

TESTING AND CONTRASTIVE ANALYSIS

NAUM R. DIMITRIJEVIĆ

University of Belgrade

Though visible efforts have been made to introduce objective measures into research in the humanities for some time the introduction of such measures and methods into certain areas of applied linguistics has proceeded fairly slowly. Not only is little attention paid to the possibilities offered by certain (non-linguistic) disciplines in the objectivization of research methodology in applied linguistics, but it also seems that in some cases basic methodological principles of research are being overlooked. Namely, subjective criteria in defining language tests are still sometimes in use; information about some important variables in research are not presented.

It goes without saying that only objective measures should be used in modern research, or rather as objective and reliable as possible in research on human behaviour. Nevertheless, in some large-scale projects in applied linguistics, generally multidisciplinary in character, there seems to be a kind of imbalance in the scientific approach adopted. Whereas on the one hand, highly refined analyses, objective and appropriate to the material and aim of the project, are applied (these are as a rule linguistic descriptions or analyses) an approach which lacks the necessary scientific rigour is adopted in the treatment of other closely connected problems. For instance, while in CA as well as in EA the linguistic analyses are often refined, explicit, objective etc., the testing of contrastive analysis hypothesis (CAH) seems to lack the same degree of scientific rigour and explicitness.

What we would like to do in this paper is to point some of these methodological requirements in the area of testing of CAH.

After a period of intensive work on CA of Ls and Lt at the level which could be called more linguistic than methodological, a period without much, or with little experimentation and testing, we have reached a point now when hypotheses have to be proved. We should no longer be content with elegantly phrased hypotheses or claims. A theory or a hypothesis, formulated

with the aim of being applied in practical situations, has to be proved; we want to see whether it works or not. One way of doing this is by testing the theory or hypothesis.

Contrastive analysis of Ls and Lt, with the final aim of application in teaching, can have different phases. For instance:

CA ———→ prediction of difficulties ———→ errors ———→ formulation of hypothesis about the errors ———→ testing of the hypothesis ———→ writing pedagogical material.

Of course, this is not the only possible order of steps; Whitman has four steps in his analysis and no testing. His procedure includes the following steps:

Description ———→ Selection ———→ Contrast ———→ Prediction (Whitman 1970).

At this time it is not our intention to discuss the possibility of predicting students' errors or linguistic behaviour on the basis of CA of Ls and Lt. As is well known, there are 'contrastivists' and those who doubt, in varying degrees, the possibilities of predicting student errors by means of CA. Our aim is simply to draw attention to the importance of testing CAH as a step in CA.

First of all, it should be made quite clear that testing cannot and should not be excluded from CA projects. It constitutes an extremely important link between the initial theoretical step to the final one — the application of CA results to teaching. The importance of testing a hypothesis is obvious and does not require explanation. However, as has been said much more attention is paid to the theoretical part of the analyses than to the practical testing of the assumption. And if there is any testing it is sometimes done without scientific rigour.

In order to prove their hypotheses some authors construct tests to trap students, tests which "... can be criticized because they were designed specifically to catch the errors the analyses predicted and no other errors" (see Whitman, Jackson (1972:29). This kind of test has only 'surface validity' and no value (either theoretical or practical) at all. Consequently, one can doubt the conclusions arrived at on the basis of such tests.

What requirements, then, should be met in order to make the testing of CAH reliable and sound, relevant to its aim and the material tested? In our brief survey of the problems of the testing of CAH we would like to discuss the following:

- the test, its psychometric characteristics, form and content.
- the sample of population to which the test is administered.
- teaching methods and techniques used with the sample of population.
- the methods used to analyse the obtained results.

Before we proceed some general remarks on language testing must be made.

In testing students in L2 (achievement and proficiency tests) we are *not* trying to discover the *cause* of students' errors. Our aim is only to find out what he knows. In a test which is supposed to help us conclude something about the origin of these errors we are operating in an area where we should know much more about language testing than we do today and also much more about language learning. This kind of testing may be objective but only to the extent of our present knowledge and there is a lot more to be learned about language learning and testing than some people think.¹ We should be more cautious therefore about the conclusions we draw on the basis of 'objective' testing.

When we start testing a CAH it means we believe we can test this hypothesis in an objective and reliable way. It seems to us that we are trying to test a hypothesis by using another hypothesis which has yet to be proved. We shall discuss this problem of testing a CAH with the assumption that this is possible or, rather, we shall be talking about the problems that those who are engaged in this kind of testing have to solve first.

