PROBLEMS IN ERROR ANALYSIS

Lewis Mukattash

The University of Jordan

1. INTRODUCTORY

The aim of this paper is twofold. First, to present a brief description of a research project on error analysis that was begun at the University of Jordan in 1976, and secondly to list some of the linguistic problems and other problems in methodology and analysis which the present writer has encountered while engaged in this research project.

My work in error analysis arose in the first instance from a recognition that Jordanian students of English progressively establish wrong linguistic habits, and that by the time they reach the university they have acquired a good number of errors. It also arose from a recognition that many of these errors are systematic and form definable patterns, and that a study of such errors may have both theoretical and pedagogic significance. To be more precise, my work or error analysis started out with the same aim as most similar investigations: i.e. an inventory was to be made of the grammatical errors found in the English of Jordanian students in order to form a basis for more efficient English-language teaching in Jordan, in particular at the post secondary level. I was first interested in the English produced by first year students at the University of Jordan, but at a later stage the research project was extended to study the performance of Jordanian students in secondary schools and teachers' training institutes.

2. SCOPE OF COMPLETED RESEARCH

2.1. PHASE (1): PILOT PROJECT

The research was begun in January 1976 and completed in September 1977, and it consists of four studies that were carried out independently of each other, and thus may be considered to complement each other.

2.1.1. *First Study*

This study (Mukattash, 1980) is based on the analysis of 6000 Yes/No questions produced in writing (in January 1976) by 600 first year students at the University of Jordan in an attempt to discover what mistakes they make in forming such questions.

Analysis of the data showed that 1516 questions were erroneous, either ungrammatical or inappropriate. Some students made the mistake of producing *Wh*-questions instead of Yes/No questions. The number of such *Wh*-questions was found to be 383, 300 of which were ungarmmatical. All the *Wh*-questions were ignored and this leaves us with 1133 erroneous questions: 18.9% of the overall number of questions. Certain erroneous questions contained more than one error and the overall number of errors was found to be 1237.

All errors were categorized and described. An attempt was also made to account for the source of errors including interference from the mother tongue. The following table shows the type and frequency of the errors described in this study:

Table I

Type of error	Number	Per cent
1. *DO* redundant	127	10.3
2. *BE* redundant	73	6.0
3. *DO* replacing *BE*	24	1.9
4. *BE* replacing *DO*	115	9.3
5. Verbal form	343	27.7
6. Tense	117	9.4
7. Word order	190	15.3
8. Inversion without *DO*	36	2.9
9. Concord	29	3.3
10. Auxiliary Deletion	57	4.6
11. Miscellaneous	126	10.2
Total	1237	100%

2.1.2. *Second Study*

The findings described in this study (see Mukattash 1978b) are based on a preliminary analysis of 4000 *Wh*-questions produced in writing (in June 1976) by 400 first year students at the University of Jordan, (J. U.).

Of the 4000 questions, 1592 were found to be erroneous (either ungrammatical or inappropriate), and the overall number of errors was found to be 1790.

This study is mainly concerned with two issues, namely: word order and auxiliaries. Other errors like verbal forms, concord and auxiliaries are listed

but not discussed since they are discussed in the preceding study. All errors involving word order and auxiliaries are discussed in detail. An attempt is also made to account for the sources of errors including interference from the mother tongue. The number of errors that involved word order and/or auxiliaries was found to be 788, of which no substantial examples are definite instances of L1 interference.

The following table shows the type and number of errors made in the 4000 *Wh*-questions:

Table II

Type of error	Number	Per cent
1. Word order	437	24.4
2. Auxiliary omission	154	8.6
3. *BE* redundancy	18	1.0
4. *BE* replacing *DO*	27	1.5
5. *DO* redundancy	73	4.0
6. *DO* replacing *BE/Aux*	79	4.4
7. Verbal form	161	9.0
8. Tense	69	3.8
9. Question words	488	27.0
10. Inappropriate questions	163	9.0
11. Miscellaneous	121	6.7
Total	1790	100%

2.1.3. *Third Study*

This study (Mukattash 1978a), deals with the identification and description of 1411 grammatical errors contained in fifty essays chosen randomly from 200 essays written by first year students at the University of Jordan in October 1977.

