AN ANALYSIS OF TRANSLATION ERRORS MADE BY BRAZILIAN STUDENTS

Uma Análise de Erros de Tradução Feitos por Alunos Brasileiros

Fehleranalyse bei Übersetzungen Brasilianischer Studenten

Maria da Conceição Magalhães VAZ DE MELLO *

SUMMARY

This paper presents an analysis of the errors which occurred in translations from Portuguese into English written by Brazilian students. This topic has been chosen because there are few studies of errors made by Brazilian students in the process of learning English and also because many language teachers still consider interference from the mother tongue the only source of errors in foreign language learning . Errors due to interference from the foreign language itself have often been disregarded.

In order to explain the causes of the errors five categories were established: errors due to L1 interference, errors due to L2 interference, errors due to L1 and/or L2 interference, errors due to communication strategies and errors of indeterminate origin. Errors due to communication strategies were classified according to three different types, proposed by Tarone (1977). The first is topic avoidance, the second is paraphrase and the third is conscious transfer. Errors of indeterminate origin are unsystematic and cannot be assigned to any of the other four categories.

My claim in this paper is that since the speech of children learning their first language contains many errors, foreign language learners should be allowed to make errors. By making errors learners provide the teacher with information about what they haven't learned.

RESUMO

Este artigo apresenta uma análise dos erros que ocorreram em versões do português para o inglês feitas por alunos brasileiros. Este tópico foi escolhido porque há poucos estudos sobre erros feitos por alunos brasileiros durante o processo de aprendizagem do inglês e também porque muitos professores de línguas ainda consideram a interferência da língua materna como a única fonte de erros na aprendizagem de uma língua estrangeira. Erros de interferência da própria língua estrangeira freqüentemente não são considerados.

Para explicar as causas dos erros cinco categorias foram estabelecidas: erros de interferência da L1, erros de interferência da L2, erros de interferência da L1 e/ou da L2, erros de estratégias de comunicação e erros de origem indeterminada. Os erros de estratégia de comunicação foram classificados de acordo com três tipos diferentes sugeridos por Tarone (1977). O primeiro é abstenção de tópico, o segundo, paráfrase e o terceiro, transferência consciente. Erros de origem indeterminada não são sistemáticos e não podem ser classificados de acordo com nenhuma das outras quatro categorias.

Um dos objetivos deste trabalho é provar que, do mesmo modo que a fala de crianças aprendendo sua língua materna apresenta muitos erros, alunos aprendendo uma língua estrangeira também deveriam poder fazer erros. É através dos erros de seus alunos que o professor consegue informação sobre aquilo que eles ainda não aprenderam.

* Faculdade de Letras
  UFMG

INTRODUCTION

This paper presents an analysis of the errors which occurred in translations from Portuguese into English written by Brazilian students. This topic has been chosen because there are few studies of errors made by Brazilian students in the process of learning English and also because many language teachers still consider interference from the mother tongue the only source of errors in foreign language learning. Intralingual errors have often been disregarded.

Besides classifying the errors found in the corpus and computing their relative frequency, this study has also established the sources of each error type.

This last task is very important because as Bickerton (1974) pointed out

"We should beware of overrigid methods of error analysis, and remember that the purpose of such analysis should not be merely the classification of different types of error, nor even the establishment of their relative frequencies; these are sterile exercises unless at the same time we are trying to find out why errors are made."

For the establishment of the sources of the errors I have followed the interlanguage hypothesis which, according to Adjemian (1976),

"grew out of the observation that adult learners of a second language, in an attempt to convey meaning, produce speech which differs from the speech of native speakers of the target language (TL) in ways which are not always the result of transfer from the learner's native language (NL)".

The term 'interlanguage' was first used by Selinker (1977) who defined it as "a separate linguistic system based on the observable output which results from a learner's attempted production of a TL norm." The terms 'approximative system' and 'idiosyncratic dialect' have been used by Nemser and Corder to refer to the same concept.

The learners whose works I have analysed in this study were all speakers of the same mother tongue and shared the same experience in learning English as a foreign language. We are thus led to suppose that they have produced the same interlanguage and that the differences which can be found in their written work can be ascribed to individual variation in intelligence, motivation and attitude.

As Corder has pointed out the learner's errors are important because of three main reasons. First, they tell the teacher what the learner has already learned. Second, they tell the researcher what processes the learner employs when he attempts to learn a foreign language. Third, they cannot be avoided because they constitute a necessary stage in the process of language learning, both in the case of the native and of the foreign language.

