Designing a search interface for a Spanish learner oral corpus: the end-user's evaluation

Leonardo Campillos Llanos

Laboratorio de Lingüística Informática
Universidad Autónoma de Madrid (Spain)
leonardo.campillos@uam.es

Abstract

This article summarizes the evaluation process of an interface under development to consult an oral corpus of learners of Spanish as a Foreign Language. The databank comprises 40 interviews with students with over 9 different mother tongues collected for Error Analysis. XML mark-up is used to code the information about the learners and their errors (with an explanation), and the search tool makes it is possible to look up these errors and to listen to the utterances where they appear. The formative evaluation was performed to improve the interface during the design stage by means of a questionnaire which addressed issues related to the teachers' beliefs about languages, their opinion about the Error Analysis methodology, and specific points about the interface design and usability. The results unveiled some deficiencies of the current prototype as well as the interests of the teaching professionals which should be considered to bridge the gap between technology development and its pedagogical applications.

Keywords: learner corpus, Spanish oral language, end-user's interface evaluation

1. Introduction

The resource here presented is aimed at the computer-aided error analysis of learners' speech, an approach to study oral interlanguage which may complement those followed in projects for English (e.g. LINDSEI, Gilquin et al., 2010) or Spanish (e.g. Díaz, 2007; or the SPLLOC corpora, Mitchell et al., 2008). In Learner Corpus Research (LCR), data has been analyzed by means of concordance programs (e.g. Wordsmith Tools) or *ad hoc* tools created for each project (see Granger et al., 2007, O'Donnell, 2008; Díaz Negrillo, 2009). Our research dealing with spoken production, we also needed to build a tool that could retrieve the audio along with the error-tagged transcriptions and be easy to use by teachers of Spanish. The suggestions made by these professionals proved to be very helpful in the development stage.

2. Corpus description

The corpus comprises forty interviews with learners of Spanish from over nine mother tongues (henceforth, L1), which typologically belong to the Romance family (Italian, French and Portuguese), the Germanic (English, German and Dutch), the Slavic (Polish), the Sino Tibetan (Chinese) or the Altaic (Japanese); besides, there is also another mixed group of students (one Finnish, one Korean, one Turkish and one Hungarian). To gather comparable data, four learners from the elementary level (A2, *Common European Framework of Reference*) or the threshold level (B1) were recorded in a semi-structured interview where they had to accomplish different tasks (retelling two stories using pictures and describing two photographs). In the final part of the conversation, learners had to give their opinions about different topics (e.g. differences in food habits) in order to obtain a more spontaneous speech. Every interview is about 15 minutes long, and more than 13 hours have been collected.

Phenomena from speech (dysfluencies, pauses, repetitions, overlappings, etc.) and from the second language acquisition process (misformations, mispronunciations, etc.) are marked in the transcriptions of the interviews, using an adaptation of the CHAT format (as it was used in the C-Oral-Rom project; Cresti and Moneglia, 2005) and the conventions used in SPLLOC (Mitchell *et al*., 2008). The XML format is used to code information about the learner, the transcription of the utterances (which have been manually synchronized with the corresponding fragment of sound) and to annotate by hand errors within every utterance (in order to retrieve the wrong contexts).

One of the goals of the project is to study the learners' oral production with the Error Analysis methodology defined by Corder (1971) –identification, classification, explanation and evaluation, distinguishing between *errors* and *mistakes*–. Although we follow Lennon's definition of error (1991), the characteristics of oral discourse are taken into account: reformulated structures were considered mistakes –thus, they were not added to the error count–, and common structures in spontaneous speech are not marked (e.g. the omission of some particles: *me acuerdo* [*de*] *que…*, 'I remember that').

