The "Blakkat" Software for Tagging Online Language Learner Corpora: Issues in SLA Assessment and Research

Simone Torsani

THE "BLAKKAT" SOFTWARE FOR TAGGING ONLINE LANGUAGE LEARNER CORPORA: ISSUES IN SLA ASSESSMENT AND RESEARCH

SIMONE TORSANI

Dipartimento della Comunicazione Linguistica e Culturale (DISCLIC),
Università degli Studi di Genova,
P.zza S. Sabina, 2, 16100 Genoa, Italy
e-mail: s.torsani@gmail.com

Abstract: This paper aims at showing, through a case study, one possible application of Computer Learner Corpus (CLC) to Network Based Language Teaching (NBLT). Research has shown how CLC can be used both for Second Language Acquisiton (SLA) research and Foreign Language Teaching (FLT), especially if they are tagged, that is, if interpretative annotations are added to the corpus (e.g. error annotations). Online learning generally takes place inside virtual environments where learners exchange mainly written asynchronous productions, which can be easily collected into a CLC. Although different tools and systems are used for coding texts, we have developed a flexible web based interface where researchers can define and apply their own set of tags to a given corpus and work at distance. As with many web-based tools, it is not the tool only that is interesting, rather the possibility to transfer the software core, that is its main algorithms, inside a Virtual Learning Environment (VLE), in order to allow teachers and researchers to code and analyse their learners productions for those uses shown in literature.

Keywords: Computer Learner Corpus, Network Based Language Teaching, Computer Assisted Second Language Research, Instructional Design

1. INTRODUCTION: CORPUS LINGUISTICS AND LEARNER CORPORA

Corpus linguistics1 has nowadays a long tradition as a discipline related to Computer-Assisted Language Learning (CALL), the first experiments of concordancer used as a software in language instruction dating back to the 1970s (Leech, 1997, p.2). Although now a major issue in Computer Applications in Second Language Acquisition (CASLA2, Chapelle, 2001, 36-38), Corpus Linguistics in language learning and teaching is still perceived by many mainly in relation to the well known paradigm of "data-driven learning", developed by Johns,

---

[1] A recent overview and reference of Corpus Linguistics is to be found in Stubbs (2004), while an up to date reference on Corpus Linguistics and language teaching is found by Mukherjee (2004).

[2] Although the acronym CALL is the most used in specialised literature (Gruba, 2004, p. 623) and is usually referred to through Levy's definition –"the search for any study of applications of the computer in language teaching and learning" - I have preferred Chapelle's acronym CASLA, "an umbrella phrase" *(ibid.)* covering CALL activities, as well as computer-assisted language assessment and second language research.

who introduced research methods in language learning as explained by his well-known maxim "research is too serious to be left to researchers" (Johns, 1991, p. 2). According to this perspective corpora are to be perused by the learner who acts "as 'linguistic researcher', testing and revising hypotheses, or as 'language detective', learning to recognize and interpret clues from context ('Every student is a Sherlock Holmes')" (Johns, 1997, p. 101).

But if we were to limit the extent of language corpora to teaching by exploiting a concordance, we might be missing an important share of their potential[3]. Computer Learner Corpora (CLC)[4] are, as it can be easily argued, collections of learners language productions and are defined as "electronic collections of authentic FL/SL [foreign language / second language] textual data assembled according to explicit design criteria for a particular SLA/FLT [second language acquisition / foreign language teaching] purpose" (Granger, 2002, p. 7). Granger, whose work is considered seminal in this respect, edited in 1998 a collection of studies dealing with learner corpora, mainly oriented at comparing the learners' language with that of native-speakers: although this seems to be the main perspective (Hunston, 2002, p. 206), learner corpora can be used for a wide range of activities which also encompass syllabus and materials design (Mukherjee, 2005, p. 17). Most of the papers in Granger's volume make use of the International Corpus of Learner English (ICLE), a database collected under the direction of Granger herself, containing corpora (200,000 words each, that is medium size corpora[5]) collecting language written productions of advanced learners from different countries, while the native-speakers corpus is the LOCNESS Corpus (300,000 words). Besides this corpus Granger also directed the collection of the Louvain International Database of Spoken English Interlanguage (LINDSEI, de Cock, 2003), comprising oral productions of advanced learners form different countries as well. Another important corpus is the Longman Learner Corpus (of intermediate and advanced learners) used by Biber and Reppen in the same volume. Mukherjee mentions also the Giessen Long Beach Chaplin Corpus.

