



V International Conference on Corpus Linguistics (CILC2013)

SCIL: a Spanish Corpus of Italian Learners

Sonia Bailini*

Catholic University of Milan, Largo Gemelli 1, 20123 Milan, Italy

Abstract

In the last 15 years research into the acquisition of Spanish as a Foreign/Second language has seen a growing interest in the building of learner corpora but, in most cases, they collect Spanish interlanguage of English-speaking learners. SCIL is a longitudinal Spanish Corpus of Italian Learners and consists of 457 written compositions (124,186 tokens) by 43 informants, whose proficiency levels range from A1 to B2 of the CEFRL. SCIL has been collected over seven months and stored to be used with *WordSmith Tools*. It allows both a longitudinal analysis, to observe the interlanguage evolution of a specific student or group, and a cross-sectional study, to observe a specific grammar or discourse feature in texts about the same topic written by different students.

© 2013 The Authors. Published by Elsevier Ltd.
Selection and peer-review under responsibility of CILC2013.

Keywords: learner corpus; interlanguage; Spanish; Italian; corpus linguistics

1. Introduction: why a Spanish Corpus of Italian learners?

In the last 15 years research into the acquisition of Spanish as a Foreign/Second language has seen a growing interest in the creation of learner corpora, i.e. collections of Spanish oral or written interlanguage of learners with different mother tongues. Ainciburu's (2009) state-of-the-art article about Spanish learner corpora, Tono (2010) and the *Learner Corpora around the world* website (2012) by the Louvain University confirm that, in most cases, Spanish learner corpora collect oral or written texts produced by English-speaking learners. Examples of these corpora are *SPLLOC: a new database for Spanish second language acquisition research* (R. Mitchell et al. 2008, University of Southampton); the *Spanish Learner Language Oral Corpus Project (SPLLOC 1): A new Corpus of Oral L2 Spanish* (N. Tracy Ventura, 2008) at <http://www.splloc.soton.ac.uk>; the *Spanish Corpus Proficiency Level*

* Corresponding author. Tel.: +39 2 72342920; fax: +39 2 72343667
E-mail address: sonia.bailini@unicatt.it

Training at <http://www.laits.utexas.edu/spt/> (D. Koike, University of Texas at Austin) and the *Proyecto CEDEL2 – Corpus Escrito de Español L2* (2007) by the *Grupo de Investigación Woslac* (Word Order in Second Language Acquisition Corpora) of the Universidad Autónoma de Madrid at <http://www.uam.es/proyectosinv/woslac/>.

Spanish learner corpora of students with native languages other than English are quite meager. Among them are the *Anglia Polytechnic University Learner Spanish Corpus* (A. Ife, 2004), which collects written texts (120,000 words) of learners with different mother tongues; the *DIAZ Corpus* (L. Díaz Rodríguez, 2007, Universidad Pompeu Fabra), which contains interviews and questionnaires given to learners whose mother tongues are German, Swedish, Icelandic, Korean and Chinese; the *Corpus de Aprendices Taiwaneses de Español* (CATE) (Lu Hui-Chuan, 2007) and the *Japanese Learner Corpus of Spanish* by Yoshihito Kamakura (2005).

Finally, the following learner corpora collect texts in different languages, including Spanish, produced by learners with the same mother-tongue: the *USP Multilingual Learner Corpus* (MLC) (S. Tagnin, 2002), which collects written compositions in English, Spanish and German by Brazilian students of the University of São Paulo and the *MiLC Corpus*, which contains letters, CVs, essays, reports, translations, chats and e-mails in English, German, Italian and Spanish produced by university students whose mother-tongue is Catalan (A. Andreu Andrés et al., 2010, Universidad Politécnica de Valencia).