CRITERIA FOR THE DESCRIPTION OF ERRORS

One of the advantages of using translations from the native into the target language for the purpose of Error Analysis is that the interpretation of the original text by the learner does not present too many problems. The difference between translation and composition, as far as Error Analysis is concerned, lies in the fact that in composition students try to use the most familiar constructions, whereas in translation they are forced to employ the constructions of the original text.

The corpus used in this study consists of fifteen translations from Portuguese into English written by Brazilian students as part of the "Lower Cambridge Examination". The students were all males and the average age was sixteen years. They had already studied English for about five years in British and American Binational Centres or in other language schools.

The translations were first corrected by six native English speakers. Since the correctors didn't know Portuguese, they have not detected errors related to lexical items which were not correct renderings of those in the original text. It was then necessary to elaborate a standard translation based on the original text. It was submitted to two native speakers of English with a good knowledge of Portuguese and to two native speakers of Portuguese with a good knowledge of English.

Johansson (1975) pointed out that the comparison between the original text and the standard translation is important if one wants to calculate the frequency and types of errors. In his view errors can be "avoided by an inexact translation or a translation which is correct from the point of view of the TL, but not a correct rendering of the original text".

According to Corder (1975) the difficulty in detecting errors "lies in the fact that what looks or sounds like a perfectly acceptable sentence may nevertheless contain errors." He defines a covertly erroneous sentence as the one which a native speaker would accept as being superficially well formed but which is not appropriate in the context in which it occurs whereas an overtly erroneous sentence is that which cannot be accepted by native speakers. In the corpus used in this study, most of the sentences containing errors corrected by the native speakers are overtly erroneous whereas most of the sentences containing errors corrected exclusively by me are covertly erroneous.

Since the translations were corrected by more than one native speaker, there were cases of items and structures which were considered as errors by some correctors, but not by others. They were not marked as errors because we cannot require a higher degree of proficiency from foreign learners than from native speakers. An example is the form *acampados* which occurs in the structure *viu um grupo de ciganos acampados perto da estação.* This item was translated in two different ways, i.e., 'camped' and 'camping'. Since both forms were

accepted by the native correctors, they were also accepted by me.

One of the most difficult problems in the field of error analysis is the classification of errors. While some errors would be classified in the same way by different people, others may be interpreted in two or more different ways.

The taxonomy of error types used in this study includes the four categories suggested by Corder (1975), i.e., omission of some required element, addition of a superfluous element, selection of an incorrect element and misordering of elements, as well as misspelling. Furthermore the items involving errors were classified in terms of the different parts of speech, that is, noun, verb, adjective, adverb, conjunction, preposition, determiner, and pronoun as well as in terms of number, case, aspect, and tense.

The advantage of using these different systems of description lies in the fact that the frequency of occurrence of the errors can be calculated either in terms of one system or in terms of a cross-association of two or more systems.

THE CAUSES OF ERRORS

In order to explain the causes of the errors detected in the corpus five categories were established:

> errors due to L1 interference
> errors due to L2 interference
> errors due to L1 and/or L2 interference
> errors due to communication strategies
> errors of indeterminate origin.

Errors due to L1 interference are those the learner makes when he relies on native language structures. An example of this type is the structure 'she swam until die' where the learner used the infinitive in English because the corresponding Portuguese structure *nadar até morrer* contains an infinitive verb form.

Errors due to L2 interference are those the learner makes when he overgeneralizes a target language rule to cases where the rule does not apply. The use of the verb 'swim' as a regular verb in the sentence 'Moema swimmed to follow it' is an example of this type of error.

Errors due to interference of L1 and/or L2 are those the learner makes when he relies on both native and target language structures. The use of the form 'Portugueses' as the plural of the noun 'Portuguese' exemplifies this case.

Errors due to communication strategies are classified according to the three different types proposed by Tarone (1977).

The first is 'topic avoidance' which, according to the same author, "occurs when the learner simply does not talk about concepts for which the vocabulary is not known." The second is 'paraphrase', which is defined as "the rewording of the message in an alternate, acceptable target language construction, in situations where the appropriate form

or construction is not known or not yet stable." An example of paraphrase is the use of the structure 'one day after that' to translate the Portuguese structure *no dia seguinte*. The third is 'conscious transfer' and refers to cases where "the learner simply uses the NL term without bothering to translate". An example is the use of the Portuguese noun *ciganos* in the structure ' group of ciganos'.