First of all it must be stressed that in spite of great advances in applied linguistics we still lack an adequate theory of foreign language testing. There are wide gaps in certain practical as well as theoretical aspects of language testing, mainly in test construction and its validation. But it is also true that we do have enough knowledge about language and test construction in general in order to prepare measuring instruments which will certainly be more objective and reliable than subjective criteria. Furthermore, modern statistical procedures can help us evaluate correctly the different results obtained on tests, which would be impossible if only raw scores were studied.

Now after this introductory warning about the incompleteness of testing theory let us see what kind of requirements must be met by the tests in the testing of CAH.

One of the main requirements of a good test is its validity. It is one of the central issues in language testing because we still do not have a satisfactory answer to the question: what does it mean to learn and know a foreign language? The answer to this question is much more complex than it may appear at first sight. This is a question we must answer before writing, for instance, achievement or placement test, a task which is much simpler than one we are dealing with here. What we are trying to do in testing a CAH is to obtain students' responses in Lt which will necessarily include incorrect responses, and in such a way as to be able to make sound conclusions about the *origin* of these mistakes.

The validity of a measuring instrument in testing a CAH must be defined first. Investigators must inform us how they have established the validity of

¹ Cf. Corder (1974).

their instruments. And it is not sufficient to use the least reliable type of validity, so-called 'face validity'. A test we use in CAH should have a precise coefficient of correlation. We must be sure (or at least as much as we can today) that we are measuring what we want and not something else.

A problem which is easier to solve than validity is establishing the reliability of a test. The reliability of a test must also be expressed in exact terms and should never be left to the teacher's or investigator's subjective feelings. Although considerable attention is paid to this requirement in testing in general and thus in language tests too, it seems to be neglected in the area of testing CAH. Without the coefficient of reliability we cannot accept the results of any CAH testing, regardless of how refined linguistic analysis may have been.

In order to show us that the hypothesis was correct some investigators give percentages of erroneous responses to a test item and say that over 90% of students made a certain mistake. First of all, we cannot accept percentages or raw scores alone in an analysis of test results. They must be processed/computed by means of appropriate statistical procedures. Secondly, one has little faith in the soundness of a test item which is solved by over 90% of students. It would be the same if the test item were answered by only 5 - 10%. In the first case it is probably too easy and in the latter too difficult. Therefore, as with validity and reliability we should have exact data on the difficulty and discrimination of tests as whole and of individual items as well.

Testing a CAH need not be practical or economical because it is a part of research, and the main point in this kind of testing is achieving an aim, regardless of cost or time. Therefore, the question whether the test will be oral or written, whether we shall insist on students' encoding or decoding, is of no importance if the test works. As in all other areas of testing CAH the adequacy of the medium has to be tried out and tested. There is a great difference between asking a student to respond orally or in writing, particularly in tests which propose to discover errors (not correct responses), on the basis of which conclusions are made about the origin of these errors.

One of the principal problems of testing a foreign language yet to be solved is that of determining a basic approach (not just techniques) to the elicitation of students' responses. Two main approaches can be discussed: an integrative and discrete item approach, while a third is being developed. None of these meets the requirements of language testing, the testing of L2 in general, let alone the very specific kind of testing we have in testing a hypothesis like CAH.² Thus only experimentation will tell us which approach will be the most appropriate to the given purpose of testing CAH.

² For an analysis of tests using different structures and design, cf. Whitman, R. L., K. L. Jackson (1972).

A very fine linguistic analysis may be wasted in a couple of 'minor' methodological issues, such as the techniques of elicitation of students' responses. In research on human behaviour one must be very careful with the classification of variables, namely labeling some of them as crucial and others as peripheral unless there are extremely solid grounds for such divisions.

The actual techniques used to elicit students' responses are of extreme importance for the reliability of the final results obtained on a test; we would like to stress the particular importance of this in tests used to prove a theoretical assumption. Though there are numerous testing techniques we still lack certain insights of relevance to the testing of CAH.³ Thus we do not know whether there are testing techniques which could be labeled "neutral", neutral as regards the students' familiarity with it. The less artificial a technique is the more neutral it will be. Therefore, it seems to us that we should aim towards the use of such techniques in language testing in general and particularly in research.

There is not enough research in this area of applied linguistics and without the exact correlations we shall be operating with instruments whose nature we do not understand very well. In such a situation the results obtained must be accepted with due caution.

