

Corpora for Linguists vs. Corpora for Learners Bridging the Gap in Italian L2 Learning and Teaching

Luciana Forti

Università per Stranieri di Perugia, Italia

Stefania Spina

Università per Stranieri di Perugia, Italia

Abstract This paper aims to shed light on how research findings stemming from Learner Corpus Research (LCR) can inform the development of Data-driven learning (DDL) pedagogical activities. By doing this, it seeks to show how the gap between corpora built to be used by linguists and those tailored for learners can be filled. It starts by defining what a corpus is and how second language learning studies can benefit from the research findings based on corpora, but also from the direct use of corpora in the classroom. Then, it provides an overview of the available native and learner corpora of Italian, and how corpora in general can be adapted for DDL purposes. Finally, it describes an example of how an LCR finding can be used to develop DDL activities. It concludes with some desiderata for the future.

Keywords DDL. LCR. Italian L2. Collocations. Pedagogical materials design.

Summary 1 Introduction. – 2 Corpora for All? – 2.1 An Overview of Italian Native and Learner Corpora. – 2.2 Towards Learner-Friendly Corpora. – 3 From LCR to DDL. – 3.1 An LCR Study on Italian. – 3.2 Building an LCR-Informed DDL Activity. – 4 Conclusions.



Peer review

Submitted	2019-06-09
Accepted	2019-11-30
Published	2020-04-22

Open access

© 2020 | Creative Commons Attribution 4.0 International Public License



Citation Forti, Luciana; Spina, Stefania (2019). "Corpora for Linguists vs. Corpora for Learners: Bridging the Gap in Italian L2 Learning and Teaching". *EL.LE*, 8(2), 349-362.

DOI 10.30687/ELLE/2280-6792/2019/02/005

1 Introduction

A corpus is widely defined as an authentic, representative and machine-readable collection of linguistic data (McEnery, Xiao, Tono 2006, 5). It is authentic because it includes only instances of language that are used by real speakers, as opposed to a single ideal speaker; it is representative because it aims to include a sample of language data that is appropriate to a given aim of inquiry, thus being able to reflect a certain genre or variety; finally, it is machine-readable in the sense that the data is stored in electronic format and is searchable through a dedicated software interface, so that large quantities of data can be processed at the one time.

L2 learning studies can benefit from corpora in a number of ways. The most notable advantage of using corpora consists in being able to have better descriptions of real language usage by native speakers, which means that teachers and, even more so, syllabus designers can rely on better sources to identify and sequence L2 learning aims. This potential of corpora has been clear since the very early stages of corpus construction, when Randolph Quirk stated that corpora were a necessity in order to deal with the inadequacy of teaching materials at the time, which kept reflecting a kind of language that learners would not find in real life communicative practices (Quirk 1960).

Another major advantage of using corpora in L2 learning studies is that corpora are able to provide empirical evidence of real language use not only in relation to native speakers, but also in relation to the learners themselves. This is the case of learner corpora, where teachers and researchers are able to detect the most frequently occurring errors, along with the most frequently occurring traits distinguishing native and non-native uses of the language (Granger 1996; 2015). Learner corpora that are constructed on the basis of more than one data collection point in time are also able to trace developmental patterns in the learning process through time, which are argued to be typically non-linear (Larsen-Freeman 1997).

In general, two main modalities in using corpora for second language learning have been identified: the indirect use, when the data derived from the corpus is not immediately visible to the learners, and the direct use, when data is immediately visible to the learners. This dichotomy was originally proposed by Leech (1997) in relation to the direct vs. indirect aspect, and later extended by Meunier (2010) in relation to when the data may be made visible to the learners, either immediately or not.

Luciana Forti is responsible for paragraphs 2.2 and 3.2, while Stefania Spina is responsible for paragraphs 2.1 and 3.1. Paragraphs 1 and 4 were conceived and written jointly by the two authors.

The indirect uses of corpora in second language learning can be seen in approaches such as Contrastive Interlanguage Analysis (Granger 1996; 2015), where varieties of native and non-native instances of language use are compared and analysed, but also in language testing, when using errors found in a learner corpus as distractors in the development of multiple-choice items. Furthermore, indirect uses are found in lexicographical practices catering for learners (Granger, Paquot 2015; Paquot 2012; Spina 2010) and also in coursebook design (McCarten 2010).