For this purpose, an error typology has been designed based on previous studies for English (James, 1998; Granger et al., 2003; Nicholls, 2003) or Spanish language (Fernández López, 1997; Vázquez, 1999). The XML error tags, which were devised from this error typology and taking into account previous research projects (e.g., Lüdeling et al., 2005; Granger, 2007), include the distinction between non-ambiguous and ambiguous errors (e.g. a mispronunciation which could also be interpreted as a formal error at the lexical level: *pequeño [peˈkiɲo]*).

Following are the criteria for error classification:

- Part of Speech (e.g. article), Phonological element (segment or suprasegmental) or Syntactic category (e.g. verb phrase).
- Target modification (i.e. the mechanism of the error):
  - Blend: e.g. *novelas de *históricos ⇒ históricas* or *de historia* ['historical novels' or 'about history']
  - Misformation: e.g. **higenias ⇒ higiene* ['hygiene']
  - Misselection: e.g. **es lejos ⇒ está lejos* ['It is far']
  - Omission: e.g. article: **camarero está* ['*waiter is']
  - Unnecessary: e.g. **un mi amigo* ['*a my friend']
  - Order: e.g. *ayuda *me ⇒ me ayuda* ['It helps me']
- Linguistic level:
  - Grammar: e.g. conjugation: **sé ⇒ soy* ['I am']
  - Lexis-Semantics: e.g. coinages: **omeleta ⇒ tortilla*
  - Pragmatics-Discourse: e.g. coherence or cohesion.
  - Pronunciation: e.g. wrong word accent.
- Type of error (e.g. indicative/subjunctive, *ser/estar*).
- Etiology (cause of the error):
  - Interlinguistic (e.g. false friends as *realizar*).
  - Intralinguistic (developmental errors).
  - Unknown (e.g. induced errors).

More details about the corpus are briefly explained in Campillos Llanos (2011).

3. Description of the interface

Together with the work of collecting, transcribing and analysing the corpus, a hypertext-based interface has been developed using XML technologies (XML markup, XSL to display the data, an XML native database and XQuery to perform every search). To select which errors to display, the user is presented with a menu to choose among criteria such as category (e.g., article errors), linguistic level (e.g. Grammar), error type (e.g. conjugation), the learner's L1 or his/her proficiency level, among others. The tool makes it possible to retrieve the utterances along with the fragment of sound where the searched error appears. Every recorded utterance can be listened to along with the explanation of the error, and it is also possible to look up information about the learner such as L1, languages spoken, time studying Spanish or time in Spanish speaking country (Figure 1).