These corpora show evidences of having been collected for SLA research: data, for instance, are organised so as to reflect an average learner, whom Granger refers to as the "archetypal learner" (1998, p. 7), and not, therefore, for any assessment practice. This is a most important point, since it is necessary to distinguish between SLA Research, Language Teaching Research and Instructional Design: this distinction, apart from academic matters, is fundamental in defining the objectives and tasks for which a CLC is investigated. SLA Research was born in the 1950/60s as a sub field of Language Teaching and gradually evolved into an autonomous discipline – that is without necessarily having practical applications such as improving teaching practices - since the 1970s (Chini, 2005, p. 19) and has now become a major field of Applied Linguistics[6]. Language Teaching Research, on the contrary, is a more practical discipline, mainly interested in developing effective methodologies, that is "the study of the system or range of methods that are used in teaching" (Adamson, 2005, p. 604). This said, however, researchers advise not to "draw strict dividing lines" (Littlewood, 2005, p. 502) between these two fields, since findings in the first can have consistent influence over the second, since, as we noticed, SLA research originally was a sub discipline of language teaching. Instructional design, finally, is a most practical activity consisting in an inquiry field whose object is defining "the rules governing the choice of the

---

[3] Another documented use of corpora in SLA is the creation of CALL exercises using them (Wilson,1997).

[4] A full list of coprora types is provided by Hunston (2002, pp. 14-16).

[5] See Chiari (2007, p. 45) for a reference table of corpus sizes.

[6] For an introduction to SLA research issues see Adamson (2005) and Chini (2005).

most suited instruction means according to the conditions of learning" (Ranieri, 2005, my translation).

2. LEARNER CORPORA IN DEPTH

Having introduced its historical and theoretical foundations, and having defined the related disciplines, it is now possible to have a closer look at CLC at work. Although quite a recent field of inquiry (Granger, 2004, p. 128), learner corpus analysis has given rise to a number of expectations and experiments SLA researchers and teachers should not neglect. As it was noted by Hunston, "the essence of learner corpora is comparison" (2002, p. 206): they are useful, in other words, as they can be compared, for instance, with native speakers corpora in order to gather evidence on SLA issues and make "interesting, qualitative generalisations" *(ibid.)*. Although certainly true as far as research is concerned, this perspective can be enlarged and it is possible to make such collections available to more practical uses, such as improving assessment, instructional design and teaching materials (Chapelle, 2001, p. 37).

Mukherjee & Rohrbach (2006) provide a recent and in-depth overview, as well as practical examples, on how to work with such materials. Their research, focused on *discourse markers*, shows how conclusions can be drawn from their corpus of German learners: the results suggest that corpora "provide a good starting point for an improvement of ELT textbooks and materials, given that – apart from *well* – all other discourse markers are notoriously underrepresented even in modern materials." (p. 216). They concentrate, then, on two further perspectives: *individualisation* and *localisation*. Learner corpora, in fact, are often viewed as "reference learner corpora" (Mukherjee, 2006, p. 18), since they collect evidence as regards an "abstract learner" (Granger, 1998, p. 7), that is the whole body of learners, careless of individual and local differences. This kind of CLC is useful for SLA research: for instance, it could be used to test the Language Acquisition Device theory, although Mukherjee & Rohrbach, on the other hand, suggest that corpus analysis could be deepened to a more individual and contestualised level.

Individual analysis "is intended to mean that for the purpose of individual assessment and analysis of the variation between learners, it would be useful to complement the learner-language- as-a-whole perspective by also taking into account the differences between learners." (Mukherjee & Rohrbach, 2006, p. 217). The use of the term *assessment* is important here, since it hints at the fact that learner corpora could be viewed not only as research materials, but also as an instrument through which teachers can assess learners, i.e. individual learners' works and productions. On the other hand, *localised* analysis is focused more on the classroom at a specific stage (p., 219-220), and is even more focused on issues of language assessment and instructional design, since it aims at evaluating the learning process at a given time. As shown by Mukherjee & Rohrbach in their example of a localised learner corpus, "one of the key advantages of a computerised local learner corpus for the teacher: he/she is now able to immediately identify patterns of use in learner language across all students in his/her class. The identification of 'typical mistakes' can now be put on an empirical footing and be based on the analysis of all learners' actual output" (p. 225).