Direct uses, on the other hand, aim to make corpus data immediately visible to the learners, so that they can explore it and use it within learning activities. This can be done both in a computer-based format, where the learners interrogate a corpus themselves (Muel-ler, Jacobsen 2016), or in a paper-based format, where concordance lines are previously selected by the teacher and printed on paper (Boulton 2010).

We can see how the indirect uses of corpora in second language learning partially overlap with Learner Corpus Research (LCR), while the direct uses of corpora mostly coincide with Data-driven learning (DDL). Our purpose here is to show how LCR can inform DDL through learner-friendly corpora.

2 Corpora for All?

As we have seen, native and learner corpora can be variously used in relation to second language learning. In the following paragraphs we will first describe the main corpora that are available for Italian, then we will identify the main traits characterising corpora that are specifically built for learners, and finally we will see what has been done so far, from this perspective, in the context of Italian L2 learning and teaching.

2.1 An Overview of Italian Native and Learner Corpora

A number of native and learner corpora have been constructed for the Italian language. As for native corpora, we can mention large written corpora like *Paisà* (Lyding et al. 2014), based on web data, the *Repubblica* corpus (Baroni et al. 2004), that includes sixteen years of the daily issues of the newspaper *Repubblica*, and CODIS (*Corpus Dinamico dell'Italiano Scritto*; Rossini Favretti, Tamburini, De Santis 2002), based on newspaper, academic, administrative and fiction texts. Among the native spoken corpora, the *Lessico di frequenza dell'italiano parlato* (LIP; De Mauro et al. 1993) was the first one to be created. A reference corpus of Italian, including both written and

spoken texts from a range of different textual genres, is the *Perugia corpus* (PEC; Spina 2014).

As for learner corpora, six main collections exist for Italian as a target language. Three were created at the University for Foreigners of Perugia: the *Longitudinal Corpus of Chinese Learners of Italian* (LOCCLI; Spina, Siyanova-Chanturia 2018; written); the *Corpus di apprendenti di Italiano L2* (CAIL2, written); and the *Corpus of Chinese Learners of Italian* (COLI, written and spoken). One was created at Eurac Bolzano: the *Merlin corpus* (Abel et al. 2014; written). One was created at the University of Torino: the VALICO corpus (*Varietà di apprendimento della lingua italiana Corpus Online*; Corino, Marello 2009; written). And one was created at the University for Foreigners of Siena: the *Lessico dell'italiano parlato per stranieri* (LIPS; Gallina 2013; spoken).

But how can native corpora be corpora for learners and not only corpora for linguists?

2.2 Towards Learner-Friendly Corpora

In order to be usable and useful for learners of a second language, corpora need to satisfy some basic requirements. First, the texts contained in the corpus need to be suitable for the learner's needs. These needs refer primarily to the difficulty level of the text: reference corpora are perhaps the most widely constructed and used corpora in any language, but because they are constructed with the purpose of describing a language in a representative way, from the perspective of native speakers, they would be usable mostly by advanced learners only, which constitute only a small proportion of the language learner population. On the contrary, the texts contained in a learner-friendly corpus would need to be suitable in terms of proficiency level, reflecting either one broad level or different proficiency or difficulty levels. Ideally, the nature of the texts contained in a learner-friendly corpus should also reflect the learners' interests, so that they are even more motivated to use the tool.

Second, a learner-friendly corpus would need to have a user-friendly interface, a learner friendly output and simple querying systems. Reference corpora built for linguists most typically require a user to formulate a query by using specific kinds of syntax. A learner would need some basic querying forms, that are graphically appealing and able to make the output easy to understand and then easy to be used for whatever learning need he/she may have.

A number of learner-friendly corpora have already been developed. One example is SCoRE (*Sentence Corpus of Remedial English*; Chujo, Oghigian 2012; Chujo, Oghigian, Akasegawa 2015), a corpus made of sentences that were extracted from a reference corpus and

then manually emended by language experts in order to make them suitable for English language learners at beginner level. Another example is SkELL (*Sketch Engine for language learning*; Baisa, Suchomel 2014), which is based on an algorithm that automatically selects 40 good examples for learners, drawn from a very large reference corpus; these examples are selected on the basis of criteria formulated with the aim to exclude, for instance, long subordinate clauses and/or words that are rare or technical. Finally, the idea of corpora based on graded readers has also been explored, but these corpora are yet to be put into practice on a systematic basis. In this case, texts are written by native experts in order to cater for different proficiency levels (Allan 2009; Gavioli, Aston 2001), though generally keeping in mind issues of authenticity and appropriateness (Hendry, Sheepy 2017).