Regarding the user definition, the decision has been to build a tool aimed at teachers of Spanish as a foreign language or specialists in Applied Linguistics. A future possibility could be to adapt it to be used with students of Spanish or Spanish teacher trainees; besides, all the data gathered could benefit NLP researchers interested in automatic analysis of learner language.
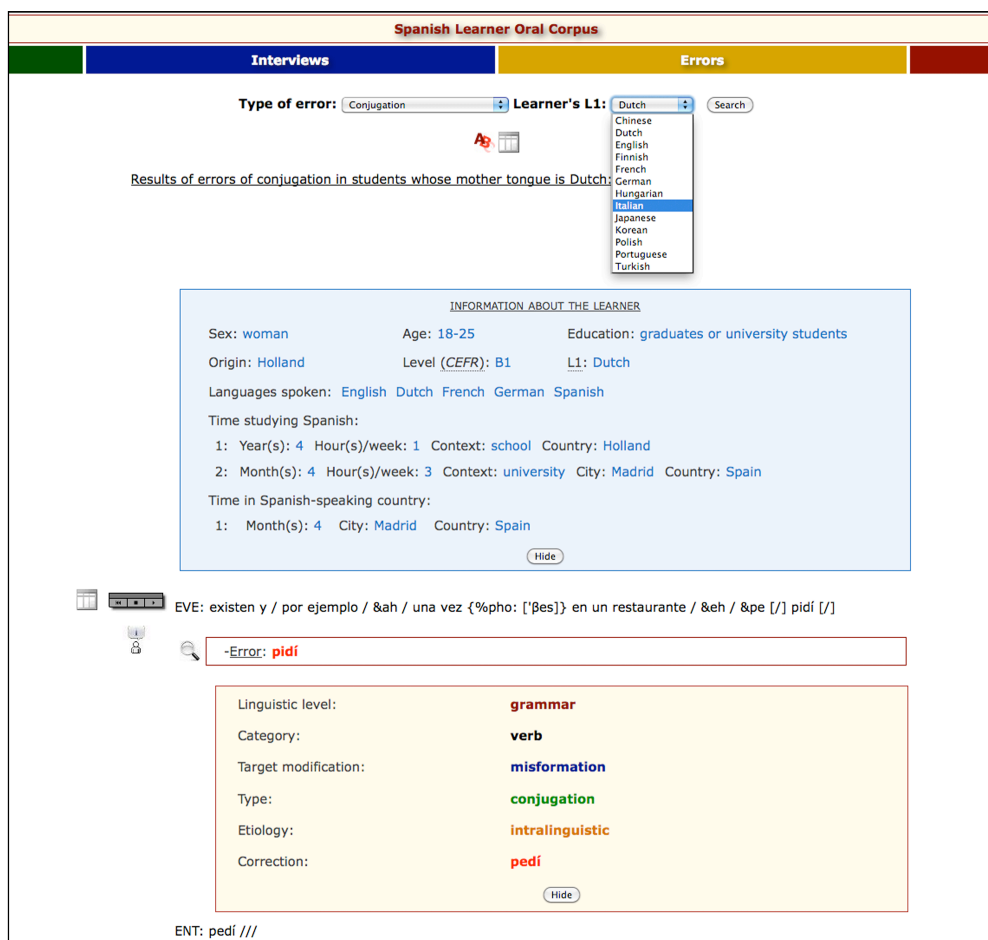

Figure 1: Screenshot of the tool with the menu for type of error, the results retrieved and information about the learner

4. Interface evaluation

The interface design and development methodology adopted has mainly been user centred (end-user design), following the recommendations of Ward (2006). Consequently, we have carried out the formative evaluation of the tool, a key stage conducted during the system development process of a Computer-Assisted Language Learning (CALL) program (Koller, 2007: 139-ss.). The aim is to incorporate user's needs and requirements while the system is being developed, in keeping with the end-user design upheld by Colpaert (2004; see Ward, 2006: 134).

4.1 Methodology

The evaluation for the current version of the corpus search interface was performed during the first months of 2011. The evaluators (N = 22), whose participation was voluntary, were teachers of Spanish as a foreign language, and they were sent the URL address of the search prototype and the same questionnaire through a mailing list aimed at these professionals (FORMESPA); afterwards, they sent these forms back by email (a summarized sample of the questionnaire with the issues here explained can be read in the appendix of this article). In brief, the topics discussed are related to the teacher's opinions about languages, the Error Analysis methodology and the design of the search interface. Evaluators were allowed to freely consult the tool, but they were not required to perform any task. This seemed to be easier for the teachers to evaluate the prototype, and even more flexible (since they made suggestions about certain points which were not previously taken into consideration). On the other hand, this lack of control could have caused some issues to be superficially checked (e.g. the correction of ambiguous errors), besides the fact that some questions were wrongly answered or left in blank.

4.2 Results

4.2.1. Evaluation participants' profile

Regarding the evaluators' knowledge or use of the new technologies, 8 out of 22 (36,4%) stated that they had an intermediate level; 11 out of 22 (50%), an advanced level, and 3 of them (13,6%) admitted to have a low level. Their profile is varied concerning their experience (some of them have just begun their professional practice, while five of them have taught between 14 and 25 years) and the group of learners they have (students with the same L1 or from different language backgrounds).