The overall number of sentence used in the fifty essays was found to be 1618, 639 of which are simple sentences and 979 complex ones. The percentage or erroneous sentences was 39.8%.

The texts were corrected not only by the author but also by a native speaker of English. Table III below shows the type of errors, their frequency of occurrence as well as the percentage of each type of error to the overall number of errors in the fifty essays. It can be seen that the bulk of mistakes (81.9%) fall into four major categories: Verbals, Articles, Nominals and Prepositions. This is, however, not surprising since these items are possibly the ones most used in the language.

The mistakes under each of the categories in Table III below were further

sub-categorized and described. Errors tat were thought to be cases of Li interference were found to be 331, that is to say 29% of the overall number of errors.

Table III

Categories	No. of errors	Per cent
1. Verbals	409	29.0
2. Articles	306	21.7
3. Nominals	228	16.2
4. Prepositions	211	15.0
5. Relatives	51	3.6
6. Adjectivals	38	2.7
7. Pronouns	38	2.7
8. Sentence Connectors	37	2.6
9. Quantifiers	29	2.1
10. Adverbs	29	2.1
11. Structure	18	1.2
12. Modals	17	1.2
Total	1141	100%

Admittedly, the percentages of errors in Table III above may be misleading: it is more revealing to look at errors as a percentage of usage, and this was done in certain cases. For instance, the percentage of errors involving the use of Modals to the overall number of errors (i.e. 1.2%) might be interpreted as meaning that Modals do not cause difficulty to Jordanian students. Actually, it is just the opposite. A usage count of the number of times the Modals were/ /should have been used was carried out, and this number was found to be 178; which means that the percentage of errors to usage is 9.6%, an indication that Modals are difficult.

2.1.4. *Fourth Study: a semi-objective test*

This test was given to over one thousand first year students at the University of Jordan. Five hundred answer sheets representing students from all faculties and all districts in Jordan were chosen for analysis.

The test included 170 items and covered eight areas of grammar. The answers were marked, coded and subjected to computer analysis. The following table shows the number and percentage of the errors made in each area. Furthermore, it shows the "absolute difficulty" of the errors under each category (the term "absolute difficulty" means the percentage of errors involving the use of a certain grammatical structure/form to the number of times that particular grammatical structure/form was used or should have been used):

Table IV

Category	1 No. of items	2 No. of errors	3 Percent-tage	4 Absolute difficulty
1. Verbals (including gerunds & infinitives)	50	17461	37.6	69.8%
2. Prepositions	40	10667	23.0	53.3%
3. Articles (including *some* & *any*)	43	9755	21.0	45.3%
4. Modals	10	2703	5.8	54%
5. Adverbs	9	2256	4.9	50.1%
6. Connectors	9	1931	4.2	42.9%
7. Quantifiers	5	1156	2.5	46.2%
8. Relative Pronouns	4	541	1.1	27%
Total	170	46470	100%	—

As in the case of the preciding study, the errors under each category were further sub-classified and all types of substitutions were discussed.

It ought to be pointed out in this connection that this test was written before the author had any clear results from the free compositions concerning the type and frequency of errors. The items tested were thought-on purely impressionistic grounds — to be the most problematic areas in English grammar for Jordanian students.

2.2. PHASE II

2.2.1. *Aims and Methodology*

The aims of the second phase of the research may be summed up as follows:

(i) The validation of the survey instruments and the finidings of the Pilot Project, which suggest that grammatical mistakes of Jordanian students are most evident in the following areas, in descending order of difficulty:

1. The verbal system, especially with reference to tense and aspect
2. Prepositions
3. Modals
4. Adverbs
5. Quantifiers
6. Articles
7. Sentence Connectors

(ii) Establishing the existence of "competence" errors in Jordanian English as opposed to "performance" mistakes caused by situational constraints of test administration and other extraneous factors.