The examples that were just outlined pertain to English language learning. With regards to Italian, two main kinds of attempts have been made. On the one hand, the texts contained in the corpora that were to be used by learners were chosen according to the specific learning aims of the learners, be it creative writing (Kennedy, Miceli 2001; 2010; 2017) or Italian for Specific Purposes (Polezzi 1993). On the other hand, an Italian version of SkELL (itSkeLL), with automatically selected examples from a large corpus, has recently become available.¹

Just like the English version, the Italian version of SkELL allows the learners to type in a word or word string into a search box and then explore it in terms of concordances, word sketches showing the words that most frequently co-occur with the word that was searched for, and finally a word cloud showing synonyms and other words that are similar to the searched one.

Other ideas as to how learner-friendly corpora can be used with learners are outlined in Naismith's article (2016). The author's primary aim is to show second language teachers how easy-to-use corpora can be integrated in second language lessons. He proposes tools such as Google Books Ngram Viewer, which uses an interface that is very similar to that of Google, which can then be easily integrated in lessons as needed. With Google Books Ngram viewer, the learner can look up the frequency of occurrence of words or word combinations throughout time. Two or more forms can be compared, for instance, in order to see which one is more frequently used in the present. The frequency is shown by means of a line graph. Another tool proposed by Naismith is Justtheword:² here, the learner can easily extract quantitative information about the usage of a word or word com-

1 <https://itskell.sketchengine.co.uk/>.

2 <http://www.just-the-word.com/>.

bination; in this case, the information is shown with examples as well as numerically, along with a wordle extension that is incorporated in the tool, allowing to visualise the results in the form of a word cloud.

These are only some of the ways in which corpora have been adapted for the needs of learners, instead of catering solely for the needs of researchers. In the following paragraph we will describe how a specific kind of learner-friendly corpus, namely SkELL, can be used to apply the finding coming from an LCR study.

3 From LCR to DDL

This paragraph describes how LCR and DDL can meet through it-SkELL. The empirical evidence emerging from the investigation of a learner corpus provides crucial data for the teacher: it can indicate error frequencies and frequent non-native usage patterns, and also shed light on how they both develop over time. In the following two sections we will describe an LCR study on Italian focused on collocations and based on the LOCCLI corpus.

3.1 An LCR Study on Italian

The study (Spina 2019) investigates the developmental patterns of phraseological errors in Chinese beginner and pre-intermediate learners of Italian in the use of noun + adjective (*tempo libero* ‘free time’) and adjective + noun (*bel tempo* ‘nice weather’) lexical combinations, and it is based on an error annotated sub-sample of the LOCCLI corpus. The main aim of the study is to verify the hypothesis that time affects errors in the combinations of nouns with adjectives produced by beginner and pre-intermediate Chinese learners of Italian.

One of the major findings is that noun + adjective and adjective + noun combinations display opposite behaviours across time with respect to the production of these specific phraseological errors: errors decrease after six months for adjective + noun combinations, while they significantly increase for noun + adjective combinations. An error of the type *Ho trovato gli spagnoli ragazzi sono non più belli di italiani ragazzi* (‘I found that Spanish boys are more good-looking than Italian boys’), where the correct form *ragazzi spagnoli* (‘Spanish boys’) is replaced by a form with the adjective wrongly preceding the noun, tends therefore to increase over time.

3.2 Building an LCR-Informed DDL Activity

The main finding of the study above consists in the observation that errors tend to increase over six months. This finding may lead the teacher to want to place a particular focus on this kind of error. This does not need to be explicit or out of context, but it can stem from the observation of an error in a learner's writing or speech. The teacher will know that this kind of error does not tend to decrease over time, but, on the contrary tends to increase. And this means that it may deserve some additional attention.

Starting from the example indicated in the previous paragraph, the teacher may point out the error to the student(s), without providing any kind of corrective feedback. Then the student(s) will be asked to open SkELL and search for *spagnoli ragazzi*. The student(s) will see an empty screen with no results.

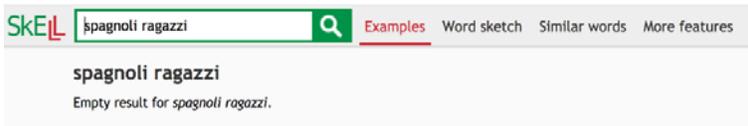


Figure 1 DDL activity – step 1

At this point, the teacher will invite the student(s) to search again by inverting the position of the two words, *ragazzi spagnoli*. In this case, the student(s) will find a number of examples and this will allow him or her to infer the regularity: in this case, the adjective always follows the noun.