4.2.2. Beliefs about languages and the EA

Teachers' opinion about the influence of the learner's L1 (or other languages spoken or learned) on the learning/acquisition process is mainly neutral. They mostly stated that any known language has both a positive influence –it can help the acquisition process–

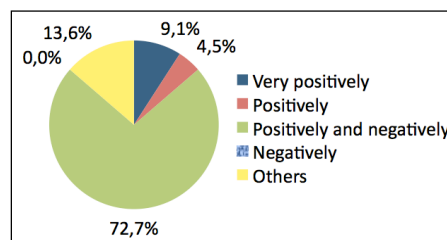and a negative influence –due to the interference of the L1 or other language (L3)– (Table 1, Figure 2).

Figure 2. Beliefs about the influence of other languages

	Frequency	
	Absolute	Relative (%)
Positively and negatively	16	72,7%
Others	3	13,6%
Very positively	2	9,1%
Positively	1	4,5%
Negatively	0	0,0%

Table 1: Beliefs about the influence of other languages

Regarding the Error Analysis approach, half of the evaluators admitted a limited degree of knowledge about the EA methodology, and less than the other half indicated a knowledge of just the basics. Many of them thought that its utility is restricted to research purposes only, and in second place, to improve teaching materials according to the learners' L1 (see Table 2 and Figure 3).
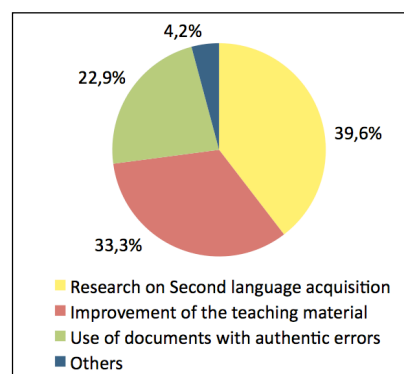

Figure 3: Opinions about the utility of the EA approach

	Frequency	
	Absolute	Relative (%)
Research on Second Language Acquisition	19	39,6%
Improvement of the teaching material	16	33,3%
Use of documents with authentic errors	11	22,9%
Others	2	4,2%

Table 2: Opinions about the utility of the EA approach[1]

---

[1] Note that, as evaluators could choose more than one option in this point, there are more than 22 answers (48 in total).

Despite that opinions about this methodology have mostly been positive (Table 3 and Figure 4), many evaluators considered that it has limitations, and disagreements have arisen about the use of documents or recordings with authentic errors; in fact, not few of the teachers made clear that the errors and its exceptions must always be explained.
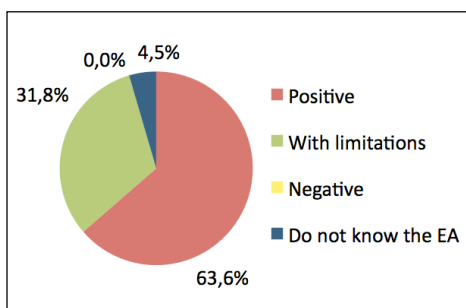

Figure 4: Opinions about the EA methodology

	Frequency	
	Absolute	Relative (%)
Positive	14	63,6%
With limitations	7	31,8%
Do not know the EA	1	4,5%
Negative	0	0%

Table 3: Opinions about the EA methodology

Other questions addressed the use of the EA specific terminology, and some recommendations were made to include clear examples of every type of error or an explanatory diagram of the error taxonomy (both of which have been included in the current version). As far as the ambiguous errors are concerned, although sometimes they were not viewed (because the search performed did not retrieved them), many evaluators reckoned that they were well treated.