(iii) Surveying the gamut of the school-leaving population — as opposed to the very small percentage which enters the University — to indicate the extent to which the latter group is representative of this age-group, and to discover and describe any variation in English language proficiency according to sex, region, type of school, type of curriculum, educational and social backgrounds, etc.

(iv) The explanation of the sources and causes of prevalent grammatical errors.

(v) The remediation of the prevalent grammatical errors by the construction of programmes of study the data for which will be provided by the findings at (i), (ii), and (iv) above. (The programme of study constitutes the final phase of the research and will be started and tested next academic year).

This research was begun in November 1977. The first step was the construction of a new test informed by the results of the Pilot Project.

The first version of the test, an objective one, consisted of 360 multiple-choice items which were tried out for purposes of pretesting and item analysis on 200 students representing all faculties in the University. All questions were based on the English Language syllabus used in government secondary schools. All items were analyzed to determine their effectiveness in terms of three criteria: item difficulty, item discrimination, and the effectiveness of distracters. Furthermore, items that included nonfunctioning and malfunctioning distracters were all excluded from the final version.

The final version of the test consisted of 150 multiple-choice items covering eleven different areas of English grammar (see Table V below). The validity and the reliability of the test proved to be very high indeed.

The administration of the test was first begun in January 1978 and it lasted for three months. The test was given to 4835 students representing three main groups of students' population: university, teachers' training institutes, and secondary schools (final year students) all over the country. The number of students is distributed as follows: 940 from the University of Jordan, 513 from five teachers' training institutes, and 3382 from 48 secondary schools in 13 geographical areas.

2.2.2. A Preliminary Analysis

After each administration of the test all answer sheets were checked and coded. The answers were then marked by the computer at J. U.. For each question we were able to get the number of examinees who answered the

question correctly as well as the number of examines who used each of the other three distracters, and the number of examinees who did not answer the question at all. Afterwards we were able to calculate the facility index (I.F.) for each item and the facility index for all items in each of the eleven grammatical areas that comprised the test.

The following table shows us the facility index for each of the eleven grammatical areas arranged in descending order of difficulty:

Table V

Area	No. of items	Facility
1. Quantifiers	12	0.37
2. Adverbs	12	0.38
3. Verbals	35	0.40
4. Sentence Connectors	11	0.44
5. Nominals	10	0.45
6. Prepositions	24	0.46
7. Modals	9	0.48
8. Relatives	8	0.48
9. Adjectivals	8	0.50
10. Articles	15	0.52
11. Pronouns	6	0.56

As in the case of previous studies all prevalent mistakes were decribed, and an attempt was made to account for their occurrences. Furthermore, an attempt was made to establish the relationship between the student's competence in English and some social and educational factors: age, type of school, district of schooling, sex, educational background, etc. (see Mukattash, forthcoming).

3. PROBLEMS OF STATEMENT AND METHODOLOGY

In what follows I shall briefly discuss some of the main problems which I have encountered over the last four years while engaged in the analysis of prevalent grammatical mistakes in Jordanian English.

3.1. THE CLASSIFICATION OF ERRORS

One of the most serious problems that face a researcher working in the field of error analysis is the description and classification of errors. Admittedly, some errors lend themselves easily to description and classification (e.g. the replacement of one preposition by another, the inappropriate use of the definite

article *the* in front of proper names), but others do not. To explain the point under discussion, let us begin by considering the mistakes in the following three examples produced by first year students at the University of Jordan:

(1) This lay in the *southern*
(2) If we want to make our country attractive to the *foreigns* we must pay a great attention to the subject.
(3) In *agriculture* field, we see the difference between agriculture in these days and in the last days.

What is involved here is the inappropriate use of a noun instead of an adjective and vice versa. But is this a syntactic error or a morphological one, or is it a lexical one? Another question that one would like to pose in this respect is: What are the causes of the errors in these deviant sentences? Contrastive analysis (Arabic and English in this case) may account for the mistake in (3) but not the mistakes in (1) or (2).