 A screenshot of the SkELL search interface. The search bar contains the text "ragazzi spagnoli" and a magnifying glass icon. To the right of the search bar are several tabs: "Examples" (highlighted in red), "Word sketch", "Similar words", and "More features". Below the search bar, the text "ragazzi spagnoli" is displayed, followed by "0.03 hits per million". Below this, there is a list of 12 search results, each starting with a number in a small box. The results are:

- 1 La prossima tappa sarà il Progetto Leonardo, che coinvolgerà **ragazzi spagnoli** già diplomati in Agraria e di età compresa tra i 18 e i 22 anni.
- 2 Video fa sorridere il mondo intero Diversi **ragazzi spagnoli** hanno lasciato la...
- 3 Yafray è impressionante ed i **ragazzi spagnoli** che lo sviluppano sono veramente molto disponibili.
- 4 L'esperienza si concluderà con un viaggio di una settimana a Madrid per gli studenti di III Media, settimana nella quale i ragazzi frequenteranno la mattina le lezioni al "Colegio San Ramon y San Antonio" insieme ai **ragazzi spagnoli**, il pomeriggio, invece, si realizzeranno visite culturali nella capitale spagnola.
- 5 Da lunedì 9 a venerdì 13 Maggio due **ragazzi spagnoli** e un italiano saranno ospiti degli uffici di via Milano della Provincia di Brescia
- 6 Almeno, è quello che hanno fatto due **ragazzi spagnoli**, Jimmy Martinez e Pablo Luiz (entrambi della Catalogna) con Immfy.
- 7 Per un'intera settimana, i **ragazzi spagnoli** accompagnati dalle docenti Aurora Figueredo e Nieves Leon, sono stati ospiti dei nostri alunni vivendo la quotidianità della vita scolastica e familiare italiana.
- 8 Frena il programma Erasmus nel 2008 e Zapatero stanza 35 milioni di euro per aumentare la borsa di studio base comunitaria per gli studenti e rilanciare il programma e la mobilità dei **ragazzi spagnoli** ...
- 9 Tutto ciò in compagnia dei sorridenti e divertenti **ragazzi spagnoli**, sempre pronti a scherzare, a ridere e comunicare, insegnandoci nuovi aspetti della loro cultura, oltre a nuove parole ed espressioni sia in castigliano che in lingua basca.
- 10 "Ospitiamo volentieri questi giovani **ragazzi spagnoli** spiega l'Assessore delle Politiche per i giovani Cristian Zara di ritorno da un così importante appuntamento per la cristianità mondiale, al quale partecipano anche molti giovani della nostra comunità.
- 11 Come da progetto consolidato, le giornate dei **ragazzi spagnoli** saranno scandite da ore di stage presso aziende del settore della ristorazione e della vendita, da ore di lezione di lingua e cultura italiana e da visite in luoghi culturali e tipici di Torino e della provincia.
- 12 I ricercatori dello Spanish National Research Council di Madrid, guidati da David Gómez Martínez hanno preso in esame 1700 **ragazzi spagnoli**, tra i quali 892 donne, residenti nelle città di Madrid, Saragozza, Granada, Santander e Murcia.

Figure 2 DDL activity – step 2

This corresponds to the simplest type of DDL activity: it guides the learner to observe whether a certain form that he or she has used in writing or speech is actually used by native speakers or not. Other activity types may involve a number of different steps building up a guided discovery process, like the many examples shown in Sinclair (2003).

The data contained in a simple activity such as the one that was shown can lend itself to be used in subsequent lessons, with the aim of recycling and increasing the frequency of input of the given structure.

For example, the set of concordance lines can be used to construct a multiple-sentence gap-fill exercise, such as the one shown in figure 3.