The error analysis performed was mainly considered suitable and evaluators did not underline the need to improve it (Table 4 and Figure 5). Nonetheless, important observations were made about what should be considered an error in speech or in colloquial register, since even the native speakers' oral discourse shows certain deviations (e.g. gender disagreement: *la plaza *nuevo,* instead of *nueva,* 'the new square'). As well, due to the lower planning level in spontaneous spoken production, phrases or words are often placed in the wrong order. These phenomena should be considered *mistakes* among native speakers; but among non-native speakers, we are not always sure whether these deviations are performance (non-systematic) errors or competence errors. Finally, dialectal varieties should be taken into account so as to avoid hypercorrection; for instance, the pronunciation of /x/ with [h] sound should not be considered a mispronunciation, given that these phenomena are common in Spanish of South America or Southern Spain.
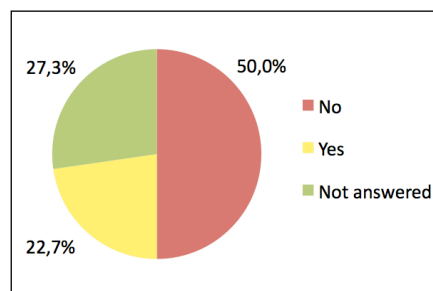

Figure 5: Need of improvement of the error analysis

	Frequency	
	Absolute	Relative (%)
No	11	50,0%
Not answered	6	27,3%
Yes	5	22,7%

Table 4: Need of improvement of the error analysis

Other remarks were made about the morphosyntactic categories or type of errors chosen for the analysis. Even though the majority of the opinions gathered did not observed the lack of any category (Table 5 and Figure 6), some specialists recommended us to further differentiate certain classes (e.g. between definite and indefinite article), or to add others from the Pragmatic or the Phonetic level (e.g. suprasegmental errors such as intonation).
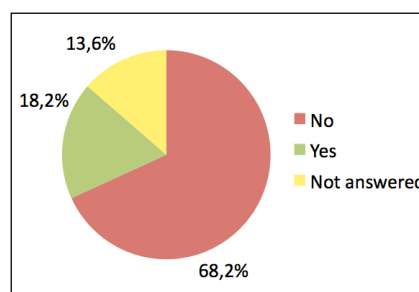

Figure 6: Lack of any category or type of error

	Frequency	
	Absolute	Relative (%)
No	15	68,2%
Yes	4	18,2%
Not answered	3	13,6%

Table 5: Lack of any category or type of error

4.2.3. Opinions about the search interface

Concerning the corpus search tool, the questionnaire addressed issues such as ease of data search and interface design. The presentation of the data (the information about the corpus, the learner's metadata, the explanation of the error or the error count) were considered "sufficient", yet some improvements were proposed with regard to the error indication, where the correction of the

error should always be shown. Furthermore, the error descriptors (the name for each error type, the morphosyntactic categories, the etiology or the linguistic level) have not always been clear or understandable enough for the evaluators (Table 6 and Figure 7). A recommendation to overcome this problem was to make a glossary with definitions for specific terminology, which has been added to the current version of the tool.
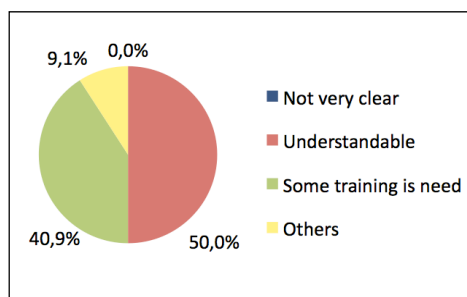
Figure 7: Opinion about the terms to describe the errors

	Frequency	
	Absolute	**Relative (%)**
Understandable	11	50,0%
Some training is needed	9	40,9%
Others	2	9,1%
Not very clear	0	0,0%

Table 6: Opinion about the terms to describe the errors

The error search criteria (for example, category or type of error) were mostly assessed as appropriate (Table 7 and Figure 8). Nevertheless, other suggestions were made to query the data in a more dynamic way: for instance, to use a checklist menu where several criteria can be selected, so that the page is simplified.
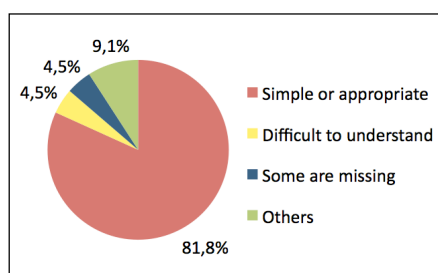
Figure 8: Opinions about the error search criteria

	Frequency	
	Absolute	**Relative (%)**
Simple or appropriate	18	81,8%
Others	2	9,1%
Difficult to understand	1	4,5%
Some are missing	1	4,5%

Table 7: Opinions about the error search criteria

The navigation scheme or the control of buttons and menus did not seem complicated (Table 8 and Figure 9).