Errors of indeterminate origin are those which are unsystematic and cannot be assigned to any of the other four categories. Corder (1975) pointed out that "there is ... no way of describing or accounting for something which is not systematic." An example is the use of the simple present in the structure 'he decide to tell her' to translate the Portuguese structure *resolveu contar a ela*.

ERROR FREQUENCY AND ERROR CAUSES

The last step in this study was to compute the frequency of the error types which occurred in the corpus and to establish their correlation to the frequency of the error causes.

In order to compute the frequency of error types and to make a statistical analysis of this frequency, I have used the Statistical Package for the Social Sciences (SPSS) which is an integrated system of computer programs for the analysis of social science data.

The computation of the errors contained in the corpus showed that the fifteen translations contained a total of 644 errors. Table 1 presents the number of errors in each sentence per student. The sentence containing the largest number of errors is sentence 5, in which a total of 234 errors, corresponding to 36,2% of all the errors, occurred. This was probably so because this is by far the longest sentence in the translation. In order to neutralize the influence of the sentence length, in each sentence the percentage of errors was divided by the number of words it contains. Sentences 6 and 7 were thus proved to be the most difficult ones, since they present the highest percentage of errors. These results are also presented in Table 2.

As it has already been stated, the taxonomy used for the linguistic description of the errors made it possible to classify them in syntactical as well as in morphological terms. Table 2 presents a classification matrix where the errors are analysed according to the categories mentioned previously in this paper, which are listed in the vertical line, while the parts of speech are listed in the horizontal line. The results indicate that the most frequent error type is that involving substitution of an item for another within the same word class.

Table 3 presents the number and percentage of errors related either to the different word classes or to the clause. The first column contains the number of times each word class occurred or should have occurred in the corpus. The second contains the number of errors in each part of speech. The third contains the percentage of errors in each part of speech in relation to the total number of errors. The fourth contains the percentage of

errors in each part of speech in relation to the number of times it occurred or should have occurred in the corpus. The last column presents this percentage multiplied by a constant factor to make its total equal to 100. An analysis of the frequency of error types enabled the researcher to make the following generalizations: first, errors involving adverbs, verbs, conjunctions' and prepositions are those which presented the four highest degrees of difficulty (more than 60%), varying from 13.5% to 17.4%; second, substitution of an item for another within the same word class was responsible for almost 50% of all errors (within this category, nouns and verbs were the word classes which presented the greatest incidence of errors, i.e. 49%); third, spelling was responsible for only a very small part of the total number of errors, (2%).

Those observations reveal that the degrees of difficulty related to adverbs, verbs, conjunctions and prepositions were about the same. The high frequency of errors involving adverbs and conjunctions can be attributed to the specific structure of the original text. Therefore verbs and prepositions can be considered the two most difficult parts of the English grammatical system for this group of learners. It is worth noting that this conclusion is confirmed by most of the studies dealing with error analysis in foreign language learning, which have shown that verbs and prepositions are the most difficult areas for students learning English as a foreign language.

THE FREQUENCY OF ERROR TYPES

Most of the Error Analysis studies which present not only the frequency of error types but also the causes of errors have shown that interference from the mother tongue was one of the greatest sources of errors in foreign language learning.

The results obtained by authors such as Aguas (1964), Richards (1977) and Duskowa' (1969) are very similar to the ones found in the present study. Table 4 presents the distribution of the different error types according to the five categories established to explain the causes of the errors. The results indicate that more than 40% of the errors were accounted for in terms of interference from the learners' mother tongue; 14%, in terms of the target language; 16%, in terms of both the native and the target language. The remaining errors were classified as errors due to communication strategies and errors of indeterminate origin, (13% each).

It is interesting to note that one of the main sources of the errors classified as substitution of an item for another within the same word class was under-differentiation, i.e., an item in the native language corresponds to two or more in the target language. Arabsky (1968) considered this kind of error as caused by 'external active interference'. One of the examples he gives is the use of the noun 'earth' in a context where 'land' should be used.

This same error occurred three times in the corpus used in this study.

Some of the errors due to interference from the target language may be classified as 'transfer avoidance'. This is a strategy the learner uses to avoid being influenced by his mother tongue. An example is the structure *homem do fogo* which should be translated as 'man of the fire' (a word for word translation). What is interesting about this structure is that among the fifteen students only one produced the correct structure. It is worth noting that errors due to 'transfer avoidance' would never be predicted by means of Contrastive Analysis.