1	La prossima tappa sarà il Progetto Leonardo, che coinvolgerà spagnoli già diplomati in Agraria e di età compresa tra i 18 e i 22 anni.
2	Video fa sorridere il mondo intero Diversi spagnoli hanno lasciato la...
3	Yafra è impressionante ed i spagnoli che lo sviluppano sono veramente molto disponibili.
4	L'esperienza si concluderà con un viaggio di una settimana a Madrid per gli studenti di III Media, settimana nella quale i ragazzi frequenteranno la mattina le lezioni al "Colegio San Ramon y San Antonio" insieme a spagnoli , il pomeriggio, invece, si realizzeranno visite culturali nella capitale spagnola.
5	Da lunedì 9 a venerdì 13 Maggio due spagnoli e un italiano saranno ospiti degli uffici di via Milano della Provincia di Brescia
6	Almeno, è quello che hanno fatto due spagnoli , Jimmy Martinez e Pablo Luiz (entrambi della Catalogna) con Immfly.
7	Per un'intera settimana, i spagnoli accompagnati dalle docenti Aurora Figueredo e Nieves Leon, sono stati ospiti dei nostri alunni vivendo la quotidianità della vita scolastica e familiare italiana.
8	Frena il programma Erasmus nel 2008 e Zapatero stanziò 35 milioni di euro per aumentare la borsa di studio base comunitaria per gli studenti e rilanciare il programma e la mobilità dei spagnoli ...
9	Tutto ciò in compagnia dei sorridenti e divertenti spagnoli , sempre pronti a scherzare, a ridere e comunicare, insegnandoci nuovi aspetti della loro cultura, oltre a nuove parole ed espressioni sia in castigliano che in lingua basca.
10	"Dspittiamo volentieri questi giovani spagnoli spiega l'Assessore delle Politiche per i giovani Cristian Zara di ritorno da un così importante appuntamento per la cristianità mondiale, al quale partecipano anche molti giovani della nostra comunità.
11	Come da progetto consolidato, le giornate dei spagnoli saranno scandite da ore di stage presso aziende del settore della ristorazione e della vendita, da ore di lezione di lingua e cultura italiana e da visite in luoghi culturali e tipici di Torino e della provincia.
12	I ricercatori dello Spanish National Research Council di Madrid, guidati da David Gómez Martinez hanno preso in esame 1700 spagnoli , tra i quali 892 donne, residenti nelle città di Madrid, Saragozza, Granada, Santander e Murcia.

Figure 3 DDL activity – step 3

The main advantage of a multiple sentence gap-fill exercise over a single sentence gap-fill is that the learner is able to test an initial hypothesis with multiple examples. In this case, there are 12 sentences containing the same word combination and gapped in relation to the same member of the word combination, namely the noun collocates.

By looking at the first sentence, the learner will be able to notice that the noun can only be masculine and plural, considering that these are the properties characterising the adjective *spagnoli*. So, ideally with a partner or with a group of peers in the classroom, the learner will start exploring options: *studenti* or *ragazzi*, for instance. The context here is that of education, since the sentence seems to evoke an educational program that caters for people aged between 18 and 22, so the missing noun could be one of the two. Then, the learner or group of learners will proceed towards the second example. Here, the context does not seem to be related to education, so of our two initial hypotheses, *studenti* or *ragazzi*, we can only retain the second one. The third example has the potential of clearing any doubts: the article that used before the noun is *i*, which is not used in front of nouns that begin with *st*. As a result, the only possible noun occurring in all of these examples is *ragazzi*.

The students may not be able to find all the clues provided by the concordance lines immediately. This is why, while fostering group work, the teacher will have provided a guided-discovery procedure that the learners can follow. This procedure most typically comes in the form of guiding questions (as shown in Sinclair 2003), but it can also come in the form of questions with a series of options, especially in cases where the proficiency level of the classroom is low, thus providing additional scaffolding. The teacher may also circulate throughout the classroom while the students engage in explorative activities such as this one, and provide even further scaffolding whenever needed.

An activity such as this one is initiated as something focused on a single word combination, which in this case is *ragazzi spagnoli*, but as the learners go through it, it can easily extend into different areas: as mentioned previously, the teacher can draw the students' attention to the use of articles depending on the how the words following them begin, but also to the verbs that this noun phrase co-occurs with. What do these *ragazzi spagnoli* actually do? Not all examples will provide an answer to this, but those that do can lead the students to observe that the *ragazzi spagnoli* students develop things (line 3), are being invited as guests somewhere (line 5), are accompanied by teachers somewhere (line 7), are part of mobility programs (line 8), are always ready to joke and laugh and talk (line 9), and take part in training programs (line 11). A teacher will easily see great potential in all of these examples, and will be able to further extend the DDL activity towards the avenues that will be deemed closer to the learners' needs.