At present there are only two Spanish corpora of Italian learners: the *Corpus para el análisis de errores de aprendices de E/LE* (CORANE) by Ana M. Cestero Mancera, I. Penadés Martínez et al. (2009) and the *Corpus de textos escritos por universitarios italianos estudiantes de ELE* by Esther Gutiérrez Quintana (2005). CORANE contains 1091 written compositions by 321 SFL students of Spanish Courses for Foreigners at the Universidad de Alcalá de Henares. There are 18 Italian informants in this corpus and they have provided 29 texts. The *Corpus de textos escritos por universitarios italianos estudiantes de ELE* consists of 44 written compositions -on one out of three topics proposed by the teacher- by 44 Italian students of the Università di Sassari.

The poor representativeness of Spanish interlanguage of Italian learners in the above mentioned corpora justifies the need for a larger corpus which could be a useful tool for both interlanguage analysis of cognate languages learners and the development of teaching materials that can meet the specific learning needs of this type of learner. The *Spanish Corpus of Italian Learners* (SCIL) consists of 457 written compositions (124,186 tokens) by 43 Italian university students of Spanish as a foreign language, whose proficiency levels range from A1 to B2 of the Common European Framework of Reference for Languages (CEFRL). SCIL is a longitudinal corpus that has been collected over seven months (from December 2008 to July 2009) and has been stored to be used with *WordSmith Tools*. SCIL allows both a longitudinal analysis, to observe the interlanguage evolution of a specific student or group, and a cross-sectional study, to observe a specific grammar or discourse feature in texts about the same topic written by different students.

Another reason for building the *Spanish Corpus of Italian Learners* is the great interest that Italian schools and universities have shown over the last ten years in learning Spanish, so that it is now the second most studied foreign language after English, especially at university, as Alonso Patiño (2007) explains in his detailed report on the Spanish learning situation in Italy in the period 2004-2007. For instance, in Lower Secondary School Spanish students went from 6.5% (64,538) in 2004-2005 to 14% (235,810) in 2006-2007, while in Upper Secondary School, in 2006-2007, there were 10,000 more Spanish students than in 2005-2006. There is no data available from the Italian Ministry of Education about university foreign languages students. To this data we must add the data of informal education, such as the Spanish courses at the *Instituto Cervantes* in Italy (Milan, Rome and Naples) and the growing number of Italians attending Spanish courses in Spain or in the Spanish-speaking countries of America.

The lack of extensive learner corpora of Italian learners' Spanish interlanguage and the increased interest in learning Spanish in Italy are good reasons for a more systematic study of the difficulties, potential and limits language affinity implies using real databases as research tools. In this view SCIL broadens the horizons of research into Spanish as a Foreign/Second language acquisition, in line with what has already been done for English learner corpora, since, as Baralo has recently pointed out, "there is still a methodological need to develop longitudinal data bases for Spanish as a Foreign Language in different learning contexts" (Baralo, 2010). We believe SCIL is a first step in this direction.

2. The Spanish Corpus of Italian Learners (SCIL): design criteria

One of the main aspects to be considered when engaging in the building of a learner corpus is to set criteria that will allow data analysis from different points of view. Tono (2002) identifies three major criteria:

- *Language-related criteria*: this category distinguishes between written or spoken corpora, text genre (letter, diary, essay etc.), style (narration, argumentation etc.) and topic.
- *Task-related criteria*: this category includes the methods used for collecting data (cross-sectional versus longitudinal), the forms through which data are elicited (spontaneous versus prepared texts), the possibility of using reference materials such as dictionaries or manuals and the existence or absence of time constraints.
- *Learner-related criteria*: refers to age, cognitive style, motivation and attitude towards the foreign language, the learning context, the proficiency, L1 influence and knowledge of other languages.

Table 1 summarizes SCIL features according to the above mentioned criteria:

Language-related criteria	Task-related criteria	Learner-related criteria
<i>Mode</i> : written, by email in e-tandem mode	<i>Data collection</i> : longitudinal, for 7 months	<i>Age</i> : mean 19,9 – median 20 <i>Gender</i> : 39 F, 4 M: total 43 <i>Cognitive styles</i> : 49% reflective 28% analytical 16% global 7% impulsive
<i>Genre</i> : informal letters	<i>Data elicitation</i> : written compositions based on a topic proposed by the teacher	<i>Motivation and attitude</i> : voluntary participation, positive attitude towards Spanish
<i>Style</i> : narration, exposition description, instruction, argumentation	<i>Use of references</i> : yes; dictionaries, grammars, manuals, etc.	<i>Learning context</i> : Spanish as a Foreign Language in a university context
<i>Topic</i> : general related to the learner's personal knowledge: geography, music, cuisine, folklore, cinema, traveling ...	<i>Time constraints</i> : none	<i>Proficiency</i> : from A1 to B2 of the CEFRL <i>Other languages known</i> : English (100%), French (25.5%), German (25.5%), Russian (7%). 56% already studied Spanish at secondary school and 44% started at university.