CONCLUSION

One should now discuss to what extent the results presented in this study can be generalized in relation to the frequency of error types and the degree of difficulty of the structures involved. One of the advantages of using translations in an error analysis study lies in the fact that every learner belonging to a particular group is forced to produce the same types of constructions.

It has been pointed out that the frequency of errors should be related to the number of possible mistakes. This will make it possible for the researcher to connect the degree of difficulty of a construction to error frequency. One of the advantages of using translations is that this can be done taking the standard translation as a basis. The results of the application of this mechanism to our corpus are shown in Table 3.

One can state that error frequency should be used as a criterion to measure the faulty performance of a group of learners in specific areas. The results thus obtained may help the teacher to determine the selection as well as the emphasis that should be given to particular items in teaching and testing.

The important thing concerning the attempt to explain the causes of errors in adult foreign language learning is that the researcher can infer the strategies adopted by the learners in their attempt to learn a foreign language.

After having established the causes of the errors, I have observed that the learners have used different kinds of strategies which seem to be common to adult learners in general whatever their mother tongues are.

One should think that errors attributed to interference from the native language would be specific to groups of learners sharing the same mother tongue. However many errors of this kind are common to groups of learners with different mother tongues. This is so because very often the native languages share some of the features related to the errors. An example is the use of the noun 'earth' instead of 'land' by Brazilian as well as by Polish students, due to the fact that in both languages there is only one noun which corresponds to the two English terms.

It has already been pointed out that Contrastive Analysis could be used to explain the errors due to interference from the mother tongue. It has been proved that transfer of native language structures does not occur whenever there is a contrast between the native and the target language structures. Therefore Ioup and Kruse (1977)

"require a new hypothesis defining precisely where transfer is expected and an explanation as to why it is confined to these aspects of language acquisition."

On the other hand errors due to interference from the target language cannot be explained by means of Contrastive Analysis. They can be found in the performance data of learners with different native languages. They are similar to those made by children in the process of acquiring their mother tongue and should be considered as an inevitable stage in foreign language learning.

We can conclude that since the speech of children learning their first language contains many errors, foreign language learners should be allowed to make errors. Errors tell the teacher how the learner progressed and what is left to be learned.

Clark (1975) suggested that "only by allowing errors to occur can we discover any discrepancies that might exist between what the student actually learned and what we intended him to learn."

TABLE 1

Number of errors in each sentence per student

Student	Sentence 1	2	3	4	5	6	7	8	9	10	11	12	13	Total	%
01	2	0	0	1	12	9	0	2	4	0	0	5	3	38	6,0
02	2	1	1	1	22	13	0	3	7	0	0	1	6	57	9,0
03	1	0	0	1	20	7	0	3	3	0	0	2	4	41	6,5
04	0	0	0	1	8	9	3	2	2	0	0	0	4	29	4,5
05	0	0	2	1	15	12	1	4	5	0	0	2	4	46	7,0
06	1	0	0	1	14	12	2	3	4	0	0	1	7	45	7,0
07	2	0	0	1	16	6	2	2	10	0	0	1	6	46	7,0
08	1	0	0	3	15	7	0	2	3	0	0	2	3	36	5,5
09	1	0	2	1	19	10	3	2	6	0	2	5	4	55	8,5
10	0	0	0	1	14	5	0	4	4	0	0	2	5	35	5,5
11	1	0	0	1	25	13	2	3	4	0	0	2	7	58	9,0
12	2	0	0	1	16	9	0	5	5	0	0	3	4	45	7,0
13	0	0	1	0	14	10	3	3	3	0	0	1	6	41	6,5
14	2	0	1	1	12	8	1	2	2	0	0	1	3	33	5,0
15	1	0	0	1	12	12	1	2	4	0	0	2	4	39	6,0
Total	16	1	7	16	234	142	18	42	66	0	2	30	70	644	100
Number of words	11	3	8	5	64	38	4	15	26	0	1	22	28	226	

TABLE 2

Number and percentage of errors in each category

CATEGORY	NOUN	ADJ.	VERBS	ADV.	DET.	PRON.	PREP.	CONJ.	CLAUSE	TOTAL	%
Word order	5	1		8						14	2,0
Tense/aspect			74							74	11,5
Addition	3	2	7	7	7	9	16	24		75	11,5
Omission	10	1	2	4	12	17	2	32	2	82	13,0
Different word classes	9	2	6	7		8	2	3	4	41	6,5
Inflection	13	8	5							26	4,0
Case	16									16	2,5
Substitution	55	4	96	12	22	23	39	20	32	303	47,0
Spelling	1	1	10					1		13	2,0
Number of errors	112	19	200	38	41	57	59	80	38	644	100,0