This would not be possible with a traditional single sentence gap-fill activity. What students most typically find themselves doing are exercises with lists of single sentences containing specific learning aims, in terms of grammatical features or lexis. Each sentence presents a cotection for the learning aim, but is ultimately devoid of context. The time spent on each item will be limited and the attention span of the learner will be likely to decrease as he or she proceeds, because each new sentence will evoke a different kind of thematic context.

In the case of a DDL activity stemming from a single word combination, such as the one we presented, the discovery process that is initiated can take a number of different turns, all of which will be logically linked. This is how the student will find him or herself in the position of a detective, a research-scientist or a traveler (Bernardini 2000; Cobb 1999; Johns 1997), while the teacher will become a demonstrator, a collaborator or a guide (Boulton 2011; Charles 2014; Frankenberg-Garcia 2012). And this way, the discovery process put into place by the DDL activity will allow the learner to gain insight into language usage and the form varieties in an autonomous way: as Cobb points out in his contribution on constructivism, "knowledge encoded from data by learners themselves will be more flexi-

ble, transferable, and useful than knowledge encoded and transmitted to them by an instructor” (Cobb 1999, 15).

However, what this study attempts to show is that a further step can be taken. If a DDL activity is linked to the empirical evidence deriving from an LCR finding, that activity is likely to be more effective, because its focus will be empirically motivated.

We have stated that bridging the gap between LCR and DDL can be done by using LCR findings to inform DDL practices. But what this means in practice is using LCR findings to decide what to focus on and how to sequence it and modulate it within pedagogical material design. If a teacher is informed about the empirical evidence that is attached to a specific kind of error, he or she will be able to make an informed decision as to how to treat it whenever it will be encountered. If an LCR finding says that a particular error is not very common, the teacher can decide to evaluate it as something temporary; on the contrary, if a finding says that an error is quite common, especially as time goes by, in this case, the teacher can decide to devote some special attention to the error, in order to counteract the learning pattern attached to a given form, unveiled by the LCR finding.

4 Conclusions

In this article we tried to explore ways in which the uses of corpora built for linguists can be merged with the uses of corpora built for learners. More fruitful exchanges between the two imply a number of desiderata. In terms of learner corpora of Italian, we need more collaborative work that is able to ensure more accurate corpus design criteria. Learner corpora for Italian need to be larger in size, their longitudinal dimension needs to be increased, and the computational tools that are used to process and extract data from them need to be more sophisticated.

On the other hand, the effectiveness of different kinds of learner-friendly corpora should be further explored, in relation to the specific needs of the different teaching contexts. Tools that are suitable for use by intermediate and lower-intermediate learners of Italian should be explored and new ones should be created, considering that this is generally the largest group of learners learning Italian. Furthermore, in order to integrate LCR findings into DDL practices, and teaching practices in general, teachers need to be aware of research findings, but researchers need to be aware of teachers’ needs: progress in language learning and teaching methods is arguably based on good research and good education policies, but also on awareness of learner needs and teacher needs. As a result, the gap that needs to be bridged is not only that between LCR and DDL, but also that between researchers and teachers.

References

- Abel, A.; Wisniewski, K.; Nicolas, L.; Boyd, A.; Hana, J.; Meurers, D. (2014). "A Trilingual Learner Corpus illustrating European Reference Levels". *Ricognizioni – Rivista di Lingue, Letterature e Culture Moderne*, 2(1), 111-26. <http://dx.doi.org/10.13135/2384-8987/702>.
- Allan, R. (2009). "Can a Graded Reader Corpus Provide 'Authentic' Input?". *ELT Journal*, 63(1), 23-32. <https://doi.org/10.1093/elt/ccn011>.
- Baisa, V.; Suchoamel, V. (2014). "SkELL: Web Interface for English Language Learning". Horák, A.; Rychlý, P. (eds), *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2014*. Brno: NLP Consulting, 63-70.
- Baroni, M. et al. (2004). "Introducing the 'la Repubblica' corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian". Lino, M.T. et al. (eds), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. Lisbon: European Language Resources Association (ELRA), 1771-4. <https://www.aclweb.org/anthology/L04-1128/>.
- Bernardini, S. (2000). "Systematising Serendipity: Proposals for Concordancing Large Corpora with Language Learners". Burnard, L.; McEnery, T. (eds), *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang, 225-34.
- Boulton, A. (2010). "Data-Driven Learning: Taking the Computer Out of the Equation". *Language Learning*, 60(3), 534-72. <https://hal.archives-ouvertes.fr/hal-00384910v2>.
- Boulton, A. (2011). "Data-Driven Learning: The Perpetual Enigma". Gozdz-Rozkowski, S. (ed.), *Explorations Across Languages and Corpora*. Bern: Peter Lang, 563-580.
- Charles, M. (2014). "Getting the Corpus Habit: EAP Students' Long-Term Use of Personal Corpora". *English for Specific Purposes*, 35, 30-40. <https://doi.org/10.1016/j.esp.2013.11.004>.
- Chujo, K.; Oghigian, K. (2012). "DDL for EFL Beginners: A Report on Student Gains and Views on Paper-Based Concordancing and the Role of L1". Thomas, J.E.; Boulton, A. (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, 130-43.
- Chujo, K.; Oghigian, K.; Akasegawa, S. (2015). "A Corpus and Grammatical Browsing System for Remedial EFL Learners". Leńko-Szymańska, A.; Boulton, A. (eds), *Multiple Affordances of Language Corpora for Data-Driven Learning*. Amsterdam; Philadelphia: Benjamins, 109-28. <https://doi.org/10.1075/scl.69.06chu>.
- Cobb, T. (1999). "Applying Constructivism: A Test for the Learner-as-Scientist". *Educational Technology Research and Development*, 47(3), 15-31. <https://doi.org/10.1007/BF02299631>.
- Corino, E.; Marelllo, C. (a cura di) (2009). *VALICO Studi di linguistica e didattica*. Perugia: Guerra Edizioni.
- De Mauro, Tullio et al. (1993). *Lessico di frequenza dell'italiano parlato*. Milano: Etas Libri.
- Frankenberg-Garcia, A. (2012). "Integrating Corpora with Everyday Language Teaching". Thomas, J.E.; Boulton, A. (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, 33-51.

- Gallina, F. (2013). "The LIPS Corpus (Lexicon of Spoken Italian by Foreigners): A Study on the Acquisition of Vocabulary by L2 Italian Learners, Between Measures of Lexical Richness and Lexical Fields". Granger, S.; Gilquin, G.; Meunier, F. (eds), *Twenty Years of Learner Corpus Linguistics Research. Looking Back, Moving Ahead*. Louvain: Presses Universitaires de Louvain, 179-95.
- Gavioli, L.; Aston, G. (2001). "Enriching Reality: Language Corpora in Language Pedagogy". *ELT Journal*, 55(3), 238-46. <https://doi.org/10.1093/elt/55.3.238>.
- Granger, S. (1996). "From CA to CIA and back: An Integrated Approach to Computerized Bilingual and Learner Corpora". Aijmer, K.; Altenberg, B.; Johansson, M. (eds), *Languages in Contrast = Papers from a Symposium on Text-Based Cross-Linguistic Studies* (Lund, 4-5 March 1994). Lund: Lund University Press, 37-51.
- Granger, S. (2015). "Contrastive Interlanguage Analysis: A Reappraisal". *International Journal of Learner Corpus Research*, 1(1), 7-24. <https://doi.org/10.1075/ijlcr.1.1.01gra>.
- Granger, S.; Paquot, M. (2015). "Electronic Lexicography Goes Local: Design and Structures of a Needs-Driven Online Academic Writing Aid". *Lexicographica - International Annual for Lexicography*, 31(1), 118-41. <https://doi.org/10.1515/lexi-2015-0007>.
- Hendry, C.; Sheepy, E. (2017). "Evaluating Lexical Coverage in Simple English Wikipedia Articles: A Corpus-Driven Study". Borthwick, K.; Bradley, L.; Thouëсны, S. (eds), *CALL in a Climate of Change: Adapting to Turbulent Global Conditions - Short Papers from EUROCALL 2017*, 146-50. Research Publishing. <https://doi.org/10.14705/rpnet.2017.eurocall2017.704>.
- Johns, T. (1997). "Contexts: The Background, Development and Trialling of a Concordance-Based CALL program". Wichmann, A.; Fligelstone, S.; McEnery, T. (eds), *Teaching and Language Corpora*. Harlow: Addison Wesley Longman, 100-15. <https://doi.org/10.4324/9781315842677-9>.
- Kennedy, C.; Miceli, T. (2001). "An Evaluation of Intermediate Students' Approaches to Corpus Investigation". *Language Learning & Technology*, 5(3), 77-90.
- Kennedy, C.; Miceli, T. (2010). "Corpus-Assisted Creative Writing: Introducing Intermediate Italian Learners to a Corpus as a Reference Resource". *Language Learning & Technology*, 14(1), 28-44.
- Kennedy, C.; Miceli, T. (2017). "Cultivating Effective Corpus Use by Language Learners". *Computer Assisted Language Learning*, 30(1-2), 91-114. <https://doi.org/10.1080/09588221.2016.1264427>.
- Larsen-Freeman, D. (1997). "Chaos / Complexity Science and Second Language Acquisition". *Applied Linguistics*, 18(2), 141-65. <https://doi.org/10.1093/applin/18.2.141>.
- Leech, G. (1997). "Teaching and Language Corpora: A Convergence". Wichmann, A.; Fligelstone, S.; McEnery, T.; Knowles, G. (eds), *Teaching and Language Corpora*. Harlow: Addison Wesley Longman, 1-24.
- Lyding, V. et al. (2014). "The PAISA Corpus of Italian Web Texts". Bildhauer, F.; Schäfer, R. (eds), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*. Gothenburg: Association for Computational Linguistics, 36-43. <https://www.aclweb.org/anthology/W14-0406/>.
- McCarten, J. (2010). "Corpus-Informed Course Book Design". O'Keeffe, A.; McCarthy, M. (eds), *The Routledge Handbook of Corpus Linguistics*. London; New York: Routledge, 413-27.

