7th International Conference on Corpus Linguistics: Current Work in Corpus Linguistics: Working with Traditionally-conceived Corpora and Beyond (CILC 2015)

A case study on oral corpus: The use of mother tongue in class by Brazilian teachers of Spanish as Foreign Language

Sabrina Lafuente Gimenez*

Universitat Jaume I; Universidade Federal de Santa Catarina; Av. JaumeI, 147. 2A – 5ªF. Benicassim, 12560, España

Abstract

On the features of the upcoming languages, some recent empirical research (Durão and Andrade, 2010) estimated that Spanish and Portuguese share approximately 85% of the lexicon in several semantic fields. Other authors such as Almeida Filho (1995) claim that among romance languages, Portuguese and Spanish are the ones that have most affinity. In consequence, by reason of the similarity (morphological, syntactic, semantic and phonetic phonological), there are no Luso-pupils considered as "beginners" in Spanish (Carmolinga, 1997). It occurs because they normally have acquired the ability to understand part of the language, spoken or writing. On the other hand, and, paradoxically, one of the greatest difficulties for such students is to overcome the existing similarities in the two languages, which facilitate the interference from the mother tongue in the L2.

It is not always easy, even in very high levels of knowledge of the foreign language, to get rid of some specific aspects of the mother tongue. In the case of Spanish teachers we have to add the concern by the influence that the L1 interference in their speech can be affected in the student's learning. For that reason, this study that is part of a doctoral thesis investigation attempts to find out if in the speech of the population analysed - Brazilian teachers of Spanish as a foreign language - there are signs of influence from the mother tongue. The work is supported by corpus linguistics. The collection of data is done through the recording and transcript of classes and audio interviews, which constitute the oral corpus (Du Bois, 1991).

Keywords: interference; corpus linguistics; corpus oral; contrastive analysis; Spanish foreign language.

* Corresponding author. Tel.: +34688990184
  E-mail address: sabrina.lafuente@hotmail.com

1. Introduction

This research is part of a doctoral thesis in process of development, fruit of a joint supervision agreement between Universidad Jaume I (Spain) and Universidade Federal de Santa Catarina (Brazil). The aim of this paper is to present some previous results of the work that is under development and, because it is an investigation that is being conducted, no final conclusions will be presented, but it will attempt to discuss the impact of these first results.

In the context of learning a foreign language, we cannot ignore mother tongues influence, widely discussed in recent decades. In the case of very similar languages, such as Portuguese and Spanish, we know that the similarity is not always a facilitator factor in learning, even at more advanced levels as we investigate in this work.

It has been studied much about transfers in the interlanguage of the learners but with different denominations (Weinreich, 1953), Crosslinguistic influence, (Kellerman, 1995) borrowing (Corder, 1981) bidirectional transfer (Ellis, 1994), among others, but there is still a long way to go when we refer to the persistent features of the mother tongue in those individuals considered to be "Advanced" within the overall framework of linguistic competence.

For these reasons, the overall objective of the thesis is to find out if the oral discourse of the studied population (Brazilian teachers of Spanish as a foreign language), presents signs of interference from their mother tongue, identify them and explain them, as well as recognizing which types are more common. In this small reduction of the research, we will present some data on the use of the mother tongue in class, feature observed in the corpus and that especially caught our attention.

2. Corpus compilation

For the compilation of the corpus we used the four steps presented by Thompson, (2004):

1. Data collection
2. Transcription
3. Mark-up and annotation
4. Access

The audio *data collection* occurred at two different times. In the first one, teachers were directed to record at least one Spanish language class and a maximum of five, in the group of their choice. We have not interfered in the choice of the group or the content because we do not believe it to be relevant to the study, since we will only focus on the linguistic record of oral production and wanted that the subjects were as comfortable as possible to do so. In other words, we aimed that the collection of data was as little intrusive as possible. Therefore, teachers were allowed to choose the group of students and the subject in which they believed they could be as natural as possible and classes could offer the optimal recording conditions. Recordings were performed under the supervision of the investigator, although it has exempted their presence in class in order to prevent alterations in the results caused by an increase in the level of anxiety of pupils and teachers. The second moment consisted of a semi-structured interview with the teachers, consisting of twenty-five questions about their training and working life. The need to arrange an interview comes from the fact that, on the one hand, we suspect that not always speaking in the classroom environment is spontaneous and, on the other hand, according to Tarone (1983), in interlanguage, the types of errors may vary according to the context which produces speech.