The informants were first and second year university students at the Faculty of Foreign Languages and Literatures at Università Cattolica del Sacro Cuore. Spanish, as well as English, was one of their specialization languages, and their curricula specialized in the use of languages for business and education. The learners dominant cognitive styles were identified through Mariani's (2000) questionnaire available at <http://www.learningpaths.org/Questionari/sxdxnuovo.htm>. Motivation and attitudes towards Spanish were checked as follows: a) first the e-tandem learning method was briefly presented; b) a survey was submitted to find out whether the students would be willing to participate in the e-tandem project; c) the students were asked about their expectations from a linguistic and cultural viewpoint and about the amount of time they would spend on the project. Eventually, students answered a short set of questions about the perceived linguistic distance between Spanish and Italian and their level of empathy towards the cultures of Spanish-speaking countries.

3. The Spanish Corpus of Italian Learners (SCIL): general features

SCIL is a longitudinal corpus consisting of 457 written compositions (124,186 tokens) by 43 Italian university students at four proficiency levels of the CEFRL (from A1 to B2):

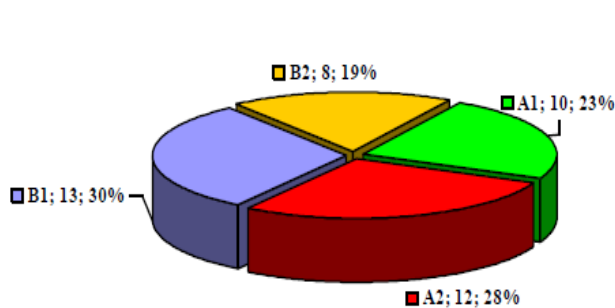


Figure 1. SCIL: learners distribution

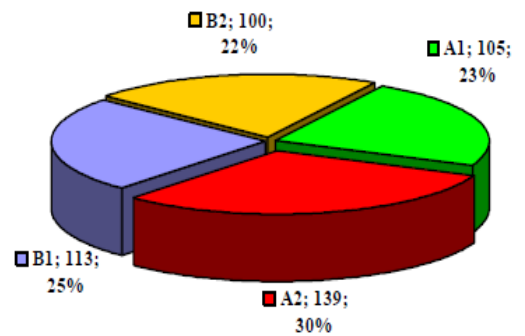


Figure 2. SCIL: texts distribution

The texts were collected over seven months (from December 2008 to July 2009) while monitoring interaction in e-tandem between Spanish students learning Italian and Italian students learning Spanish. SCIL only contains the texts written in Spanish by Italian learners (Figure 3); at a second stage of the research the texts written in Italian by Spanish learners will also be included in order to use interlanguage corpora of cognate languages to identify problematic learning areas. It is worth pointing out that SCIL only contains written compositions on a topic proposed by the teacher or chosen by the participants; therefore, it does not include translations, exercises, summaries or other texts that cannot be considered spontaneous samples of interlanguage. This variable reduces the ratio between the number of texts per student and the period of data collection, but does not affect the observation of interlanguage evolution. Moreover, it is worth mentioning that some participants, for various reasons, did not maintain the textual exchange with their partner during the entire period of data collection and this explains the low production of texts of some of them. However, if the productions of these students may be less relevant to the longitudinal analysis, they can be useful for the cross-sectional study.