Learner corpora, then, can be examined through the most common concordance software and they become even more interesting and revealing if they are tagged. Although a learner corpus can be accessed "as is", that is, even if it is a raw collection of (learners) text, applying a set of tags to this data increases its potential and allows quantitative data to provide qualitative ones (Granger, 2004, p. 129). Different experiments have been conducted: an instance of corpus tagging can be glimpsed in Nicholls (2003), where an extremely accurate set of tags is shown along with the advantages of a coded corpus over a raw one. In Nicholls

errors are generally tagged with a two-letter code, where the first letter defines the "type of error" and the second the "word class": this results in a very exact error-tagging operation, where, for instance, erroneous forms can be separated from correct ones, and the different kinds of errors extracted and considered separately. Researchers use different tools in order to tag they corpora, especially POS taggers[7], software coding the text automatically (Granger, 2004, p. 128). The analysis of one such collection is that common corpus software, Wordsmith[8] for instance, allows users to define a set of tags to be used along with their tagged corpus.

If, as we noticed, learner corpora can be used either for research or for pedagogical purposes, it is also evident, however, that research is normally allowed much more time and resources than assessment or instructional design are. If teachers were to perform such complex operations as those described in the literature in order to assess their students' errors, it goes without saying that it would be very hard to convince them to adopt this practice. There is evidence, nevertheless, that such practice is of great value as regards pedagogical matters. In the following paragraphs I will show the case of online language learning; inside Virtual Learning Environments, in fact, language exchanges generally take place inside web forums, that is ready-to-use and organised electronic text. In this case a longitudinal perspective[9] could easily be applied, provided the adequate tools are available, to small learner corpora in order to assess the whole learning process, either at individual or general level (Mukherjee & Rohrbach, 2006).

3. THE BLAKKAT SOFTWARE

The Blakkat software was originally conceived and developed for the analysis of a corpus of forum messages in order to allow different researchers to cooperatively assess the social presence in a post-degree course and not, therefore, for a strictly linguistic investigation[10]; it was soon found, however, that it could be effectively used also for assessing issues of language learning. The software allows one or more researchers to define and apply an arbitrary set of tags to a given corpus and then extract and compare data through different filters.

Every project has a coordinator who defines the set of tags and the documents to be coded. Tags are identified through name, colour (tag colour is very important since it is through colour that tags are printed on the screen while a text is coded) and a short description, and can be exported/imported for recycling within other projects. Text, on the other hand, can be written/pasted inside a form or imported from an XML file that can be easily produced by many database management systems: most Virtual Learning Environments, say Moodle, are based on PHP technology that relies on a MySQL database.

---

[7] POS (Part Of Speech) taggers apply a part-of-speech label to words in a corpus, such words as *light* might appear in different contexts as a noun, an adjective, a verb or an adverb. See Hunston (2002) pp. 79-92 for an overview of corpus annotation and parsing.

[8] http://www.lexically.net/wordsmith/

[9] In SLA research "longitudinal" refers to a diachronic analysis of the learners' development; for a recent overwiew of this issue see Ortega L. & Iberri-Shea (2005).

[10] Lupi V., Pozzi, F, Torsani S., *Développer la dimension sociale dans un master post-universitaire à distance: outils, animation et analyse des interactions*. The paper, presented at the EPAL conference in Grenoble (June 2007, in press) contains a brief overview of the software.

Difficult as this may seem, it is a short and easy operation, provided one has access to the database and a basic knowledge of this software.

Every text of the corpus, is associated with a title, a timestamp[11] date, so that timelines can be produced with the data, and an extra field where an arbitrary value, the text's author for instance, can be inserted: in this way comparative timelines can be produced. The tagging mask is very easy to use: the text to be coded is in a form field allowing formatting and, for each tag defined for the project, a button is available inside the form through which users can tag a test portion in a visual way, i.e. they highlight the interested portion with the colour corresponding to a given tag.