For instance some claim that translation from Ls into Lt is the best or the most appropriate technique for testing a CAH; others maintain that better results are achieved if multiple-choice items are used. For a reliable assessment we cannot be satisfied with statements like these (though there are a few methodologically sound research projects in this area).⁴ What we need are exact correlations not only for a couple of techniques (translation vs. multiple-choice) but for many more of them perhaps trying a combination of translation from Ls into Lt and multiple-choice items.

The problem of elicitation techniques in language research is not a peripheral one, though some may think so. How a student will react depends, obviously, on what he is exposed to and on the situation we place him in. If we need students' reactions to prove a hypothesis this proof should be well grounded, which again depends on the use of appropriate research methodology as mentioned at the beginning of this paper.

The content of the language course taken by our subjects is also an important variable for the final assessment of the results. For example, if a particular structure is not practiced, is not given adequate attention, we can predict a certain number of errors in the use of that structure. The opposite is also true. If a structure, very different in Lt from its corresponding one in

³ The only fairly exhaustive, but not complete, list of testing techniques is "Sample Test Items" by K. Radovanović (1974).

⁴ Cf. for instance, Whitman R. L., K. L. Jackson (1972).

Ls, is given ample time and practice in the teaching, the results may be contrary to the hypothesis (that students will make a lot of mistakes, owing to the difference between the two structures).

Therefore the test must reflect the content of the course not only in terms of presented material but also in terms of the time devoted to the material presented during the course. In other words, the selection of the material for the test which tests a CAH is of special importance, we would say of greater importance than for an achievement or proficiency test.

In conclusion of this brief discussion of the measuring instrument used in testing CAH the following can be said:

In CA and in testing CAH much more attention is being paid to linguistic and language analyses of the material than to certain psychometric and methodological aspects of the problem. It is commendable that investigators are concerned about the model of the description which is to be used; obviously, without an adequate model and its adequate application the rest of the analysis could easily lead us astray. However, if the measures used in testing hypotheses are not adequate, then again, but in another way, the final results will blur the picture we are trying to bring into focus.⁵

The sample of population on which a CA, or any other, hypothesis will be tested must be very carefully selected. Every report on the results of CAH testing should include information on the population studied. In any language experiment, and testing is a kind of experiment, students' knowledge of the language is of evident importance. Both the researcher and the reader of the report should have as clear an understanding as possible of the level of the students' knowledge of Lt. It is only too natural that this variable affects the subjects' reactions to the stimuli in the test and in that way directly influences the results of the test, which in turn is connected with the main point of the report — accepting or rejecting the hypothesis.

However, this variable, like a few others, mentioned here, seem to have escaped the attention of some of those engaged in testing CAH. Either no information is given at all or only enough to give serious doubts about the value of the testing (without further analysis of the results).

The students' language proficiency may be defined in a very vague and subjective manner, for example, the "students studied Lt for four years" or "our subjects were from an intermediate/advanced level of instruction". Obviously, this kind of information does not mean much to anybody, either to those who read the reports or, which is even more important, to the researchers themselves. One can interpret errors, their significance, nature etc., only if the stu-

dents' general knowledge of the language is known. In other words, the same error does not carry the same weight if made by a 'beginner' or a student at a very 'advanced level of instruction'.

Because of the importance of this information on the students' language ability, standardized language tests should be used for a precise 'placement' of students. If this kind of measure is not available and locally produced tests are administered, detailed information about the test should be supplied. In addition to this information other subjective factors may add to the meaning of the non-standardized test, e.g. students' grades,⁶ the set books used in the course (this may be more meaningful if well-known textbooks are used). These measures, subjective as they are, may, together with the test, give a much better picture of students' language proficiency than just the number of years.

In short, errors which students make on a test used to test a hypothesis (in our case CAH) are meaningful only in relation to their knowledge of the language being tested, or to the level of their interlanguage.

Knowledge of Lt is not the only variable connected with the subjects in a language experiment of this kind. The type of school, students' motivation to learn the language etc., can and indeed do, influence the results of any kind of testing.⁷

The results of a test may also depend on the students' familiarity with a particular teaching (not only testing) technique or method, which may be similar or even the same as testing technique. For instance, familiarity with translation from the mother tongue, as a testing technique, can definitely influence the results of a test which includes translation from Ls into Lt. Of course, the opposite is also true. Students who have never translated from Ls may show poorer results on a test which requires translation. Here we should not forget the fact that we are dealing with *students*, with those whose knowledge of Lt is in constant change.