Considering what has been explained above, Figure 4 shows the average texts per student:

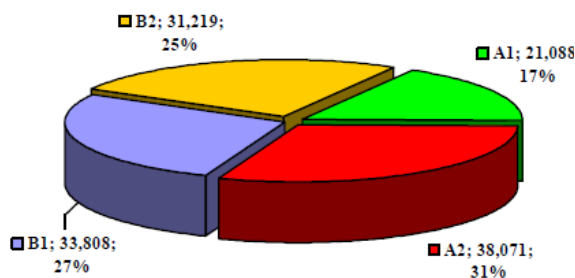


Figure 3. SCIL: number of tokens at each proficiency level

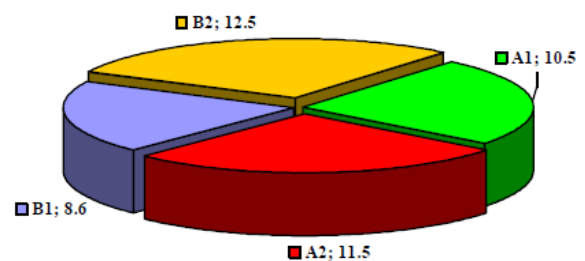


Figure 4. Average N° of texts per student

Since our aim is to provide a corpus that can be used for different kinds of analyses, the condition for a text to be included in the corpus was set at a minimum of 5 texts produced by the same informant. Table 2 shows the distribution of texts per student:

Table 2. SCIL: Text distribution per student

N° texts	5	6	7	8	9	10	11	12	13	14	15	16	17	22	25
N° informants	2	4	3	7	5	3	5	4	2	1	1	2	2	1	1

If we observe this data we note that only 2 students wrote as few as 5 texts and only 2 of them wrote more than 20. Apart from these 4 students, 49% of the remaining 39 informants wrote less than 10 texts and 51% of them more than 10. The longitudinal distribution of the texts per student in each level is illustrated in Table 3:


Table 3: SCIL: N° of texts per student in each level

A1	N° students (10)	1	1	3	1	1	1	2		
	N° texts (105)	6	7	8	11	12	13	16		
A2	N° students (12)	1	4	1	3	1	1	1		
	N° texts (139)	5	9	10	11	13	17	25		
B1	N° students (13)	1	3	1	3	1	1	1	1	1
	N° texts (113)	5	6	7	8	9	10	11	12	17
B2	N° students (8)	1	1	1	2	1	1	1		
	N° texts (100)	7	8	10	12	14	15	22		

Table 3 also shows that most students at A2 and B2 proficiency levels wrote more than 10 texts and that B1 proficiency learners present the lower ratio of number of texts per student.

The topics range from descriptions of people, environments, objects, usages and customs to narration of experiences (travels, childhood memories, everyday life events etc.) and expression of likes and dislikes and from opinions and comments on recent events to instructional texts (recipes, directions on how to do something or how to go somewhere etc.). Finally, it is important to emphasize that being texts addressed to a real partner, they maintain the freshness and authenticity of spontaneous communication, which is highly relevant to a pragmatic analysis of learner speech. Figure 5 shows a sample of instructions for a composition task:

Mi casa es tu casa



Para profundizar tu conocimiento del español este verano has decidido ir un mes a España. Tu compañero/a Tándem te ha ofrecido su casa, ya que él/ella piensa venir a Italia para hacer lo mismo. Entonces tú le has ofrecido la tuya, ya que sabes que estará libre en el mes de agosto. Este intercambio os permitirá ahorrar mucho dinero. Ahora, lo primero que tienes que hacer es describirle tu casa: donde está, cómo es, cuántas habitaciones tiene, cuál es la parte que más te gusta de ella, si hay algo que no te gusta.... Si quieres, puedes agregar una foto o un dibujo.

.....

Déjale una nota a tu compañero/a en la que le das instrucciones sobre cómo funcionan los electrodomésticos, si se pueden usar o no determinadas cosas, lo que se puede o no se puede hacer.... Acuérdate también de aclarar qué cosas te molesta que se hagan en tu casa y si hay algo que te gustaría que él/ella hiciera mientras esté allí.