By way of further exemplification, let us consider the mistake in the following example:

(4) Most *of people* work in trading

The mistake here may be looked at from two different angles: first, inappropriate deletion of the definite article *the* (viz. *Most of the people work in trading*); secondly, redundant use of the preposition *of* (viz. *Most people work in trading*). Out of context, and without reference to L1, it would be difficult to prefer one alternative to the other. As it happens, a comparison between L1 and L2 would make us favour the first interpretation, for this reflects the facts of L1. But resource to L1 is not always helpful. How can we describe the mistake in:

(5) *The another* cause for this custom is their culture.

Again, the mistake here may be interpreted in two different ways: (i) the use of *another* instead of *other*, and (ii) the redundant use of the definite article *the*. Resource to L1 is not helpful in this particular case, for both interpretations are possible in Arabic.

A somewhat different problem can be exemplified by the following deviant sentences:

(6) The television *has just repaired* and it is working very well
(7) When the manager arrived, he found that the safe *had opened* by the thieves.

At first sight, one is tempted to view the mistakes in the preceding two examples as ones in "Voice" (i.e. the use of the active voice instead of the passive). But this is not necessarily the case. In fact, it is not clear whether we should consider these mistakes as mistakes in "voice" or in "form". Most probably

they are mistakes in "verbal form". This observation is based on two facts. First, the two contexts in (6) and (7) above require the use of the passive form of the verb in Arabic. Secondly, there are mistakes in the opposite direction, where the Arabic context requiers the use of the active voice: cf.

(8) They *are sung* because they have passed their exams
(9) Education in Jordan *has been reached* a high level because you will find schools in every village

The only way to decide whether these are mistakes in "voice" or "form" would be to ask the individuals who produced such sentences to translate them into L1.

3.2. SOURCES OF ERRORS

If the description of errors proves to be difficult, the justification of their occurrence is certainly more difficult. Again, one must admit that certain errors may be easily accounted for in terms of L1 interference or in terms of intralingual interference, but certainly there are errors which are difficult to account for, and it would not be very helpful to refer to such errors as "ambiguous goofs". This statement is based on the following observations:

(i) Many of the grammatical errors which Jordanian students make have been recorded to have been made not only by children acquiring English as their mother tongue, but also by foreign learners of English with different language backgrounds. One of the most prevalent mistakes in Jordanian English involves word order in *Wh*-questions. The following are illustrative examples (Mukattash, 1978b):

(1) *What the group is doing?
(2) *How the students are studying?
(3) *When our friends will arrive?
(4) *Where Mrs. Smith was cooking?

What is involved here is the absence of the obligatory inversion of the auxiliary and the subject NP. At first sight, one might fall under the impression that the errorneous questions above are direct translation from Arabic. Indeed, the word order of the deviant questions above reflects the structure of both Modern Standard Arabic and spoken Jordanian Arabic. In fact I did discuss these mistakes with some colleagues who have been teaching English at the university level for a long time, and they unanimously believed that the deviant questions listed above are cases of L1 interference. But such deviant questions which include the absence of inversion of the auxiliary and the subject NP have been noticed to occur in the speech of children whose native language is English. Menyuk (1969 : 75—6) cites the following examples:

(5) a. *What you are writing?
   b. *Why they are here?
   c. *What they are doing?

Dulay and Burt (1974 : 111) also cite similar examples:

(6) a. *When John will read the book?

What adds to the complexity of the situation is that deviant questions like the ones produced by Jordanian students (i.e. 1—4) are reported to have been made by foreign learners with different language backgrounds. Ravem (1974 : 141) records the following *wh*-questions produced by his children whose L1 is Norwegian:

(6) b. *What he's doing?
   c. *How they can't talk?

Richards (1974 : 187), on the other hand, records similar deviant questions produced by native speakers of various languages (Japanese, Chinese, Burmese, French, Czech, Polish, etc.): cf.

(7) a. *What she is doing?
   b. *When she will be fifteen?
   c. *Why this man is cold?