Still, one teacher suggested including (in the introduction section) an explanation about the aims of the tool, the potential user and some usage recommendations.
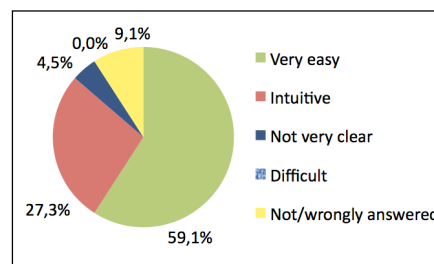
Figure 9: Opinion about the control of buttons or menus

	Frequency	
	Absolute	**Relative (%)**
Very easy	13	59,1%
Intuitive	6	27,3%
Not/wrongly answered	2	9,1%
Not very clear	1	4,5%
Difficult	0	0%

Table 8: Opinion about the control of buttons or menus

The menus with selectable lists of options for each criteria were mostly assessed as "Simple or appropriate" (Table 9 and Figure 10), in spite of certain negative comments, such as the long list of options or the need to look up a glossary to understand the terms.
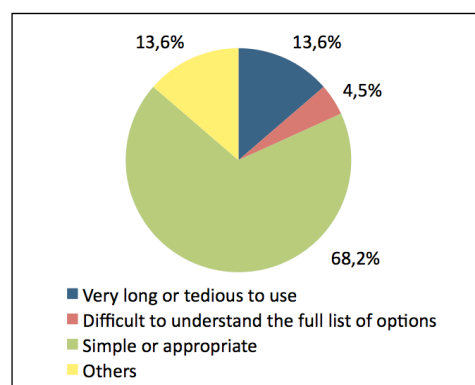
Figure 10: Opinions about the selectable list and menus

	Frequency	
	Absolute	**Relative (%)**
Simple or appropriate	15	68,2%
Very long or tedious to use	3	13,6%
Others	3	13,6%
Difficult to understand	1	4,5%

Table 9: Opinions about the selectable list and menus

The aesthetic design, though regarded as "sober" or even "boring" by few evaluators, was mainly judged as "appropriate" (Table 10 and Figure 11). Some recommendations were provided to improve the visual

distribution and the search results, which can be confusing when they are numerous.
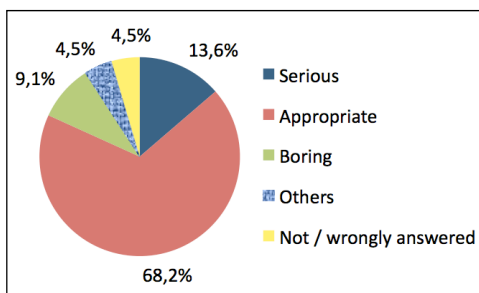

Figure 11: Opinion about the aesthetic design

	Frequency	
	Absolute	**Relative (%)**
Appropriate	15	68,2%
Serious	3	13,6%
Boring	2	9,1%
Others	1	4,5%
Not/wrongly answered	1	4,5%

Table 10: Opinion about the aesthetic design

Other points evaluated were the ease of listening to the interviews and the utility of the help texts. About the display of transcriptions and full recordings, more than half of those polled would not improve the visualization format (Table 11 and Figure 12).
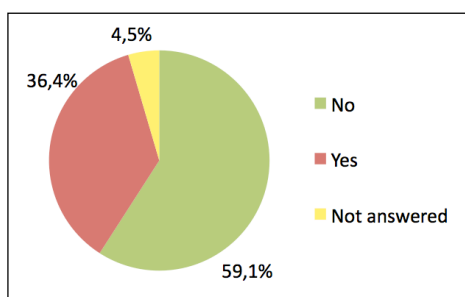