Querido/a

Figure 5: Sample of instructions for a composition task

SCIL has been stored for use with *WordSmith Tools* and it has to be seen as a first attempt to collect and systematize a corpus of Spanish interlanguage of Italian learners which makes it possible to analyse data both longitudinally (observation of the evolution of the interlanguage of a student or group over a period of time), and

cross-sectionally (observing the acquisition of a structure in different students at a certain step of their interlanguage).

4. The Spanish Corpus of Italian Learners (SCIL): possible exploitations

A corpus-based interlanguage analysis makes it possible to clearly identify specific linguistic features and to give a comprehensive description of the interlanguage system from different viewpoints. This section provides an example of longitudinal and cross-sectional analysis of the Spanish structure *hay/estar*. Spanish distinguishes between *'hay'* to indicate the existence of something (equivalent to English *there is/there are*) and *'estar'* to indicate its location. This construction is quite problematic for Italian learners because the Italian verb *'stare'* (i.e. *to stay*) does not have the same usage as its Spanish cognate *'estar'* nor is it as common; moreover Spanish *'estar'* performs functions for which Italian requires either *'c'è/ci sono'* (i.e. *there is/there are*) or the verb *'essere'* (i.e. *to be*). Nevertheless, in some cases there is a total correspondence between Spanish and Italian; this means that in such contexts positive transfer plays an important role. A corpus-based interlanguage analysis makes it possible to observe and assess the full extent of positive and negative transfer and understand the contexts where the Spanish grammar rule is correctly applied, offering a more reliable research method for qualitative analysis of specific acquisition issues.

As an example of a corpus-based interlanguage analysis let us observe how the construction *'hay/estar'* is used by A2 learners in the SCIL Corpus; indeed, it is at this proficiency level that this construction is usually taught. For a longitudinal analysis, the first and the last three texts produced by the same students have been included in order to assess the students' progress with this structure from the beginning to the end of the observation period (seven months). Only 11 informants out of 12 have been analyzed because one of them only produced as few as 5 texts so that her production was too low for a longitudinal analysis, which has therefore been carried out on two sets of 33 texts each. Obviously, concord research with *WordSmith Tools* has included all the inflected forms of the verbs involved in the construction analysed, that is *'hay/había/hubo/habrá'* and all their possible incorrect forms -in this context- *'*habían/hubieron/habrán'*. The same research criteria has been used with *'está/estaba/estuvo/estará'* and their respective plural forms *'están/estaban/estuvieron/estarán'*, even if their usage is incorrect in this construction. Moreover, in order to include all the possible variants of this structure, verb forms that normally require a word stress were also searched without it. The first 33 texts contain 37 utterances with *'hay/estar'*: 21 of them (57%) indicate a correct use of the construction due to positive transfer, 13 (35%) show a correct use due to rule application, 2 utterances (5%) show the effect of negative transfer and 1 (3%) of intralingual error. Concord research with *WordSmith Tools* makes it possible to see in which contexts the construction is used: positive transfer shows with *'hay + indefinite article'*, *'hay + numbers and indefinite quantifiers'* (*mucho/s, poco/s, algún/algunos*, etc.) and *'hay + generalizations'* (ex: *'no hay edificios'*). We can also notice that students are generally quite careful with the grammar rule regarding *'estar'* to locate spaces and things (*'Mi pueblo está cerca de Mantova'*; *'Brescia está en Lombardia'*) and that negative transfer usually shows with *'hay + definite article'* (indeed, this construction is correct in Italian, but not in Spanish; ex: *'*en la primera planta hay la cocina'* (Sp.) → *'al primo piano c'è la cucina'* (It.)). The intralingual error is due to the use of *'estar + indefinite quantifier'* instead of *'hay'*: *'*están muchos artistas'* instead of *'hay muchos artistas'*. In the last 33 texts, 70 utterances containing the construction *'hay/estar'* have been found: 42 of them (60%) show that positive transfer is working effectively, while the halved percentage of sentences with the correct grammar rule (17%, 12 utterances) clearly shows that the rule is broken by negative transfer. In these contexts, negative transfer amounts to 20%, with a total of 14 occurrences, all of which are due to the use of *'hay + determined article'*; and finally, the percentage of intralingual errors remains very low (3%, 2 occurrences with *'estar + indefinite article'*). To conclude, positive transfer plays an important role to speed up the acquisition of this construction and negative transfer shows its influence slowly as it grows persistently along time, becoming more visible just when learners perceive it as an acquired and well settled rule and divert their attention away from it.