Clearly, the teaching techniques used with the subjects taking part in CAH are of considerable significance. Nevertheless some projects lack this information and thus leave the reader in a state of doubt as to the reliability of the whole experiment.

In a general evaluation of the results obtained from testing CAH the time when the students are tested plays a very important part. There will certainly be a difference in the results, in terms of the kind and number of errors, immediately upon the presentation of a particular language item and after a cert-

⁶ If grades are given, they should be accompanied by a description of their value, see Harris (1969:84).

⁷ How important motivation can be in language learning, see Gardner, R. C., W. E. Lambert (1972).

⁵ It is surprising that authors of books on testing do not, as a rule, discuss the problem of the testing of CAH; see, for example, Lado (1961); Valette (1967); Harris (1969); Clark (1972); Heaton (1975).

ain period of time.⁸ In order to understand these errors we should know how they correlate, what the differences of occurrences signify etc. We do not have such data as yet and this poses important question. Is there a period of time which the student must be given in order to assimilate new Lt material, a period during which he himself is hypothesizing about the Lt systems, finding solutions and "fixing" them, all this before we test our CAH? During this period of hypothesizing the student is necessarily making mistakes and making them not only under the influence of Ls, but other mistakes as well. He makes all kinds of inter and intralingual errors, etc.⁹

If, during 'hypothesizing period' the student necessarily makes mistakes how can we be sure which are those mistakes for which we can claim to be the result of Ls negative influence? Perhaps it would be safer to say that such mistakes are made under the influence of Ls only after a period of assimilation has passed. This question, like certain other areas of errors analysis has yet to be clarified. However, this does not mean that we should not be experimenting and looking for the right answers.

Let us summarize the main points of our paper:

— Outlining some difficulties and methodological problems of testing a CAH it was not our intention to dispute the usefulness of CA in foreign-language learning/teaching. It definitely has its role and place in methodics and even in methodology. How predictive it may be remains to be discovered.

— More objective and more powerful tests and other kinds of measures should be used in applied linguistics research in general and particularly in projects and experiments whose aim is to prove a theoretical assumption.

— Some measuring instruments are used, even in large scale projects, under the name of tests without having the necessary characteristics of a true test. In this way they only pay lip service both to the researcher and to the whole discipline.

— Testing CAH is of great importance both for the practical aspects of foreign-language teaching and for a more precise and explicative formulation of certain theoretical assumptions about foreign-language learning.

⁸ The negative transfer from Ls to Lt is not the same at all language levels. Most researchers are concerned about syntactic and phonological analyses. Yet it seems to us that negative transfer is just as strong at the lexical level but it has not been recognized. Ls lexis affects Lt in a more subtle way though. A research into the problem of the intensity of Ls influence at different language levels should constitute a part of any broadly conceived project of CA and CAH testing. It would be very good if we had information not only about the intensity of the errors, their frequency but their relative importance in communication too.

⁹ There are different types of incorrect uses of Lt (see the classification Corder makes, Corder (1974)). However, when the errors are supposed to be the result of the negative influence of Ls one generally speaks of 'errors', without making any classification.

— If tests are to be used to verify (confirm or discard) a CAH they should conform to standard psychometric and methodological requirements, which, among other things, include the following:

a) exact statements about the validity, reliability, objectivity and difficulty of the instrument used in the project/testing;

b) deeper insights into validity of language tests which measure a CAH are essential in order to give a final formulation of a test which would correspond to its aim. We lack such insights and experimental research should be undertaken in that direction too;

c) a detailed description of the sample of population on which the hypothesis is tested;

d) a detailed description of the teaching techniques and methods used with the sample;

— Further development and elaboration of elicitation techniques are needed as one of the basic conditions in any kind of experimental work and thus in the testing of CAH.

— The correlation of different testing techniques must also be established;

— The results obtained from testing a CAH should be processed according to standard statistical procedures and should not be left to the subjective evaluation of the teacher/researcher.

— For a more complete understanding of students' errors, their origin and nature we should know more about the nature of foreign-language learning, regardless of the tests we use in proving or rejecting an assumption or hypothesis. Today, we seem to lack such knowledge.

— Instead of insisting on the linguistic aspects of these errors it would be better to aim at a more complete CA by continuing our efforts in developing objective measuring instruments and gaining a deeper understanding of foreign language learning. Linguistic analysis alone is no longer sufficient.