(ii) I examined many of the deviant sentences recorded in Jain (1974 : 208 —13) which were produced by Indian students whose L1 is Hindi, and discussed these sentences in the classroom. In most cases my students made the same mistakes, and when I discussed these mistakes with them, they thought that many of these errors were caused by interference from Arabic. The following is a representative sample:

(8) a. The girls need three more *scissors*
   b. I *was seeing* all this happen in front of my own house.
   c. Entry will be by *tickets*
   d. We *enjoyed* at the theatre for two hours

(iii) Many of the mistakes which Richards (1974 : 182—88) reports to have been produced by speakers of different languages are also prevalent in Jordanian English. Here are some typical errors in the verbal system:

(9) a. Be+V // V
      *We *are live* in this hut
   b. M+V — en // M+V
      *I *can saw* it
   c. V — en // Be+V-en
      *He *born* in England

   d. V-ing // Be+V-ing
      *They *running* very fast
   e. V-ing // V-en
      *I *am interesting* in that
   f. Be+V-ing // V
      *I *am not liking* it

Other mistakes involving the use of prepositions, articles and auxiliaries, which Richards reports are also typical of Jordanian learners of English: e.g. *reach at a place; at the evening; angry on him; play on the piano; entered in the room; the friendship; after the breakfast; the science; did he went*, etc. Indeed, I would have considered many of the mistakes which Jordanian learners of English make to be cases of L1 interference were it not for the fact that many of these errors have been recorded to have been made by other foreigners with different language backgrounds and by children acquiring English as their mother tongue.

3.3. DEGREE OF ACCEPTABILITY

A third problem that faces a researcher working in the field of error analysis is determining the accepability of a given utterance. This is particularly true if the researcher is not a native speaker of the target language. For example, some of the erroneous sentences listed in Jain (1974 : 210) and Richards (1974 : 185) have been judged by some native speakers to be perfectly acceptable. Here are some examples:

(1) I *am having* a very heavy work-load this semester
(2) I *was feeling* as if I was in a hell
(3) She has *a* very important *work* to do
(4) He *said* that there *is* a boy in the garden
(5) I met *with* her

Secondly, I myself have had a difficult experience with two research assistants who are native speakers of English (one is a native speaker of British English and the other a native speaker of American English). The two are linguistically trained, but quite often they disagreed with respect to the acceptability of a good number of sentences. Each of the following examples was attested by one and rejected by the other:

(6) They cannot perceive *of* the difference
(7) Mr. Smith left *to* England yesterday
(8) I *didn't see* him since Christmas
(9) The professor spoke *with* a high voice

(10) We decided to go to the beach *by* foot

(11) I prefer this book *than* that one

Thirdly, there are problems in the opposite direction, namely determining the acceptability of sentences in the mother tongue (Arabic in this case). In fact, this proved more difficult than determining the acceptability of English sentences. This is due to two factors. First, Modern Standard Arabic is rarely used in everyday life and is restricted to very formal situations; and though most educated Jordanians have a passive knowledge of the language, the overwhelming majority of them cannot speak the language with any reasonable degree of fluency. Secondly, there are many dialectal variations in Jordan, which display differences on all linguistic levels. Thus it would not be possible for a researcher to decide on the acceptability of a given utterance in Arabic; and consequently it would not be possible for him to determine whether a given mistake in Jordanian English is a function of L1 interference or not.

3.4. PROBLEMS OF METHODOLOGY

It is quite obvious to any researcher working in the field of error analysis that no one testing technique or elicitation procedure is adequate for the purposes of the specification, description and justification of prevalent grammatical errors in the second/foreign language of any speech community.

The shortcomings of any testing technique will become obvious only after one begins to analyse his data. Generalizations based on the use of one test-type are bound to be inaccurate and misleading. Briefly, various types of tests are to be used in order to neutralise the extraneous factors which are involved when any one is used exclusively. Furthermore, for any generalization to be reasonably valid, various testing techniques are to be employed systematically within the same grammatical area.