The second part, the transcript, was done manually since the complexity of the corpus, mainly in reference to the recordings of the classes, where, in addition to the interventions of the students that result in joint and cross talk there is a constant code change by the teacher, and even mixes between Portuguese and Spanish in his speech. Therefore, using specific programs for transcribing resulted in an inefficient task. The only human transcriber has been the investigator of the study and focused on the orthographic level only, since it will provide us with the ideal material for the type of analysis we intend to do. Some signs of Convention were adopted and adapted from

Marcuschi (1986) in order to organize the corpus and facilitate interpretation and treatment of data. On the other hand, the formulated principles of vertical organization of transcription by Edwards (1993) were taken into consideration, what lead to a linear database which facilitates the search for information and facilitates the understanding of the context.

Special care has been taken with the protection of personal data of both teachers and students. Although the teachers have accepted to participate in the study voluntarily and have signed a letter in which they yielded the rights to use the compiled material, all sources of identification of the participants in the study have been replaced by the sign [*], including the names of the students and institutions, as well as any other source of identification of any person or public or private company associated with the study as we can see in the examples below:

   a. D5: (…) Silencio... [*] guarda el cuaderno ahora...
   b. D9: [*]... preste atención... primera respuesta... lunes...
   c. D3: no... nunca... El año pasado tuvimos una maestra a través de [*] [[nombre de la secretaría de educación]]... que trabajó con nosotras un semestre... veinte horas… vino de [*] [[nombre de la universidad]].

The speech represented in examples "a" and "b" shows the deletion of the name of the students to whom they are directed. Example "c", as the context may not offer us the reference, small notes were introduced in the same corpus to allow the understanding of the text.

In the third part, concerning the *annotations*, it was used the minimum of them necessary for the understanding of the context, using always the same convention. As we can see in the example "c" above, two brackets [[ ]] were used so that they were easily identifiable as components added by the transcriber. Other cases in which notes have been used were, for example, events that occurred during the recording which might affect the speech as interruptions, external noise and behavior of pupils, etc.

On the last part, *access*, we must say that for now, the corpus is not published in printed or digital. Because it is a small and very specific corpus, we do not believe it can be published at present, although we do not discard that it might be incorporated to a bigger and specific corpus in the future.

3. Population: sociological and educational profile

The studied population is made up of ten teachers of Spanish as a foreign language. They work in formal schools of three towns in Brazil and are between one and 15 years in the profession. The institutions where they teach are of primary and secondary education, public or private. All of them are originally from the southern region of Brazil, aged between 29 and 51 years old, female and have taken the same degree at the same university, with exactly the same Spanish language teachers. Most research subjects (8 out of 10) have learned Spanish as a foreign language during their degree, since there is no minimum language level for access to university studies.

Six of them have some kind of postgraduation in the area of education or languages. Only one has done a training course in a foreign university and two of them resided in a Spanish speaking country for periods of one and seven years. Two of them have never had contact with a native speaker of Spanish language.

4. Characteristics of the corpus

Corpus has a total of approximately 58,000 words, of which 65% come from the context class and 35% from the interviews, as we can see in the chart below where are presented the total numbers by subject and collection environment. It is important to emphasize that were counted only the turns of teachers, therefore, speeches of the students and the interviewer were excluded.

Table 1. Number of corpus words

Participants	Classes	Interviews	Total
D1	1604	1297	2901
D2	2345	2525	4870
D3	1077	1080	2157
D4	2889	924	3813
D5	3789	2132	5921
D6	5847	1492	7339
D7	5078	3397	8475
D8	6505	3084	9589
D9	5081	1049	6130
D10	3311	3451	6762
TOTAL	37526	20431	57957

In this article we will discuss one of the aspects that most attracted attention in the first analysis of the corpus, which is the use of the mother tongue, especially in the foreign language classes but also in the interviews. Thus, we have classified Portuguese recurrences in seven types, which have been entered in the corpus through the following tags (both for its count, as for the subsequent analysis, we used the program WordSmith Tools):

Table 2. Tags from mother tongue interventions

	Tag	Description
1	[comt. port]	Fragment / comment in Portuguese by the teacher
2	[transl. port]	Fragment translated into Portuguese by the teacher
3	[expl. port]	Fragment explained in Portuguese by the teacher
4	[cuest. port]	Question made in Portuguese by the teacher
5	[order port]	Order given by the teacher in Portuguese
6	[neg. port]	Negation in Portuguese
7	[afirm. port]	Affirmation in Portuguese

Labels have been important to try to understand how and when the teacher made the direct use of the native language in class. From the tags inserted in the corpus, and with the help of the context we have been able to analyze more objectively that phenomenon. Here is an example:

D4: *[[lectura]] Escriba los numerales que hay en el texto... [[fin lectura]]* ¿Qué numerales <u>que</u> aparecen? <u>Aparece</u> el once y doce… *[[lectura]] Escribe por extenso... [[fin lectura]]* ¡Mucha charla! [trad. port]... *[[lectura]] Busque pronombres personales... [[fin lectura]]* ¿Qué pronombres aparecen en el texto? (…)

[afirm. port]… Tú… Aquí hablamos de saludos... ¿tá? (…) Vosotros... él... usted... tú… [expl. port]... (…) <u>No puede</u> ir al baño… [orden port].

