied number of correctors. The Colorado Project researchers (1963) for instance, besides asking 12 college professors to correct their corpora, counted as acceptable or nonacceptable only those structures which had the agreement of at least 3/4 of the judges. The others being disputable, and consequently subject to arbitrariness, were not taken into account.

2.3. *Linguistic categorization*

The fact that linguistic categorization is a difficult task is known to every one who has tried a hand at EA. The so-called multiple errors may be classified in two or even three valid ways; other errors defy categorization in any linguistic way. However, for practical purposes, the classification of multiple errors does not constitute a very serious problem. The descriptive framework proposed by Halliday (1964: 119) opens way to a good solution.

Due to their high frequency of occurrence, some unclassifiable errors seem to deserve special attention. Duskova (1969) calls them "nonce errors", and claims that they are unique in character, non recurrent and not readily traced to their sources and therefore of no importance for EA. But we cannot help finding it surprising that only 34% of the preposition errors found in her corpus were worth being further studied. Unfortunately, she does not make any other comment on this subject.

2.4. *Explanation of errors*

Politzer (1973: 58), based on Dulay and Burt, reports that "to pin down the precise cause of an error inevitably involves some guessing about the underlying psychological process". We may continue and say that the explanation of errors is probably the most vulnerable area of EA.

Some optimistic authors claim that interlingual errors can be rather easily detected, and explained by contrastive methods. It happens that only one to two thirds of the total number of errors were traced to this cause in the studies carried out by Richards (1971), George (1972), Dulay and Burt (1972); all others were either intralingual, performance or nonce errors. These three kinds of errors have a common characteristic: they are not always traceable to a particular linguistic area. As can be expected, this leads to debatable, if not arbitrary, classifications to which no solution is yet envisaged.

2.5. *Statistical treatment*

The original assumption that the frequency of errors reveals their degree of difficulty has been seriously questioned. But, even so, the frequency of errors continues playing an important role, mainly in the arrangement of materials in an appropriate order for practical teaching purposes, that is, for grading.

We have observed that many researchers provided very little statistical information. Most of them limited themselves to establishing the absolute and the relative frequencies of errors, that is, they made a count of each kind of error and calculated the percentage it constituted of the total number of errors. The fact that these procedures cannot give a real picture of the im-

portance of the errors found in a given corpus was pointed out by Stormzand in 1924. The notion of error quotient is the result of his search for a measuring instrument which would not exaggerate the importance attributed to certain items, nor underestimate the seriousness of others. It may be described as being the ratio of the actual number of wrong occurrences of an item to the total number of opportunities for that item to occur.

Nevertheless, other statistical possibilities seem to be overlooked. Not many European Error Analyses established the ratio between the number of errors and the number of words per sentence, or the ratio between the total number of words in the corpus and the number of errors. This means that important pieces of information have been neglected, to the detriment of the reliability of EA.

3. *Conclusions*

The material we analysed shows that an EA which aims at providing information for language specialists is not so economical and pragmatic as some researchers would like it to be. The following are some of the reasons which lead us to this conclusion.

3.1 The corpus used should constitute a fair sample of the students' performance, and be representative of the level of learning being analysed. This means that the generalizations derived from the final results should hold true for all students in similar learning conditions. The corpus should also be large enough to remain meaningful after the unavoidable loss of information that takes place in the detection and categorization of errors.

3.2 We believe that for the sake of reliability, subjectivity should be avoided as much as possible: dubious cases should not be taken into account; the detection of errors cannot be based upon prescribed rules, nor be the product of a single person's judgement.

3.3 The importance of the model adopted for the linguistic description of errors should not be underestimated. Superficial description of errors as omissions, additions, wrong selection or ordering has little explanatory value. Corder (1973 : 277) tells us that "linguistic explanations... imply a linguistic theory in which the notion of deep and surface grammar play a part". The fact that Duskova could not classify 66% of the preposition errors found in her corpus seems to be an important argument for the use of more explicit language models in EA studies.