A cross-sectional interlanguage analysis of the acquisition of the same structure (*'hay/estar'*) at a certain proficiency level takes into account all the texts produced by the A2 learners at a certain step in their learning period. In this case all the texts (139) of the A2 learners have been analysed at the end of the observation period

with the aim to assess the students' progress with this construction and the incidence and nature of errors at this proficiency level. On a total of 207 utterances with the construction 'hay/estar' the data shows 138 instances of correct use (67%) due to positive transfer, 30 instances of correct rule application (14%), 37 instances of error (18%) due to negative transfer and 2 intralingual errors (1%): the cross-sectional interlanguage analysis confirms the beneficial role of positive transfer and the hindering effect of negative transfer, which thus prevails over grammar rules. In other words, L1 influence is very strong and heavily affects progress in a cognate language. These findings are not new nor surprising, but it is worth remarking that they arise from a more rigorous and reliable research method, which makes use of large data: this is what makes our results generalizable and useful for a more systematic research on cognate languages interlanguage.

5. Conclusions

Nesselhauf (2004) remarks both the potential and the limits of learner corpora: among the former is the possibility of studying the frequency of certain errors, observing pragmatic use of language and identifying the influence of L1 or other foreign languages; as regards the latter Nesselhauf points out that learner corpora only let you see what the learner *uses*, which does not necessarily correspond to what he/she *knows*. Nesselhauf also underlines that learner corpora do not enable to observe the learner perception of acceptability regarding his/her productions, nor do they make it possible to analyse the role of the interaction or the effectiveness of a learning method. Nevertheless, learner corpora are a useful tool because they are authentic and extensive databases, opening the way for more comprehensive and reliable qualitative analysis. SCIL fits into corpus linguistics research applied to the acquisition of Spanish as a Foreign/Second language, a line of research that has yielded significant results with learners with different mother tongues, but that, as far as Spanish and Italian are concerned, still has a long way to go. Hopefully, SCIL will open a new path in a still unexplored area.

6. References

- Ainciburu, C. (2009). Al día *Revista Nebrija de Lingüística Aplicada*, 5
http://www.nebrija.com/revista-linguistica/revista_5/aldia_1.html [03/05/2013].
- Andreu, M. et al. (2010). Analysing EFL learner output in the MiLC project: An error it's*, but which tag? In M. C. Campoy, B. Belles-Fortuno & M. L. Gea-Valor (Eds.), *Corpus-Based Approaches to English Language Teaching* (pp. 167-179). London: Continuum.
- Baralo Ottonello, M. (2010). La investigación en español como lengua segunda: necesidad de un corpus de español no nativo, Congreso Internacional de la lengua española, Valparaíso, Chile
http://congresosdelalengua.es/valparaiso/ponencias/lengua_educacion/baralo_marta.htm [22/04/2013].
- Cestero Mancera, A., Penadés Martínez, I., et al. (2009). *Corpus de textos escritos para el análisis de errores de aprendices de E/LE (CORANE)*. Alcalá: Universidad de Alcalá.
- Díaz Rodríguez, L. (2007). *Interlengua española. Estudio de casos* (with a corpus in attached CD). Barcelona: Printulibro.
- Granger, S. (2004). Computer learner corpus research: current status and future prospects. In U. Connor & T. Upton (Eds.), *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 123-145). Amsterdam & Atlanta: Rodopi.
- Granger, S. (2007). A Bird's-eye View of Computer Learner Corpus Research. In W. Teubert & R. Krishnamurthy (Eds.), *Corpus Linguistics: Critical Concepts in Linguistics* (pp. 44-72). London & New York: Routledge.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 13-32). Amsterdam & Philadelphia: John Benjamins.
- Grupo de Investigación Woslac. (2007). *Proyecto CEDEL2 – Corpus Escrito de Español L2* <http://www.uam.es/proyectosinv/woslac/> [3/5/2013]
- Gutiérrez Quintana, E. (2005). Corpus de textos escritos por universitarios italianos estudiantes de ELE. *Linred Lingüística en la red*, 1, 11
http://www.linred.es/informacion_pdf/informacion9_01112005.pdf [22/04/2013].
- Hui-Chuan, L., Fang-I, P. (2007). Construcción de corpus de aprendices taiwaneses de español (CATE), XXV International Aesla Conference. Murcia: Spain.
- Ife, A. (2004). The L2 learner corpus: reviewing its potential for the early stages of learning. In M. Baynham, A. Deignan & G. White (Eds.), *Applied Linguistics at the Interface: British Studies in Applied Linguistics* (pp. 91-104). London: Equinox.
- Kamakura, Y. (2005). Japanese Learner Corpus of Spanish – L2 and L2 interlanguage contrastive analysis, 4th International Contrastive Linguistics Conference. Santiago de Compostela: Spain.
- Koike, D. *The Spanish Corpus Proficiency Level Training*, <http://www.laits.utexas.edu/spt/> [3/5/2013]
- Learner Corpora around the world*, <https://www.uclouvain.be/en-cecl-leworld.html>, last updating: December 2012 [03/05/2013].