In this example we can see a strong trend in the routine of nine of the ten participating teachers of the study. It is very common to use the mother tongue to give orders, or explanations of the tasks. When it comes to activities that are found in textbooks, it is common to use the foreign language to make the reading of them, and then the mother tongue to make the explanation, without being the vehicle of communication in the foreign language.

It avoids that spontaneous talk be collected since, as we have seen in the examples, good part of the speech in the foreign language of the teachers is camouflaged between the readings of the exercises and the use of the mother tongue when there is the need for communication.

Table 3. Use of L1 in class

	Classes	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Total
1	**[comt. port]**	45	44	81	44	48	66	4	18	13	13	**376**
2	**[expl. port]**	45	30	2	106	35	29	15	71	28	6	**367**
3	**[transl. port]**	8	10	1	16	13	6	8	73	19	10	**164**
4	**[cuest. port]**	19	5	3	18	6	3	1	28	26	8	**117**
5	**[order port]**	16	22	0	7	41	4	2	2	16	5	**115**
6	**[afirm. port]**	0	12	2	1	3	2	0	0	3	9	**32**
7	**[neg. port]**	1	1	0	0	1	1	0	1	0	1	**6**
	Total	**134**	**124**	**89**	**192**	**147**	**111**	**30**	**193**	**105**	**52**	**1177**

We can see that the *comments (***[comt. port])** are responsible for the majority of relapses of the use of the mother tongue, with 32% of the total. That happens, we believe, because the comments are extra content of the class information, made by the teacher, with the intention of reporting any activity, occurrence or situation that has nothing to do specifically with the class or activity with which they are working. Thus, the teacher understands that this message should not be given in a foreign language, but in the language of the participants of the event, what appears to us to be quite natural and, somehow, even expected.

Likewise, we understand that, perhaps, the teachers prefer to make *explanations* (**[expl. port])** of activities or aspects of the language specifically in the L1, since they are more certain that the message is being conveyed in a way that can be understood, keeping only the possible doubts on the issue and not about the language used for communication. However, we have to say, many times the mechanics performed in the explanation was to explain in the L2, and then clarify it in the L1, as we can see in the following extract:

D4: (…) Ahora ustedes <u>vão fazer</u> la interpretación de esta viñeta... ¿tá? La primera... Traduce libremente... Ya hicimos la lectura del texto... <u>Fídeos</u>... [trad. port] [expl. port]... La número tres es una pregunta personal. *[[lectura]] La historieta nos presenta algunas situaciones desagradables que a menudo ocurren en los supermercados... ¿Cuál de ellas te parece la peor y por qué?* [expl. port]... *Cuatro... Identifica el texto... una frase que se traduce una ironía... una amenaza... un pedido de disculpa... una resignación... [[fin lectura]]* [expl. port]... una sospecha... [expl. port]... y la número cinco [expl. port] (...)

In the submitted extract, we can clearly identify the tendency for explanations in both languages, as well as directly in Spanish. Hence we can interpret two different situations. One, in which the teacher, who has the students

in front, realizes that they are not understanding what is being explained, and therefore repeat it later in their mother tongue to make sure that the message is understood. And, on the other hand, might have a preconceived idea of who will not understand it, and to avoid any doubts or confusion, in a class where the subject is taught to, usually about thirty students, to get used to using the two languages. Anyway, such practice may lead students to a relaxation in the effort to understand the speech of the teacher in the foreign language, since they already dominate the dynamics and know that after the explanation they will have its version in the mother tongue.

What draws our attention is the fact that there is a high incidence of *questions* and *orders* in the native language. Such attitudes lead us to sense that there is a need to speak in native language so that the teachers are understood by students, which occurs less frequently in interviews, as we can see in the chart below.