Figure 12: Need of improving the transcriptions display

	Frequency	
	Absolute	**Relative (%)**
No	13	59,1%
Yes	8	36,4%
Not answered	1	4,5%

Table 11: Need of improving the transcriptions display

Notwithstanding that, some of them would incorporate the following features: to use a livelier sound player to make it easier to access to certain points of the recording, to allow downloading of the file or to simplify transcription marks (an interesting alternative could be to offer the option to choose between a full or a simplified version).

5. Conclusions

The formative evaluation, despite being positive overall, has proved to be useful to unveil deficiencies, especially regarding those aspects more related to the pedagogical aims of the application. In fact, the following suggestions have been already included in the current version:

- Phonetic categories to describe pronunciation errors more accurately (segments and suprasegmentals).
- Brief diagram of the error taxonomy (with examples).
- Explained glossary of terms (with examples).
- Error frequency count (expressed as a percentage) per group of students with the same L1.

Following are other salient recommendations received:

- To include a further explanation of the aims of the tool and the type of user.
- The production of a teaching guide to be used along with the computer interface.
- To provide the option of displaying the simplified transcriptions (without marks for the speech phenomena) as well as the full transcriptions.
- A more dynamic menu design to simplify the page where the search criteria are selected.
- An improved display of the search results, by limiting the number of sound fragments retrieved per page.
- A sound player for the full interviews which make it easier to access to specific parts of the recording.
- To display the correction next to the indication of the error and to simplify the way to show the explanation.

Even though it is difficult to meet every user's needs –particularly those expressed by few specialist linguists interested in specific branches of Linguistics–, the implementation of some of the recommendations will benefit the final users (i.e. teachers of Spanish) to make use of this tool.

6. Acknowledgements

7. References

Campillos Llanos, L. (2011) "A XML-tagged Spanish learner oral corpus for learner language research". Poster presented in Learner Corpus Research Conference. 15-17 Sept 2011. Louvain-la-Neuve, Belgium. www.lllf.uam.es/ESP/Publicaciones/XMLLearner.pdf

Colpaert, J. (2004) *Design of online interactive language*

courseware: conceptualization, specification and prototyping. PhD Thesis. Universiteit Antwerpen.

Corder, P. (1971) Idiosyncratic Dialects and Error Analysis. *International Review of Applied Linguistics,* 9(2): 147-60.

Council of Europe. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge University Press.

Cresti, E. & Moneglia, M. (2005) *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages.* Studies in Corpus Linguistics, 13. Amsterdam/Philadelphia: John Benjamins.

Díaz Negrillo, A. (2009) *EARS: A user's manual.* LINCOM Academic Reference Books 01.

Díaz Rodríguez, L. (2007) *Interlengua española: estudio de casos.* Barcelona: Printulibro Intergrup.

Fernández López, S. (1997) *Interlengua y Análisis de Errores en el aprendizaje del español como lengua extranjera.* Madrid: Edelsa.

Gilquin, G., De Cock, S. & Granger, S. (2010) *Louvain International Database of Spoken English Interlanguage.* Presses universitaires de Louvain.

Granger, S. (2003) Error-tagged Learner Corpora and CALL: a promising synergy. *CALICO Journal,* 20 (3), pp. 465-480.

Granger, S., Kraifa, O., Pontona, C., Antoniadisa, G. & V. Zampa (2007) Integrating learner corpora and natural language processing. *ReCALL,* 19, pp. 252-268.

James, C. (1998) *Errors in Language Learning and Use. Exploring Error Analysi*s. London/NY: Longman.

Koller, T. (2007): *Design, Development, Implementation and Evaluation of a Plurilingual ICALL System for Romance Languages Aimed at Advanced Learners.* PhD Thesis. Dublin: Dublin City University.