- Lozano, C. (2009). CEDEL2: Corpus Escrito del Español como L2. In C. M. Bretones et al. (Eds.), *Applied Linguistics Now: Understanding Language and Mind/La Lingüística Aplicada actual: Comprendiendo el Lenguaje y la Mente* (pp. 197-212). Almería: Universidad de Almería.
- Lozano, C., Mendikoetxea, A. (2013). Learner corpora and Second Language Acquisition: The design and collection of CEDEL2. In N. Ballier, A. Díaz-Negrillo & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins <http://wdb.ugr.es/~cristobalozano/wp-content/uploads/Lozano-and-Mendikoetxea-2012-CEDEL2-versi%C3%B3n-p%C3%A1gina-web.pdf> [23/04/2013].
- Mariani, L. (2000). *Portfolio*. Bologna: Zanichelli, <http://www.learningpaths.org/Questionari/sxdxnuovo.htm> [21/05/2013]
- Mitchell, R., Domínguez, L., Arche, M., Myles, F., & Marsden, E. (2008). SPLLOC: A new database for Spanish second language acquisition research. In L. Roberts, F. Myles & A. David (Eds.), *EUROSLA Yearbook* 8 (pp. 287-304). Amsterdam: John Benjamins.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. McH. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 125-152). Amsterdam & Philadelphia: John Benjamins.
- Patiño Alonso, M. (2007). Situación del español en Italia 2006-07. *Cuadernos de Italia y Grecia*, 6, 9-32. https://sede.educacion.gob.es/publivena/descargas.action?f_codigo=13412&codigoOpcion=3 [28/2/2012]
- Tagnin, S. (2002). *A Multilingual Learner Corpus (MLC) in Brazil* <http://www.fflch.usp.br/dlm/comet/artigos/A%20multilinguagem%20de%20aprendizagem%20de%20segunda%20lingua%20em%20brasil.pdf> [03/05/2013]
- Tono, Y. (2003). Learner corpora: Design, development and applications. In D. Archer, P. Rayson, A. Wilson, & T. McEnery (Eds.), *Proceedings of the 2003 Corpus Linguistics Conference UCREL*. Lancaster University: United Kingdom, UCREL Technical Paper #16
- Tono, Y. (2010). *Corpora, Collections, Data, Archives*, <http://www.uow.edu.au/~dlee/corpora.htm>, last updating: April 2010 [25/02/2013].
- Tracy Ventura, N. (2008). *The Spanish Learner Language Oral Corpus Project (SPLLOC 1): A new Corpus of Oral L2 Spanish* <http://www.splloc.soton.ac.uk> [3/5/2013]