Table 4. Use of L1 in interviews

	Interview	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Total
1	[comt. port]	2	1	2	1	6	0	0	0	0	0	12
2	[cuest. port]	0	0	1	0	5	0	0	0	0	0	6
3	[neg. port]	0	1	0	0	3	0	0	0	0	1	5
4	[expl. port]	0	0	3	0	1	0	0	0	0	0	4
5	[afirm. port]	0	0	0	0	1	0	1	1	0	0	3
6	[transl. port]	0	0	0	0	0	0	0	0	0	0	0
7	[order port]	0	0	0	0	0	0	0	0	0	0	0
	Total	2	2	6	1	16	0	1	1	0	1	30

As we can see, in interviews the subjects use the mother tongue less frequently than they do in classes. It seems to be related more closely to a communication resource that respondents use so that the other part can understand them when they do not recognize how to express themselves properly in the L2, since both have the same L1. The comments here play a different role than those of classes, since in the interviews the comments mean that they introduced a new mention to something related to the topic, and not related to something external as we observed in classes. The comments are not related to another topic necessarily, but on the contrary, a mention of something which, for some reason, they do not want to do in the foreign language, whether because they do not dominate the vocabulary or the necessary structures for the development of the conversation, or because they feel more comfortable using their native language for certain matters, as in the example below, where the teacher tries to explain a recent change in the structure of basic education in Brazil.

E: …y empieza en el sexto año... que es el quinto de los antiguos…
D5: [comt. port]...
E: ¿entonces empieza en el sexto año?
D5: [comt. port]
E: ¿y en la educación infantil ya no hay español?
D5: no... solamente inglés…

Maybe what leads teachers to use Portuguese in situations similar to the one above, are the comfort or the lack of knowledge of the appropriate terms and the doubt about its use. Other frequent features were the unfinished responses when they were in doubt about any word or expression that they needed to use and the questioning to the interviewer about the appropriate form in the L2. Anyway, considering the number of incidents detected, we can say that, perhaps, in the interview the teachers used strategies such as paraphrase responses and they have the opportunity to think with a little more time that in class, and it can lead them do not need use native language as in another situations.

5. Conclusions

The development of an oral corpus is a complex task and requires a meticulous task of preparation before you begin your analysis. The objectives of the research must not be forgotten in this laborious task, in all its stages. Consistency is the key to ensure uniformity in the corpus and it can respond to the objectives of the study. This includes defining the conventions of transcription with much attention and be prepared for changes and rehabilitation of corpus during the task of transcription.

As already stated (Cook, 1995) the record remains inevitably partial. Therefore, the annotations must be especially careful, since are an important part of the work and often open new doors to research. In some studies the correct use of annotations may lead to data that was not expected during its the design, but that may be the key to reach the objectives more concisely, as we believe happened with our study.

As mentioned in the introduction, this research is still in the development phase and, therefore, it is still early to present final results. However, it seems clear that the use of the mother tongue in class is more recurrent than latest methods suggests. The mechanics seem to reflect, on the one hand, the need for the teacher to use the mother tongue in certain situations extra content class, but, on the other hand, a dynamic of translations and explanations in both languages in every moment does not seem to be clear provides benefits to the development of the student's interlanguage. On that aspect, we should investigate further the input that the students would be receiving and, perhaps, in a longitudinal study, verifying its influence on the development of the mother tongue of L2 learners.

References

Almeida Filho, J. C. P. (1995). *Português para estrangeiros – interface com o espanhol*. Campinas: Pontes Editores.

Carmolinga, R. (1997). A distância da proximidade - A dificuldade de aprender uma língua fácil -. *Intercâmbio. Revista Do Programa de Estudos Pós-Graduados Em Linguística Aplicada E Estudos Da Linguagem*.

Cook, G. (1995). Theoretical issues: transcribing the untranscribable.No Title. In G. Leech, G. Myers, and J. Thomas (Eds.), *Spoken English on Computer* (pp. 35–53). Harlow: Longman.

Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.

Du Bois, J. W. (1991). Transcription design principles for spoken discourse research. *Pragmatics*, *1 (1)*, 71 - 106.

Durão, A. B., and Andrade, O. G. (2010). Algumas questões referentes à aproximação da lingüística contrastiva e as ciências do léxico. *Revista Trama*, *1 (1)*, 9 - 18.

Edwards, J. (1993). Principles and Contrasting Systems of Discourse Transcription. In J. Edwards, and M. Lampert (Eds.), *Talking Data: Transcription and Coding in Discourse Research* (pp. 3–32). Hillsdale: Lawrence Erlbaum Associates.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Kellerman, E. (1995). Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics*, *15*, 125 – 150.

Marcuschi, L. A. (1986). *Análise da conversação*. São Paulo: Ática.

Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, *4 (2)*, 142 – 164.

Thompson, P. (2004). Spoken Language Corpora. In M. Wynne (Ed.), *Developing Linguistic Corpora: a Guide to Good Practice* (pp. 1-12). Oxford: Oxbow Books.

Weinreich, U. (1953). *Languages in Contact. Findings and Problems.* New York: Linguistic Circle.