Secondly, there are other questions which one would like to pose with respect to the informants who provide the linguistic data: (i) How many informants does one need in order to be able to make reasonable generalizations about the "competence" errors in the foreign/second language of a particular speech community?; (ii) What type of informants should a researcher choose (e.g. students, friends, people who use the foreign language regularly!)?; (iii) How can one guarantee that his informants will co-operate, taking into consideration the time and effort needed?; (iv) Should the various testing techniques be used over a short or an extended period of time? As far as I know, none of these questions has been adequately answered so far.

Thirdly, in any research in error analysis, certain socio-linguistic and psycho-linguistic factors should be taken into consideration. Such factors

include (i) the age at which the student begins to learn L2; (ii) the learner's linguistic and cultural background; (iii) the status of the mother tongue; (iv) the learner's attitude towards L2 and its speakers; (v) the relevance of L2 to the learner's needs and objectives; (vi) the prestige of L2 locally; (vii) the learner's first experience with L2; (viii) the textbooks used and the objectives of the syllabus; (ix) methods of teaching; (x) modality of exposure to L2; (xi) individual differences in language aptitude, intelligence, interest and perseverance. Other factors that are relevant in this respect are: (1) the type and calibre of the teachers of L2; (2) the attitude of L2 teachers to the teaching profession in general; (3) the nature of the educational system in the country (see Mukattash, forthcoming, and the references he cites).

To conclude this paper, there seems to be no way of proving that any given error-type has a specific cause, whether this be L1 interference, learners' communication strategy, or target-language interference. The most that can be claimed is that any given error-type can be plausibly attributed to one, or to a combination, of these causes. Progress in the resolution of such indeterminancy can only be assayed as James (1978) suggests "through small-scale in-depth psycho-linguistic research".

REFERENCES

Burt, M. K. and Kiparsky, C. 1974. "Global and local mistakes". In Schumann, J. and Stenson, N. (eds). 1974. 71—80.

Dulay, H. C. and Burt, M. K. 1974. "You can't learn without goofing". In Richards, J. (ed.). 1974. 95—123.

Jain, M. P. 1974. "Error analysis: source, cause and significance". In Richards, J. (ed.). 1974. 189—215.

James, C. 1978. A report on the Error-Analysis Research Project in Jordan, the University of Jordan.

Menyuk, P. 1969. *Sentences children use*. Cambridge, Mass.: the M.I.T. Press.

Mukattash, L. 1977. *Problematic areas in English syntax for Jordanian students*. Unpublished MS, the University of Jordan.

Mukattash, L. 1978a. "Common grammatical errors in Jordanian English". *The Interlanguage Studies Bulletin 3*. 250—291.

Mukattash, L. 1978b. "Wh-question in English: a problem for Arab students". Paper read at the 5th International Congress of Applied Linguistics under the title "Wh-questions in English: a problem for foreign learners", Montreal, 1978. To appear in *IRAL*.

Mukattash, L. 1979. *Further studies in Jordanian English*. Unpublished MS, the University of Jordan.

Mukattash, L. 1980. "Yes/No questions and the contrastive analysis hypothesis". *E.L.T.J*, XXXIV, 2. 133—145.

Mukattash, L. Forthcoming. "English language proficiency in Jordan: educational, socio-economic and other factors", to appear in *Dirasat* (The University of Jordan Journal).

Nickel, G. (ed.). 1971. *Papers in contrastive linguistics*. Cambridge: Cambridge University Press.

Ravem, R. 1974. "Language acquisition in a second language environment". In Richards, J. (ed.). 1974. 124—133.

Richards, J. (ed.). 1974. *Error analysis*. London: Longman.

Richards, J. 1974. "A non-contrastive approach to error analysis". In Richards, J. (ed.). 1974. 172—188.

Schumann, J. and Stenson, N. (eds.). 1974. *New frontiers in second language learning*. Rowley, Mass.: Newbury House.

Svartvik, J. (ed.). 1973. *Errata: papers in error analysis*. Lund: CWK Gleerup.