When the texts are tagged, a set of filters allows the user to extract the collected data in different ways, in order to:

- generate statistics on one or more tags
- generate timelines on the arbitrary value (e.g. the user, so that a timeline of different tags is produced)
- export the tagged corpus and the corresponding set of tags so that it is possible to analyse it through a specific corpus software (e.g. Wordsmith).

4. CASE STUDY OF CLC APPLICATION TO NETWORK-BASED LANGUAGE TEACHING: TAGGING A FRENCH BEGINNERS CORPUS

The impact of the Internet on language teaching in general and on CASLA in particular has been, as it can be easily imagined, a great one and soon became the subject of a vast literature. A most important point in this history is the volume, edited in 2000 by Warschauer and Kern, "Network-based Language Teaching: Concepts and Practices". Although the NBLT scenario has somehow changed since the volume publication, some of the ideas underlying its whole content are still valid and constitute a basic reference for all those interested in web based language teaching: in particular, the concept that "NLBT represents a new and different side of CALL, where human-to-human communication is the focus" (Warschauer & Kern, 2000, p. 1) is an essential one. To what extent this principle is central is well explained by Gruba (2004, p. 624), who points out how pedagogical approaches have changed along with the introduction of the Internet as a CASLA tool, and by the various works focusing on such issues such as Computer Supported Collaborative Learning (CSCL) in SLA (see. Chapelle, 2001, pp. 31-32) or Computer-Mediated Communication (CMC), one of the main topic areas of *Language Learning & Technology*, a professional online journal devoted to CASLA[12].

If, by the time Warschuer and Kern edited their volume, online learning environments were only a perspective or still at a prototypical stage (Chun & Plass, 2000), today literature illustrates how this field has vigorously developed, and how instruction increasingly takes place online inside various learning environments. Since socio-collaborative approaches are becoming dominant in SLA practice (Gruba, 2004, p. 624), a number of collaborative tools have been developed, planned or whished-for to foster collaboration and exchange among learners (Calvani et al., 2005); the web forum, however, remains the most used tools for such

---

[11] A timestamp is a measure of time consisting in the number of seconds since January 1st 1970: data transmitted in this format can be differently transformed into dates or easily used for computation.

[12] http://llt.msu.edu/index.html

tasks and different works focus on web forums as the pivot location where learners interaction takes place[13].

4.1. Tagging case study

In this case study I have collected a small sample corpus (2500 words approximately) of 11 second year French learners and analysed it as regards simple errors (spelling, missing words and words to be replaced), that is through a set of tags defining such errors. Small as this corpus may seem, however, it must be borne in mind that CLC tagging and analysis for assessment does not happen on large collections. The corpus consists of different forum messages written in a time span of four months (November 2006 - March 2007) in the Moodle VLE; students were required to interact with French students attending a Master FLE course, that is an university course in teaching French as a Foreign Language. I have easily extracted the messages from the database, since the University server administrators have chosen (like many) the phpMyAdmin administration system to access the MySQL database which allows users to export data in the XML format. I have defined a set of tags, in order to extract relevant data according to this perspective: since the interface is flexible, however, I could have also defined a set of "L1 (that is Italian) transfer tags" aiming at coding influence of the learners L1 on their L2 productions[14], which would be more suited for SLA research. Then I have imported the XML file containing the messages into the Blakkat interface, through a form allowing users to associate a given XML tag to a given field: the "message" tag was associated with the "text" field, etc. so that the system recognises this text as the text to be tagged.

FIGURE 1 Forum messages as XML data

I have also added a user id as extra arbitrary field, so that tagging data can be associated with a user and, therefore, an *ad personam* analysis is also possible. Now this small corpus is ready along with its own set of tags and can be coded.

---

[13] An interesting example of web forum analysis, especially as regards socio-constructivist theories, is in Ligorio (2005).

[14] Influence of L1 on SLA is one of the main topics of SLA research and has been differently interpreted according to the various approaches (see Chini, 2005 pp.54-59). It is not my intention, however, to draw any conclusion form the data collected in this case study– the analysis being inaccurate and unstructured: the only aim of this experiment is to show the possibility of integrating the activities described inside the teaching practice.