Lennon, P. (1991) Error: Some Problems of Definition, Identification, and Distinction. *Applied Linguistics,* 12 (2): 180-196.

Lüdeling, A., Walter, M., Kroymann, E. & Adolphs, P. (2005) Multi-level error annotation in learner corpora. *Proceedings of the Corpus Linguistics Conference 2005.* Birmingham: University of Birmingham.

Mitchell, R., Dominguez, L., Arche, M. J., Myles, F. & Marsden, E. (2008) SPLLOC: A new database for Spanish second language acquisition research. *EuroSLA Yearbook,* 8, 287-304.

Myles, F. (2005) Interlanguage corpora and Second Language Research. *Second Language Research*, vol. 21(4): 373-391

Nicholls, D. (2003) The Cambridge Learner Corpus – error coding and analysis for Lexicography and ELT. In Archer *et al.* (eds.) *Proceedings of the Corpus Linguistics 2003 Conference (CL2003).* Lancaster University: UCREL, pp. 572-581.

O'Donnell, M. (2008) The UAM CorpusTool: Software for corpus annotation and exploration. *Proceedings of the XXVI AESLA Conference, Almería, España, 3-5 Abril 2008.*

Vázquez, G. (1999) *¿Errores?¡Sin falta!* Madrid: Edelsa.

Ward, M. (2006) "Using Software Design Methods in CALL". *Computer Assisted Language Learning*, Vol. 19, No. 2-3, pp.129-147.

Appendix: Sample questionnaire for the evaluation (summarized)

1. Teacher's profile

-Languages spoken:

-Teaching experience:

-I have principally worked with students who speak the following languages: (*underline or mark*)

Arab	English	Greek	Japanese	Portuguese	Swedish
Chinese	French	Hindi-Urdu	Korean	Rumanian	Cooficial languages (*indicate*):
Czech	German	Italian	Polish	Russian	Others:

-Other comments:

2. Beliefs about languages and opinions about the Error Analysis methodology

-The influence of the mother tongue or other languages known influences the acquisition of a foreign language…

Very positively   Positively   Positive and negatively   Negatively   Very negatively

-My opinion about the Error Analysis (EA) methodology is…

Positive   With limitations   Negative   I do not know the EA.

-I consider that the utility of Error Analysis (EA) is… (*you can choose more than one*)

 Research in Spanish as a Second/Foreign Language Acquisition.

 Improvement of the teaching material according to the learners' L1.

 The use of documents or teaching material with authentic errors in class.

 Others (*indicate*):

-The terms used in the search tool for each type of error, the etiology or the linguistic level where it appears are…

  Not very clear.   Understandable.

  Some training is needed.   Others (*indicate*):

-Is there any lacking category or a type of error which has not been considered? No. Yes.
 If so, which one? (*indicate*)

-Would you improve the error analysis performed so far?   No. Yes.
 If so, in which aspects? (*indicate*)

3. Opinions about the search interface

-Regarding the use of the new technologies, you consider yourself a user with this level:

   Very low   Basic   Intermediate   Advanced

-Becoming familiar with the control buttons, the navigation scheme or the interface menu is:

   Very easy   Intuitive   Not very clear   Difficult

-The aesthetic design of the interface is:

   Serious   Appropriate   Boring   Others (*indicate*):

 Would you improve anything of the aesthetic design? (*optional answer*)

-The error search criteria (e.g. category, L1, type of error, linguistic level, etc.) are:

  Simple or appropriate.   Some are missing (*indicate*):

  Difficult to understand.   Others (*indicate*):

-Selectable lists and menus to consult the errors (e.g. the options for "type of error": conjugation, *ser/estar*...) are:

  Very long or tedious to use.   Simple or appropriate.

  Difficult to understand.   Others (*indicate*):

-Would you improve the display of the transcriptions or the full recordings? No. Yes.
 If so, in which aspects? (*indicate*)

-What other functionality or features would you like to include in the search tool?

-Free comments: