

# ERROR ANALYSIS IN A LEARNER CORPUS. WHAT ARE THE LEARNERS' STRATEGIES?

Willelmira Castillejos López<sup>1</sup>  
*Universidad Autónoma Chapingo (México)*

## Abstract

This corpus-based work is intended to find transfer and intralingual errors made by native Spanish speakers with a high proficiency level of the English language, particularly when the production is written. Recent trends in language teaching are in tune with second language acquisition (SLA) research results, so studying learner outputs is important to help define parameters on the way a second language is learned. At the same time, this means a contribution to teaching. By analyzing errors of learners' production it is possible to distinguish the strategies they trigger in their process of SLA.

The learner corpus used for this research is named ICLE (International Corpus of Learner English) and belongs to the Center for English Corpus Linguistics from the Université Catholique de Louvain (UCL). This corpus contains over three million words, a collection of essays written by learners of English with diverse mother tongues. The Spanish sub-corpus was error-tagged by researchers at the UCL, this was the material used for the study.

**Key words:** learner corpus, transfer error, intralingual error, error-tagging.

## I. INTRODUCTION

### I.1. Corpus linguistics and error analysis

For many years linguistics was criticized for lacking methods that demanded scientific rigour. Linguistic studies were considered part of other disciplines: logics, philosophy, sociology, psychology, etc (Corder, 1992: 81). Fortunately, the works of prominent linguists before and after Saussure have granted linguistics its scientific character. If one of the starting points in the scientific method is the observation of phenomena and the elaboration of hypotheses, followed by the experience and result analysis that will confirm or discard the hypotheses made, then linguistics finds a perfect companion to develop its scientific trajectory in the field of corpus linguistics. Most scholars agree on considering corpus linguistics as a methodology rather than a theory. Mc Enery, Xiao and Tono (2006: 7) consider corpus linguistics as “a whole system of methods and principles of how to apply corpora in language studies and teaching/learning, it certainly has a theoretical status.” These authors claim for the theoretical side of corpus linguistics in comparison to other social science methodologies but also agree that it “should be considered as a methodology with a wide range of applications across many areas and theories of linguistics” (ibid: 8). I believe corpus linguistics brought about a revolution in the linguistic field. It is revolutionary in the

---

<sup>1</sup> **Address for correspondence:** Universidad Autónoma Chapingo (México) Fray Pedro de Gante 321, Col. San Mateo. Texcoco, Edo. Méx. C.P 56110. Tel. 0052 5959555441. [williecastillejos@hotmail.com](mailto:williecastillejos@hotmail.com)

sense that it deals with the study of real language data and not longer with the linguist's intuition. Under corpus studies, linguistics confirms its scientific character through the complete and accurate analysis of real language. Corpus linguistics is also a serious scientific method because it deals with large quantities of data. Statistically speaking, if we want to draw conclusions on an entire population, we must care about the sample selection so as to obtain a size whose proportion and characteristics will allow inferring information that we can apply to the entire population. If language constitutes an enormous, infinite-tending population, the least we can do to infer information on the language is to analyze large quantities of data. The marriage of linguistics and computer science has made this fact possible and now thanks to the tools a researcher can run in corpus treatment software, information on the language that could take days, or even months, if it were obtained manually, is accessed in seconds. Moreover, with the support of computational tools, a corpus can be coded for specific information in the form of tags or parses. Taggers and parses help draw accurate conclusions when the size of data collection is large (Meyer, 2002). All of these aspects can only be qualified as revolutionary.

As for error analysis (EA), corpus linguistics could have a 'vindicating' role. EA has been largely criticized for both methodological and theoretical reasons. The former reason stands, in general, for the weaknesses in error evaluation judgments, lack of precision in defining the point of view under which an utterance is considered erroneous, difficulty to find the interlingual or intralingual source of error, classification and interpretation of errors (see Els, Bongaerts *et al* 1984; and Ellis, 2003). The latter reason involves the comprehension of EA techniques as a means to infer the process of learning a second language. According to this position, EA does not provide a complete idea of the learners in terms of their competence and performance. For EA to be efficient, it needs to study both erroneous and non erroneous forms.

Another criticism is in the SLA area and refers to the cross-sectional nature of research studies (in opposition to longitudinal studies). For this position, EA cannot explain the development of the learner during his/her process of acquiring a second language since analysis is done on a static text, a sort of language photo taken at a certain moment under certain circumstances; however, as Ellis (2003) states, no dispute between longitudinal studies and EA exists since the study of errors through different stages of the learning process can contribute to the understanding of the strategies learners put into practice when acquiring a second language. Another limitation conferred to EA is the learners' tendency to avoid structures they believe to be erroneous, and use those structures they consider as correct. Although this is true, it is also worth remarking that in producing the target language, learners cannot avoid structures all the time. Some part of the output can be the result of avoidance but a great part of is also 'authentic'. Furthermore, if corpus linguistics supposes the study of large language collection, conclusions drawn cannot be that neglected because statistically there is a great part of 'authenticity' in the data. In short, second language learners cannot be 'acting' the whole time they write or speak the target language.

Data is the convergence point of corpus linguistics and EA; both of them are not theories of language acquisition but methodologies for dealing with data. Corpus linguistics provides the object of study and EA determines the technique of study, both of them holding their own rules but making mutual contribution in order to offer results that will enrich SLA theories and language teaching.

In conclusion, together with corpus linguistics, EA is still alive. EA has contributed to understanding important facts in SLA as well as to supporting language pedagogy. What is crucial in exercising EA is the definition of objectives and the setting of the analysis nature. Although contrastive, performance and discourse analysis can provide great contributions to the study of language, EA has also a value on its own.

## II. THE PERTINENCE OF ANALYZING ERRORS

First of all, by analyzing errors we can obtain information about how a language is learned, errors reflect the learner internal constructs, which for Selinker constitute an independent language system called *interlanguage* (Selinker, 1972), and the amount of knowledge a learner has of a language. However, the evaluation of the competence of a learner goes beyond the analysis of errors alone, focus on other aspects such as avoidance of difficult structures is an indicator of the progress made in the learning of a language (Yip, V., 1962: 5). The analysis of errors is however crucial in the SLA field because it allows the observation of actual learner output and gives SLA researchers the possibility to explain how learning progresses. For Ellis (2003), error analysis is also important because it provides the researcher with a methodology to study learner language.

Another reason that justifies the analysis of errors is language teaching. In the evolution of language teaching methods, the study of errors has played a very distinctive role, the first formal methods considered errors as the evidence of *mislearning* a language, they were avoided at all costs, and impeccable oral and written production was the objective for learners. The recent trends of communicative methods focus mainly on developing communicative skills, with errors being of less importance as long as they do not hinder communication. Both scenarios are quite divergent, as one approach considers errors as a demon to exorcise and the other as an eternal unavoidable companion. The truth is that the study of errors offers great advantages for improving language pedagogies; EA results cannot be out of fashion since they evidence those areas of the language teachers need to focus on, areas such as grammar, lexis, discourse, etc. In other words, it is by analyzing errors that important suggestions for language method design can be made, this involves all the areas of the pedagogical design, from syllabus to materials.

Traditionally, language courses are based on a list of items adapted to the contents of a certain book. Without questioning the linguistic quality of books, they hardly ever pay attention to problems that specific users of the target language may have, these specific users are referred to in terms of their mother tongue. The mother tongue of English learners is a crucial fact of homogeneity when designing pedagogies and materials as it is the most important language learning experience subjects have had. In the framework of this work,

the errors studied are from learners with Spanish as a mother tongue, so all the results from error analysis, and their implications on language teaching are mainly attributable to Spanish native speakers, although they can be of use to other groups of learners. In a global pedagogical sense, research studies carried out with error analysis techniques are pertinent because of the implications on the organization of the language syllabus and in the classroom itself since, even though communication is considered the main language function and should be an objective of language teachers, when communication is carried out neatly, with the appropriate structure, lexis, register and style the communication objective is undoubtedly best accomplished. In general terms, the pertinence of error analysis in language teaching could be situated in the middle of the *no errors* tendency (at the expense of communication) and the *only-communication matters* tendency (at the expense of linguistic quality in the message).

This work concentrates on the study of learner errors but errors are not to be considered as language learning defects but as part of the strategies developed by learners in the process of acquiring a second language. The description of the nature of the learner language has interested linguists such as Corder, but also Selinker and Richards among several other SLA researchers. An important conclusion made by these authors is that learner language is a particular system with its own code and its own rules, that keeps a dynamic process, hence its unstable nature. Part of the characteristics cited for Corder's *idiosyncratic dialect* can be explained as Selinker's *interlanguage* although *interlanguage* can be also explained as the result of an interaction of two language systems, usually the mother tongue and the target language, an interaction that produces a language in its own.

When describing learner language, the observation of the learner output conveys certain generalizations on how a second language is learned, but what researchers look for is the reasons for the phenomena observed. In this perspective, Selinker (1972) attributes five reasons to the process of learning a second language; these reasons are briefly explained as follows:

- Language transfer: the output is explained as a result of the interaction with the learner mother tongue.
- Transfer of training: the output is explained in terms of the type of training to learn the language the learner has had.
- Strategies of second language learning: the output is explained by the association the learner makes with the material to learn.
- Strategies of second language communication: the output is explained by the association the learner makes with communication with native speakers of the target language.
- Overgeneralization of target language linguistic material: the output is explained by the overgeneralizations the learner makes of syntactic rules and semantic aspects of the target language.

In a paper written for a volume on EA, Richards (1977) distinguishes *language transfer* as the main source of deviant forms of the learner second language. The other sources cited by this author are *intralingual interference*: learners produce forms that are not

a consequence of interference with their mother tongue but of a partial exposure to the target language, *sociolinguistic situation*: the social context of the target language is ignored by the learner (aspects of motivation are also considered in this point), *modality*: the learners performance is different if their role is as language producer or as language receiver, *age*: the language output differs according to the learners age, *successions of approximative systems*: these systems are unstable and vary from one learner to another and *universal hierarchy of difficulty*: certain linguistic and semantic features are complex *per se* no matter the mother tongue or particular capacities of the learner.

In analyzing learner language all these factors explain the nature and origin of learner errors but apart from these explanations, what is relevant for the objective of this work are the conclusions that can be drawn on the basis of regularity of certain error patterns. Stevens (1969) considered the gradual modification of these error patterns as an evidence of effective learning. These observations find a perfect match in the corpus linguistics work, where the study of regularities is crucial since it determines the most salient phenomena to observe. In this case, the most frequent (and regular) patterns of English learners errors are detected, the sense of regularity becomes even more important when the group of learners is homogenous, i.e. they all have the same mother tongue (Spanish).

### III. THE ERROR-TAGGED CORPUS

After having explained some of the main features to have in mind when developing a work on corpus linguistics, attention is focused on corpus mark-up. This characteristic of the corpus can be defined as the set of codes added to the text so as to obtain information about it and be able to manipulate the processing of the document. Corpus mark-up could be a discouraging aspect when one has a first contact with corpus linguistics; in fact, watching the amounts of ‘incomprehensible’ symbols, numbers and letters around a text provokes a feeling of ignorance but once we are able to understand the codes, we can appreciate the great value of this system.

A term that is related to corpus mark-up is corpus annotation. If the mark-up of a corpus provides great advantages in terms of linguistic research studies, the advantages increase when a corpus is annotated as the information added to the text is linguistic, e.g. it provides information of the grammar class of words.

Tagging can be automatic (the computer program tags the corpus) or manual (there is human intervention in the tagging with the aid of a computer program). Hunston (2002: 18) points out the level of accuracy achieved depending on the automatic or manual tagging, and states that even though in the first case a corpus can be totally tagged, the level of accuracy is lower than that obtained through manual tagging, however this is only feasible when the corpus is small.

The corpus used in this study was a small one (62,982 tokens), it was error-tagged by a group of researchers of the ICLE project, who developed a Windows-based error editor and designed a generic based error tagset based on seven categories of error (Dagneaux *et al*, 1996). A ‘tagset’ is not to be confused with ‘a tagger’. The former term refers to the set of

symbols that represent various parts of speech while the latter concerns the software program inserting the tags in the corpus (Meyer, 2002: 86).

Error-tagging is a specific type of annotation and the corpora on which error-tagging is done is normally associated to second language learners. Maybe the best exploitation of an error-annotated corpus is to observe the frequency of errors made by second language learners and their type. More interesting information is obtained when the study of errors is done with an emphasis on a specific language background, like in this case, where the purpose is to find about the frequency of errors and error types made by learners of English with Spanish as a mother tongue.

In general, tagging can be done in two ways: rule-based or probabilistic; rule-based respects the rules of grammar written into the tagger, probabilistic tagging is “based on the statistical likelihood that a given tag will occur in a given context” (Meyer, 2002: 88). However, for the case of error-tagging the task is more than complicated since determining error patterns can become very arbitrary, and on the other hand, the frequencies of error differ from one group of learners (with a specific language background) to another. That is why the effort of error-tagging learner corpora from ICLE is greatly appreciated; otherwise this analysis would not have been possible.

#### **IV. THE ASPECT OF FREQUENCY**

The type of frequency concerned here is the number of error-tags generated in the corpus. A detailed description of every error type goes beyond the scope of this work, so focus on the highest frequencies is considered. The categories for error analysis are chosen according to the linguistic description that corresponds to the error-tags. Researchers from the Center for English Corpus Linguistics at the Université Catholique de Louvain defined seven main categories for error classification, the categories are: form, grammar, lexico-grammar (that is to say, error incidences on both lexical and syntactic properties of a word), lexis, word redundant, word missing and word order (the three ‘word’ adjectives with the three nouns integrate a single category), register, and style (Dagneaux, E. *et al*, 1996). These categories involve concrete linguistic areas where errors can occur, for instance, the category ‘form’ involves two linguistic aspects which originate errors: morphology and spelling.

The reason to consider only the top frequency of each category is explained in terms of a global learner language, it is interesting to dedicate all the efforts to only one language aspect, the form for instance. But part of the enthusiasm for this work resides in observing as much as possible. Exactly as when we observe a landscape, there is a desire to contemplate it all.

In language teaching, frequency is certainly crucial as items in the syllabus should be chosen on the criteria of how frequent students will be faced to a certain structure, word or language function. It is not very useful to spend time in studying a language aspect that students may hardly deal with, it is better to furnish the syllabus with those aspects of the language that are more likely to occur in the L2 contextual experience of students. With learner errors it is about the same, teachers should be especially aware of those error-prone

elements students tend to fall into more often. Learner corpora constitute an unbeatable instrument to study the learner output, and prediction of the type or errors to be made according to the mother tongue of students becomes in this way an invaluable aspect of language teaching. In the case of a learner corpus, the more errors are observed, the better suggestions for language improvement can be offered.

## V. METHODOLOGY

### V.1. Corpus description

The *International Corpus of Learner English* (ICLE) is the generic name of the corpus used in this study. ICLE is the result of collaborative work among universities from different countries, the whole corpus of over three million words, contains the writing of learners of English from 21 different mother tongues. As a matter of utility, the different groups of learners are stored in different sub-corpora (21), it is exactly part of these sub-corpora that was used in this work: the error-tagged Spanish corpus. The name given to this sub-corpus is ICLESP. It contains the writings of advanced-level students of English as a foreign language who wrote argumentative essays to enrich the big ICLE corpus. The sample used was the following size: 62,982 tokens.

Important conclusions on the learning process of English can be drawn out of this corpus as it is homogenous in the mother tongue of the learners and the writing style, they are all argumentative essays on similar topics.

### V.2. Process of extraction

The information about the size in tokens was provided by one of the tools in *Wordsmith: Wordlist*, which generates statistic data about specific files in the program, providing not only the total of tokens but also information on the length of sentences, lexical richness, headings, etc. The error search was conducted in the following way: since there exists an error manual (See Dagneaux *et al.*, 1996) that explains the different codes used in the corpus annotation, the process to obtain the list of errors was done through *Concord*, the main *Wordsmith* tool used in this work.. The search ‘word’ was each of the error codes from the seven categories and the sorting was 2R and 2L (context search horizons 5L and 5R), ascending distribution. In the cases of article errors, three elements were contextualized: *the*, *the \$0\$* and *0 \$the\$*, and the sorting was 5R and 5L. The categories and subcategories together accounted for 42 searches (pointed out on Table 1). Here is the list of codes and the type of error referred according to the taxonomy used in the error tag set of Dagneaux *et al.*, 1996

**Table 1. Codes and types of error (Source: Dagneaux *et al.*: 1996)**

CODE	TYPE OF ERROR
FM	form – morphology
FS	form- spelling
GA	grammar- articles
GN	grammar – nouns
GNC	grammar - noun case
GNN	grammar - noun number
GP	grammar – pronouns
GADJO	grammar - adjective order
GADJN	grammar - adjective number
GADJCS	grammar - comparative/superlative
GADVO	grammar - adverb order
GVN	grammar - verb number
GVM	grammar - verb morphology
GVNF	grammar - non-finite / finite verb forms
GVV	grammar - verb voice
GVT	grammar - verb tense
GVAUX	grammar – auxiliaries
GWC	grammar - word class
XADJO	lexico-grammar - erroneous complementation of adjectives
XCONJCO	lexico-grammar - erroneous complementation of conjunctions
XNCO	lexico-grammar - erroneous complementation of nouns
XPRCO	lexico-grammar - erroneous complementation of prepositions
XVCO	lexico-grammar - erroneous complementation of verbs
XADJPR	lexico-grammar - adjectives used with the wrong dependent preposition
XNPR	lexico-grammar - nouns used with the wrong dependent preposition
XVPR	lexico-grammar - verbs used with the wrong dependent preposition
XNUC	lexico-grammar - nouns: uncountable/countable
LS	lexis - lexical single
LSF	lexis - false friends
LP	lexis - lexical phrase
LCL	lexis - logical connectors
LCLS	lexis - single logical connector
LCLC	lexis - complex logical connector
LCC	lexis - coordinating conjunctions
LCS	lexis - subordinating conjunctions
WR	word redundant
WM	word missing
WO	word order
R	register
S	style
SI	style - incomplete
SU	style - unclear

Each of these searches gave a different size of error lists, so relative frequencies were calculated in order to compare the type of error Spanish speaking learners are more likely to make when expressing written English. The statistics are provided in the results section.

### V.3. Process of analysis

This section explains how the process of EA was carried out. Corder (1974) suggests three steps in the study of EA: recognition, description and explanation. He also emphasizes the importance of the data to analyze, particularly of how this data has been collected. On this basis and on the fact that this is a corpus linguistics work, an essential task is data analysis in order to draw conclusions of linguistic value, the route to conduct the results was then:

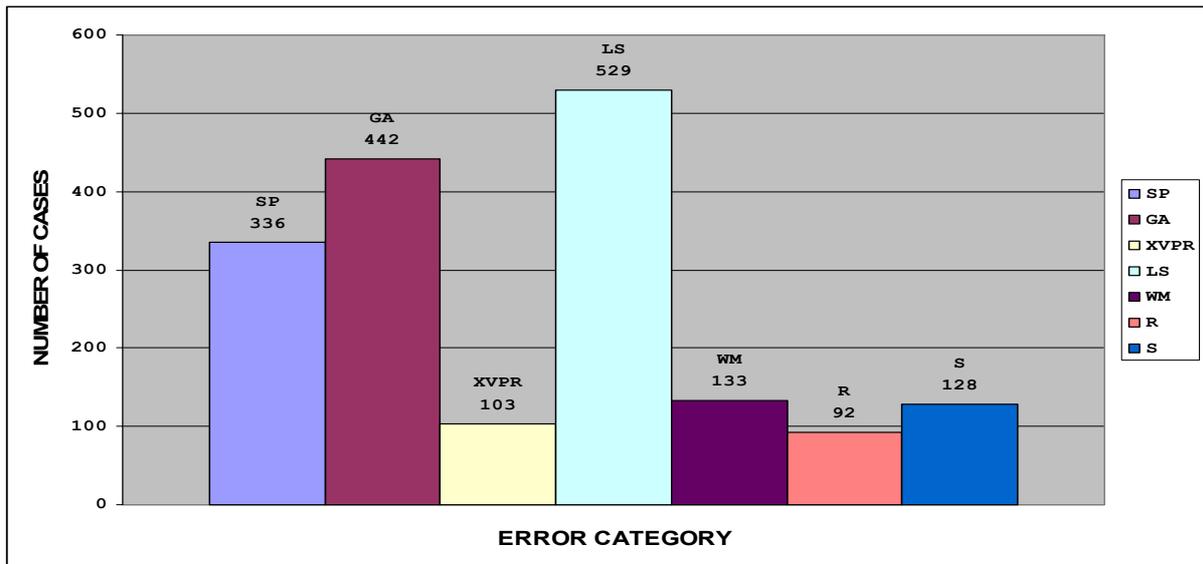
1. Collection of the sample to analyse
2. Recognition of errors
3. Description of errors
4. Explanation of errors

The four steps mentioned above have a space in this work but the level of effort dedicated to each is quite different since great part of the task had been done by the group of researchers working at the Center for English Corpus Linguistics of the Université Catholique de Louvain. It is then assumed that errors have been recognized and described in a sample collected by the research center so what follows is the application of tools to derive statistical results and analyze the nature of errors.

Focus was on two types of errors: *transfer* (as the result of the mother tongue influence) and *intralingual* (influence coming from the target language). If transfer errors are caused by the structure of the first language (that is why they are also called interference problems), intralingual errors do not have an origin in the first language but in the target language itself. As Els, Bongaerts *et al.* state, intralingual errors cannot be submitted to contrast with the first language, they “relate to a specific interpretation of the target language and manifest themselves as universal phenomena, in any language learning process (1984: 51)

## VI. RESULTS

According to concordances run for each of the seven error categories (and their sub-categories), the top frequencies were as shown on Figure 1.



**Figure 1. Top-error frequencies per category.**

SP: Spelling, GA: Articles, XVPR: Wrong verb complementation, LS: Lexical single, WM: Word missing, R: Register, S: Style.

Having each error category a different linguistic nature and a different number of cases, there is not a single or uniform way of explanation, so the main idea that should guide all the analysis is the focus on the search of both transfer and intralingual sources of errors. Transfer and intralingual are not the only causes of error, so not all of the cases occurred in the concordances were considered for analysis. Figure 1 shows the bare results of the concordances but transfer and intralingual proportions are presented in the Table 2. These types of errors were manually counted after close observation of the concordances.

Table 2 below illustrates the absolute and relative frequencies of transfer and intralingual errors found in the corpus, the last column to the right indicates the total percentage of these types of errors. Note that the total does not account for 100% of the errors in the corpus, but only for those whose nature could be explained or justified in terms of a transfer or an intralingual operating principle.

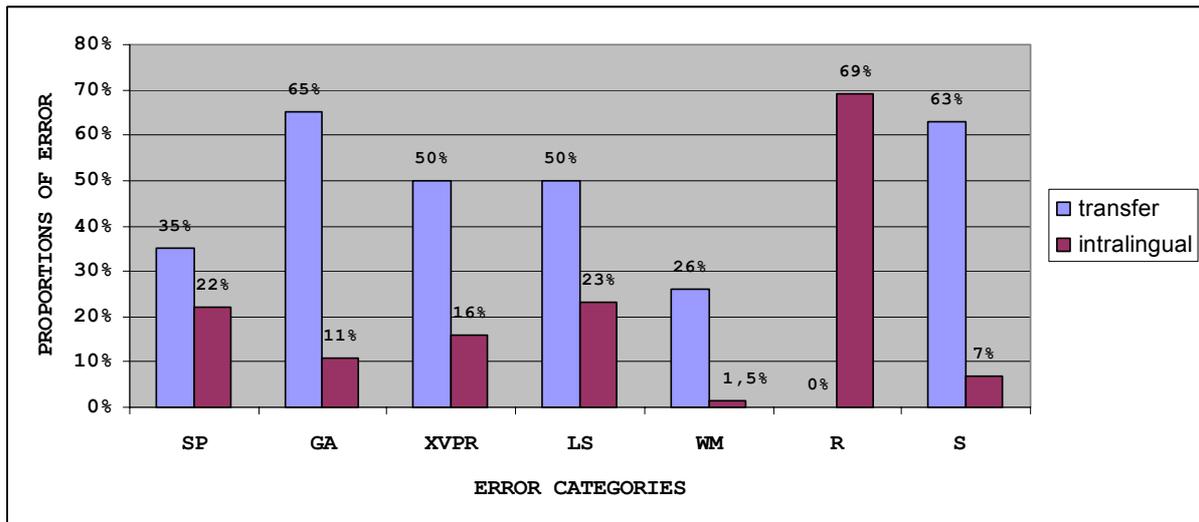
Category of error	Absolute and relative frequencies (%)		Total (%)
	Transfer	Intralingual	
Spelling	105 (35%)	67 (22%)	57%
Articles	287 (65%)	48 (11%)	76%
Wrong verb complementation	50 (50%)	16 (16%)	66%
Lexical single	267 (50%)	122 (23%)	73%
Word missing	35 (26%)	2 (1.5%)	27.5%
Register	0 (0%)	63 (69%)	69%
Style	80 (63%)	9 (7%)	70%
<b>Average (<math>\bar{x}</math>)</b>	<b>41%</b>	<b>21%</b>	<b>62%</b>

**Table 2. Absolute and relative frequencies of transfer and intralingual errors.**

Percentages are obtained out of the total of errors considered for analysis, which do not include concordance noise. However, this does not happen very often, so most of the times the large total corresponds to the number of entries resulted in the concordance.

As observed on Table 2, the total of errors by a transfer or an intralingual reason accounts for 62% of the total of errors in the corpus, but the transfer source doubles the percentage of intralingual sources, so transfer was the main cause of error found in the corpus, which means that four out of ten times an error occurred were due to this cause (cause of error on the L1), and two out of ten were due to intralingual (cause of error on the L2). The implications these results should have on teaching deserve a closer attention on transfer and intralingual phenomena, as McEnery, Xiao and Tono (2006: 65) state “learner corpora are immediately related to the classroom”. It is important to notice that proportions should not be taken on a global perspective; such is the case of register errors, for example, whose proportion refers only intralingual causes. Susceptibility of error depends on the category in question, articles and lexical singles are more numerous in items because they tend to occur more frequently in texts. In addition, objective linguistic characteristics are more detectable than subjective characteristics of the language, such as register and style categories.

In order to have a visual panorama of error tendencies, Figure 2 represents the contrast between transfer and intralingual proportions according to each category.



**Figure 2. Transfer and intralingual error proportions**

P: Spelling, GA: Articles, XVPR: Wrong verb complementation, LS: Lexical single, WM: Word missing, R: Register, S: Style.

The over-presence of transfer errors does not diminish the importance of intralingual errors; they were rather high in relation to other types of errors, except for transfer.

## VII. CONCLUSIONS

It is not easy to determine if an error is a cause of transfer or intralingual nature but as far as linguistic and psychological explanations can be provided we can expect to attribute certain validity to results. What is true is that transfer (on a large scale) and intralingual (on a small scale) do play a role in the learning of an L2 by adult people. Ellis (2003) provides a series of percentages of inference errors reported by different researchers, the proportions are considerably varied:

Study	% of interference errors	Type of learner
Grauberg 1971	36	First language German— adult, advanced
George 1972	33 (approx)	Mixed first languages— adult, graduate
Dulay and Burt 1973	3	First language Spanish— children, mixed level
Tran-Chi-Chau 1975	51	First language Chinese— adult, mixed level
Mukkatash 1977	23	First language Arabic— adult
Flick 1980	31	First language Spanish— adult, mixed level
Lott 1983	50 (approx)	First language Italian— adult, university

**Table 3. “Percentages of interference errors reported by various studies of L2 English grammar” (Taken from Ellis 2003: 302)**

The study carried out here reported 41% of transfer errors (cf Table 2) so it confirms the tendency of some of the research results on Table 3. Notice that the only case of study with children (Dulay & Burt, 1973) differed considerably in relation to the other studies. These facts can make us state that every time adult learner language is to be studied, transfer operating principles will emphatically appear. That being so, L2 course designers should take into account the role of transfer in the learning process. Textbooks do not usually consider learners particularities coming from their MT, they are designated to heterogeneous audiences and global markets. In order to solve this “inconvenience”, teachers are exhorted to make a reasonable planning of the syllabus, materials and activities in the classroom. Teachers should have in mind that adult learners will inevitably base the new learning (L2) on the grounds of the previous one (L1). If transfer takes place, it is time for language teaching to develop more materials aimed at facilitating the L2 learning. On Lado’s work, Selinker concludes the following “[...] the most important new thing in the preparation of teaching materials is the *comparison* of native and foreign language and culture in order to find the *hurdles* that really have to be surmounted in the teaching” (Lado, 1957: 2 in Selinker, 1992: 9-10).

One feature of transfer is selectivity. SLA research studies can contribute to the improvement of language teaching by providing those structures which are potentially transferable among the different languages. In order to achieve this goal, contrastive studies are essential.

It would not be responsible to develop a work on EA without assuming the criticisms made to this method. One of the most important ones is the indeterminate character of error, a problem that might not be solved because it deals with researchers’ own conception of error, i.e. what is an error to someone may not be such for another one. Some areas are more delicate than others in language matters. Register and style errors, for example, are more difficult to be agreed on than morphological or grammatical errors. In the same way, indeterminacy<sup>1</sup> can be found at the level of the proposed correction. When I analyzed the data, I sometimes thought of other solutions, different to the ones proposed in the tagged-corpus. However, indeterminacy should not discourage work on EA because analyzing errors represents an objective form of observing learners’ display of strategies and means to interpret and express a FL/SL. In spite of the problems and the criticism posed to EA and contrastive studies, the results they have provided have made great contributions to applied linguistics. The study of transfer and intralingual phenomena, for instance, has contributed to recontextualize the role of learner errors: whereas under the audiolingual (behaviourist) view errors were an impediment for learning, under the communicative and cognitive perspectives, errors are part of the resources learners trigger during the learning development. On the other hand, the study of errors in a big corpus reflects that learners are certainly attached to rules of a universal grammar and hardly ever make nonsense errors. Except for some isolated cases, all of the errors found have (at least) one logical explanation for what they have occurred. The problem is agreeing on the right source, on the nature of the error. However, EA works agree on the fact that errors of this type: *She was absolute to*

*call me mad* are not likely to occur. As James states (1998: 15) “[...] people’s intuitions tell them that such things cannot be said in such ways in human languages.”

EA can not so disappear as long as there are learner errors, as James (1998: 63) puts it: “Ignorance is not synonymous with incompleteness, but as long as there is incompleteness or failure to attain full NS-like knowledge of the TL, there will be EA.” And EA can now be explained through patterns resulted from a large computer corpus.

Corpus linguistics is in direct contribution to language teaching. It can help teachers visualize the language as source, target and something in between. In the end, language teaching is influenced by the way teachers understand the language but teachers hardly ever make their own conceptions of language explicit, corpus linguistics may represent a means to achieve this goal.

Studying learner errors could be a first step to introduce teachers to the knowledge of learner’s language, but it is just a starting point to discover the multiple nuances learning an L2 entails. At the same time, studying learner errors involves approaching learning in an intimate way; this would enable teachers to promote appropriate pedagogical tools: it is by understanding the nature of a phenomenon that we can better explain it and tackle it. Teaching an L2 demands an effort of continuous search, but it is such a passionate task that all efforts are worth it.

---

## NOTES

<sup>1</sup> Some cases of error indeterminacy can be found in James (1998: 11), the author mentions research works on EA developed by experienced teachers of EFL/ESL.

## ACKNOWLEDGEMENTS

Special thanks to the staff of the Center for English Corpus Linguistics at the Université Catholique de Louvain, especially to professors Fanny Meunier, whose guidance and observations were crucial to this work; and Sylviane Granger, for her authorization to work at the Center and use a section of the ICLE corpora.

## BIBLIOGRAPHY

- Andersen, Elaine S. (1992). *Speaking with Style*. New York: Routledge.
- Celce-Murcia, M. and L. McIntosh (1979). *Teaching English as a Second or Foreign Language*. Massachusetts: Newbury House Publishers, Inc.
- Corder, S. P. (1974). *Error Analysis*. In Allen J. P. B. and Pit Corder (1974, editors). *Techniques in Applied Linguistics* (The Edinburgh Course in Applied Linguistics). London: Oxford University Press.
- Corder, S. P. (1978). *Language-Learner Language*. In Richards, J.C. (1978, editor). *Understanding Second and Foreign Language Learning*. Massachusetts: Newbury House Publishers, Inc.
- Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press
- Corder, S. P. (1992). *Introducción a la lingüística aplicada*. México: Limusa
- Courchêne, R.J. et al (1992, editors). *Comprehension-based Second Language Teaching. L’enseignement des langues secondes axé sur la compréhension*. Canada: University of Ottawa Press.

- 
- Dagneaux, E., S. Denness, S. Granger and F. Meunier (1996). *Error Tagging Manual*, Version 1.1. Center for English Corpus Linguistics. Louvain-la-Neuve: Université Catholique de Louvain.
- De Bot Kees, Lowie Wander and Verspoor Marjolijn (2005). *Second Language Acquisition*. London and New York: Routledge.
- Douglas Brown H. (2000). *Principles of Language Learning and Teaching*. NY: Longman.
- Ellis, Rod (2003). *The Study of Second Language Acquisition*. UK: Oxford University Press.
- Els Theo Van, Theo Bongaerts *et al* (1984). *Applied Linguistics and the Learning and Teaching of Foreign Languages*. London: Edward Arnold.
- Gass, S. and C. Madden (1985). *Input in Second Language Acquisition*. Massachusetts: Rowley.
- Gass, S. and L. Selinker (1992, editors). *Language Transfer in Language Learning*. Amsterdam/Philadelphia: John Benjamins Publishing Co.
- Halliday, M.A.K. (1979). *Language as Social Semiotic*. London: Edward Arnold.
- Hendrickson, J. (1979). *Error analysis and Error correction in Language Teaching*. Singapore. Seameo Regional Language Center. Occasional papers No. 10
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- James, C. (1998). *Errors in Language Learning and Use*. England: Longman.
- Jaworski, A. and Nikolas Coupland (2000, editors). *The Discourse Reader*. London: Routledge.
- James, C. and Peter Garrett (editors, 1995). *Language Awareness in the Classroom*. New York: London.
- Johnson K. and H. Johnson (2000). *Principles of Language Learning and Teaching*. New York: Longman.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. England: Longman.
- King, C. and N. Stanley (1983). *Building skills for the TOEFL*. Boston: Nelson.
- Matthiessen C. (1993). *Register in the Round: Diversity in a Unified Theory of Register Analysis*, in Ghadessy M. *Register Analysis* (1993). London: Pinter Publishers.
- McEnery, Tony and Andrew Wilson (2001). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery Tony, Richard Xiao and Yukio Tono (2006). *Corpus-Based Language Studies*. London and NY: Routledge.
- Meyer, Charles F. (2002). *English Corpus Linguistics*. Cambridge: Cambridge University Press.
- Nunan, D. (2000). *Language Teaching Methodology*. Malasia: Longman.
- Odlin, T. (1996). *Language Transfer*. Cambridge: Cambridge University Press.
- Ramón, Noelia (2006). Multiple Modification in English and Spanish NP's. *International Journal of Corpus Linguistics*. Volume 11. No. 4 (p. 463-495). John Benjamins Publishing Company.
- Richards, J. C. (1971). A non-contrastive approach to error analysis. *English Language Teaching* No. 25 (p. 204-219).
- Richards, J. C. (1997). *Error analysis, Perspectives on SLA*. London: Longman.
- Richards J. and T. Rodgers (2003). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.
- Schachter J. (1974). An Error in Error Analysis. *Language Learning. A Journal of Applied Linguistics*. Volume 24. No. 2 (p. 205-214). Edited by Douglas Brown.
- Seco, R. (1982). *Manual de Gramática Española*. Madrid: Aguilar.
- Selinker, L. (1992). *Rediscovering Interlanguage*. London: Longman.

- 
- Steer, J.M. and Karen A. Carlisi (1997). *The Advanced Grammar Book*. Massachusetts: Henle and Henle Publishers.
- Taylor, G. (1986). Errors and Explanations. *Applied Linguistics*. Volume 7. No. 2 (p.144-166). Oxford University Press.
- Van Roey, J. (1990). *French-English Contrastive Lexicology*. Série Pédagogique de l'Institut de Linguistique de Louvain. Leuven: Peeters.
- Yip, Virginia (1995). *Interlanguage and Learnability / from Chinese to English*. USA: John Benjamins Publishing.
- ZhaoHong Han (2004). Fossilization: five central issues. *International Journal of Applied Linguistics*. Volume 14. No. 2 (p. 212-242). Blackwell Publishing Ltd.