- McEnery, A.; Xiao, R.; Tono, Y. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. London; New York: Routledge.
- Meunier, F. (2010). "Learner Corpora and English Language Teaching: Check-up Time". *Anglistik: International Journal of English Studies*, 21(1), 209-20.
- Mueller, C.M.; Jacobsen, N.D. (2016). "A Comparison of the Effectiveness of EFL Students' Use of Dictionaries and an Online Corpus for the Enhancement of Revision Skills". *ReCALL*, 28(1), 3-21. <https://doi.org/10.1017/S0958344015000142>.
- Naismith, B. (2016). "Integrating Corpus Tools on Intensive CELTA courses". *ELT Journal*, 71(3), 273-83. <https://doi-org.eres.qnl.qa/10.1093/elt/ccw076>.
- Paquot, M. (2012). "The LEAD Dictionary-cum-Writing Aid: An Integrated Dictionary and Corpus Tool". Granger, S.; Paquot, M. (eds), *Electronic Lexicography*. Oxford: Oxford University Press, 163-86.
- Polezzi, L. (1993). "Concordancing and the Teaching of ab initio Italian Language for Specific Purposes". *ReCALL*, 5(9), 14-18. <https://doi.org/10.1017/S0958344000004067>.
- Quirk, R. (1960). "Towards a Description of English Usage". *Transactions of the Philological Society*, 59(1), 40-61. <https://doi.org/10.1111/j.1467-968X.1960.tb00308.x>.
- Rossini Favretti, R.; Tamburini, F.; De Santis, C. (2002). "A Corpus of Written Italian: a Defined and a Dynamic Model". Wilson, A.; Rayson, P.; McEnery, T. (eds), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*. Munich: Lincom-Europa, 27-38.
- Sinclair, J.M. (2003). *Reading Concordances*. London: Pearson.
- Spina, S. (2010). "The Dictionary of Italian Collocations: Design and Integration in an Online Learning Environment". Calzolari, N.; Choukri, K.; Maegaard, B.; Mariani, J.; Odijk, J.; Piperidis, S.; Rosner, M.; Tapias, F. (eds), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valetta: European Language Resources Association (ELRA), 3202-8.
- Spina, S. (2014). «Il Perugia Corpus: Una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione». Basili, R.; Lenci, A.; Magnini, B. (eds), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-It 2014 & the Fourth International Workshop EVALITA 2014*, vol. 1. Pisa: Pisa University Press, 354-9. <http://clc2014.fillel1.unipi.it/proceedings/vol1/CLICIT2014168.pdf>.
- Spina, S. (2019). "The Development of Phraseological Errors in Chinese Learner Italian: A Longitudinal Study". *Widening the Scope of Learner Corpus Research. Selected Papers from the Fourth Learner Corpus Research Conference*. Louvain-la-Neuve: Presses universitaires de Louvain, 95-119.
- Spina, S.; Siyanova-Chanturia, A. (2018). "The Longitudinal Corpus of Chinese Learners of Italian (LOCCLI)". Poster presented at the 13th Teaching and Language Corpora conference, University of Cambridge.