Applied linguistics, which includes CA and EA as well as their testing, will establish itself as a scientific discipline only if principles of scientific research are strictly followed, if experimentation adheres to the rules of objective observation and verification. The procedures we use must be not only explicit but also explicative.

It is our firm belief, founded on work done in the recent past and even more so on what is being done today that applied linguistics and methodics will definitely gain the status they deserve alongside other linguistic disciplines, and will be accepted as such even by the exclusive supporters of theoretical disciplines. It is up to us (if I may paraphrase the words of a well known applied linguist) not to others, to achieve this.

REFERENCES

- Alatis, J. E. (ed.). 1968. *19th annual round table contrastive linguistics and its pedagogical implications*. Washington, D. C.: Georgetown University Press.
- Allen, J. P. B., Corder, S. P. (eds). 1974. *Edinburgh course in applied linguistics, Vol. 3, Techniques in applied linguistics*. London: Oxford University Press.
- Clark, J. L. D. 1972. *Foreign language testing: theory and practice*. Philadelphia: The Center for Curriculum Development.
- Corder, S. P. 1973. *Introducing applied linguistics*. Harmondsworth: Penguin.
- Corder, S. P. 1974. "Error analysis". In Allen, J. P. B., Corder, S. P. (eds). 1974. 122 - 154.
- Davies, A. (ed.). 1968. *Language testing symposium, a psycholinguistic approach*. London: Oxford University Press.
- Dimitrijević, N. R. (ed.). 1974. *Teach and test your students*. Beograd: N. U. "Veselin Masleša".
- Duškova, L. 1969. "On sources of errors in foreign language teaching". *IRAL* 7.1. 11 - 36.
- Filipović, R. (ed.). 1974. *Pedagogical materials 2*. Zagreb: Institute of Linguistics, Faculty of Philosophy.
- Gardner, R. C. Lambert, W. E. 1972. *Attitudes and motivation in second language learning*. Rowley, Mass.: Newbury House.
- Good, C. V. 1959. *Essentials of educational research*. New York: Appleton-Century-Crofts.
- Guilford, J. P. 1956*. *Fundamental statistics in psychology and education*. New York: McGraw-Hill.
- Harris, D. P. 1969. *Testing English as a second language*. New York: McGraw-Hill.
- Houston, J. B. 1975. *Writing English language tests*. London: Longman.
- Lado, R. 1961. *Language testing*. London: Longmans.
- Mužić, V. 1973. *Metodologija pedagoškog istraživanja*. Sarajevo: Zavod za izdavanje udžbenika.
- Nickel, G. (ed.). 1971. *Papers in contrastive linguistics*. London: Cambridge University Press.
- Nickel, G. (ed.). 1974. *Association de linguistique appliquee third congress, Copenhagen, 1972, Proceedings, Volume I: Applied contrastive linguistics*. Heidelberg: Julius Groos Verlag.
- Pickett, G. D. 1968. "A comparison of translation and blank-filling as testing techniques". *ELT* 23.1. 21 - 26.
- Raabe, H. (Hrsg.). 1974. *Trends in contrastiver Linguistik, Band I*. Tübingen: Verlag Gunter Narr.
- Radovanović, K. 1974. "Sample test items". In Dimitrijević, N. R. (ed.). 1974. 11 - 66.
- Richards, J. C. (ed.). 1974. *Error analysis, perspectives on second language acquisition*. London: Longman.
- Ritchie, W. C. 1967. "Some implications of generative grammar for the construction of courses in English as a foreign language". *Language Learning* 17. 1 - 2. 45 - 69, and 17. 3 - 4.
- Townsend, J. C. 1953. *Introduction to experimental method*. New York: McGraw-Hill.
- Valotte, R. M. 1967. *Modern language testing*. New York: Harcourt, Brace and World.
- Vilke, M. 1974. "Engleski kontrastivni projekt i nastava engleskog jezika". *Strani jezici* 3. 3 - 12.

- Whitman, R. L. 1970. "Contrastive analysis: problems and procedures". *Language Learning* 20. 2. 191 - 197.
- Whitman, R. L., Jackson, K. L. 1972. "The unpredictability of contrastive analysis". *Language Learning* 22.1. 29 - 41.
- Wilkins, D. A. 1972. *Linguistics in language teaching*. London: Edward Arnold.