3.4 The relevant data for EA cannot be reduced to the wrong occurrences only. The correct ones play an important role both in the categorization of the stages of learning (which Corder (1973 : 271) calls stage of random guess-

ing, stage of systematic errors and stage of post-systematic errors), and in the determination of error quotients.

3.5 More sophisticated statistical treatment is available, and there is no reason why the improvements they offer should be neglected.

We hope to have provided here some elements which will prove helpful to the evaluation of the degree of reliability of EA studies, and so avoid some sterile and misdirected work. After all, there seems to be no point in conducting projects which are too elaborate to be of use for the language teacher who intends to prepare remedial work for his students, and which, on the other hand, are not reliable enough to be of any use for language teaching specialists.

REFERENCES

- Aguas, E. F. 1970. *English composition errors of Tagalog speakers and implications for analytical theory*. Unpublished Ph.D. dissertation, University of California, Los Angeles.
- Arabski, J. 1968. "Linguistic analysis of English composition errors made by Polish students". *SAP* 1. 71-89.
- Buteau, M. 1970. "Students' errors and the learning of French as a second language". *IRAL* VIII. 133-145.
- Castelo, L. M. 1962. *Structural differences between English and Tagalog verbs: a study designed to improve the teaching of English to advanced Filipino students*. Unpublished Ph.D. dissertation, Columbia University.
- Chaplen, E. F. 1973. Extracts from *An appraisal of the Syndicate's written examination in English as a foreign language together with some suggested modifications*. Internal paper, University of Cambridge, Local Examinations Syndicate.
- Chau, T. T. 1972. *The concept of difficulty in second language learning and teaching*. Ph. D. thesis, University of Toronto.
- Corder, S. P. 1973. *Introducing applied linguistics*. Aylesbury, Bucks.: Hazell, Watson and Viney.
- da Rocha, F. J. 1973-1974. "Error analysis: a working bibliography with special emphasis on English". *Cahiers de Linguistique* II/1-2. 87-97.
- Dulay, H. C. and M. K. Burt. 1972. "Goofing: an indicator of children's second language learning strategies". *Language learning* 22. 235-253.
- Duskova, L. 1969. "On sources of error in foreign language teaching". *IRAL* VII. 11-36.
- George, H. U. 1972. *Common errors in language learning*. Rowley, Mass.: Newbury House.
- Halliday, M. A. K. and A. McIntosh. 1964. *The linguistic sciences and language teaching*. London: Longmans, Green and Co.
- Hunn, B. E., Silvers, S. L., Ribbens, D. D., Nebel, D. A. and R. W. Speer. 1963. *The Colorado project*. Unpublished dissertation, Colorado State College.
- Jackson, K. 1971. "The verification and comparison of contrastive analysis". *The PCCLLU papers* 3. 201-210.
- Nickel, G. 1971. "Variables in a hierarchy of difficulty". *The PCCLLU papers* 3. 185-194.

- Politzer, R. L. and A. G. Ramirez. 1973. "An error analysis of the spoken English of Mexican-American pupils in a bilingual school and a monolingual school". *Language learning* 23. 39-63.
- Richards, J. 1971. "Error analysis and second language strategies". *Language sciences* 17. 12-22.
- Routhledge, J. L. 1972. *An analysis of the errors in the written English of advanced Portuguese learners*. Unpublished dissertation, University of Lancaster.
- Ruiz, M. B. 1963. *Weighting and sequencing English tense-aspect modifications for Hili-gaynon speakers*. Unpublished Ph. D. dissertation, University of California, Los Angeles.
- Stormzand, M. J. and M. U. O'Shea. 1924. *How much English grammar?* Baltimore: Warwick and York.
- Strevens, P. 1969. *Two ways of looking at error analysis*. Washington, D. C.: ERIC Document Reproduction Service.