TABLE 3

Number and percentage of errors in each word class

WORD CLASS	NUMBER OF TIMES IT OCCURRED IN THE CORPUS	NUMBER OF ERRORS	PERCENTAGE	DEGREE OF DIFFICULTY (*)	NORMALIZED DEGREE OF DIFFICULTY (**)
Adverb	135	38	5,90%	28,15%	17,48%
Verb	780	200	31,06%	25,64%	15,92%
Conjunction	315	80	12,42%	25,40%	15,77%
Preposition	270	59	9,16%	21,85%	13,56%
Pronoun	360	57	8,85%	15,83%	9,83%
Noun	780	112	17,39%	14,36%	8,91%
Adjective	135	19	2,95%	14,07%	8,73%
Determiner	450	41	6,37%	9,11%	5,66%
Clause	570	38	5,90%	6,67%	4,14%
Total	3.795	644	100%	—	100%

(*) Degree of difficulty means the percentage of errors involving the use of certain grammatical structure or lexical item to the number of times that particular grammatical structure or lexical item was used or should have been used.

(**) Normalized degree of difficulty means the degree of difficulty multiplied by a factor that makes its total equal to 100%.

TABLE 4

Error types and error causes

	L1 INTERFERENCE	L2 INTERFERENCE	L1 AND/OR L2	COM. STRATEGIES	IND. ORIGIN
Change of word order	8	6			
Change of verb tense and aspect	11	8	45		10
Addition of superfluous elements	28	10	10	1	26
Omission of essential elements ..	41	11	5	15	10
Different word classes	8	15	7	3	6
Inflection	5	5	15		1
Change of case	2	13			
Substitution of noun	11	3	8	32	1
Substitution of adjective	3			1	
Substitution of adverb	11			1	
Substitution of conjunction	19			1	
Substitution of preposition	34	4			1
Substitution of determiner	8	1	1		12
Substitution of pronoun	19				4
Substitution of verb	51	8	10	20	7
Substitution of clause	18		3	9	2
Misspelling	5	5			3
Total	282	89	104	83	83
%	44,0	14,0	16,0	13,0	13,0

BIBLIOGRAPHY

ADJEMIAN, C. On the nature of interlanguage systems. *Language learning*. Michigan, 26(2) : 297-320, 1976.

AGUAS, E. F. English composition errors of Tagalog speakers. UCLA, 1964.

ARABSKY, J. A linguistic analysis of English composition errors made by Polish students. *Studia Anglica Posnamiensia*, 1 : 71-89, 1968.

BICKERTON, D. Cross-level interference; the influence of L1 syllable structure on L2 morphological error. In: PERREN, G. E. & TRIM, J. L. M., eds. *Applications of linguistics*. Cambridge, Cambridge Univ. Press, 1974.

CLARK, R. Adult theories, child strategies and their implications for the language teacher. In: ALLEN, J. P. B. & CORDER, S. P., eds. *The Edinburgh Course in Applied Linguistics. Papers in Applied Linguistics*, London, 1975.

CORDER, S. P. *Introducing applied linguistics*. Harmondsworth, Penguin, 1975.

DUSKOWA, L. On sources of errors in foreign language learning. *IRAL* 7(1):11-36, 1969.

IOUP, G. and KRUSE, A. Interference versus structural complexity in second language acquisition; language universals as a basis for natural sequencing. In: BROWN, H. D., YORIO, C. A. and CRYMES, R. H., eds. *On Tesol '77 Washington, D.C. Teachers of English to speakers of other languages, 1977*.

JOHANSSON, S. The uses of error analysis. In: *Papers in contrastive linguistics and language testing*. Lund, CWK Gleerup, 1975.

RICHARDS, J. C. A non-contrastive approach to error analysis. *Error analysis: perspectives on second language acquisition*. London, Longman, 1977.

SELINKER, L. Interlanguage. In: RICHARDS, J. C., ed. 1977. Op. Cit.

TARONE, E. E. Conscious communication strategies in interlanguage; a progress report In: BROWN, H. D., YORIO, C. A. and CRYMES, R. C., eds. 1977, Op. Cit.