The set of tags for the "error analysis" is shown in table 1 and is a subset of the system defined by Nicholls (2003):

TABLE 1 The set of tags for the error coding

Error type	Error code/colour
Wrong verb spelling, e.g. *je parl* instead of *je parle*	<#FV>[15]
Wrong noun spelling, e.g. *francaise* instead of *français*	<#FN>
Wrong adjective spelling, e.g. *dificil* instead of *difficile*	<#FJ>
Wrong adverb spelling, e.g. *bientot* instead of *bientôt*	<#FY>
Something is missing, e.g. *Sont des chansons...?* instead of *Est-ce que ce sont des chansons...?*	<#M>
Word should be replaced, e.g. *difficile pour un principiant* instead of *difficile pour un débutant*	<#R>

Since the text is coded inside a form field it can be freely edited, in the case of error tagging, then, it is also possible for the research to follow the convention of adding a correction to the learner's error in the form <#CODE>wrong word/phrase | corrected word/phrase</#CODE> defined by Nicholls.

---

[15] As the tags are defined they must not contain the tag signs "<" and ">", since these are automtically added by the software.
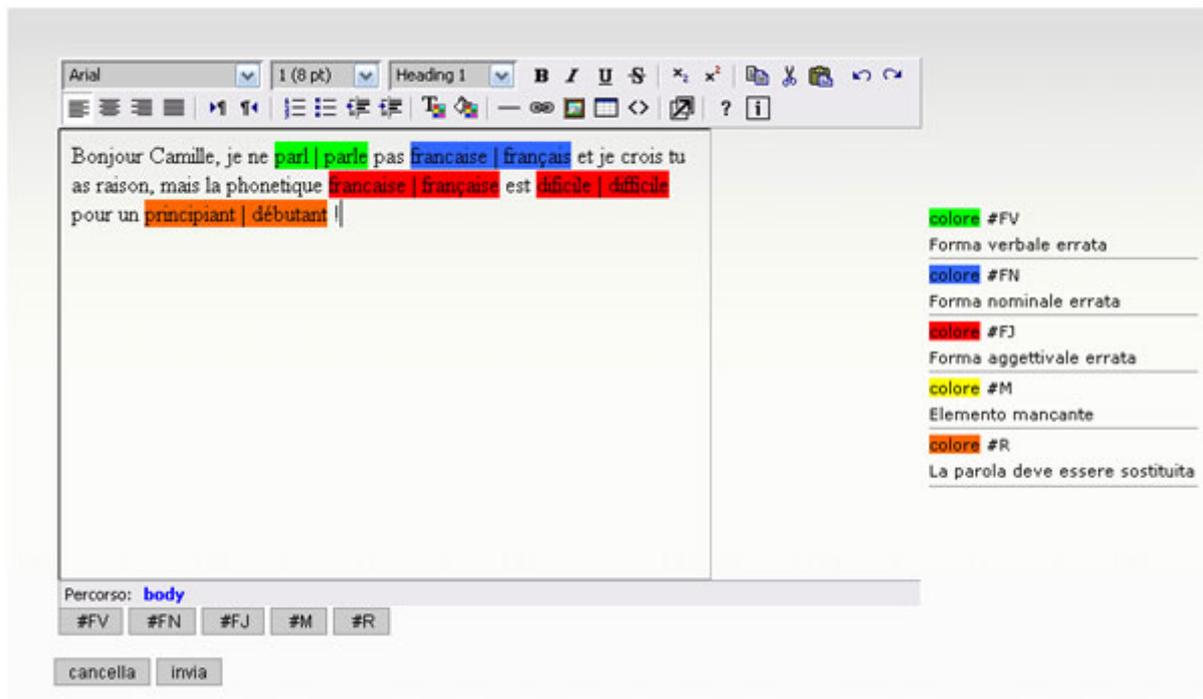
FIGURE 2 Blakkat: tagging window

Once the coding has been completed, results can be extracted through different filters, so that data relevant to a given issue can be analysed. For instance, it was found that:

- noun errors tend to occur more often than others (28.13%)
- user 3 is the most tagged: she made more errors if compared to others (31.25%), but also wrote more (9 messages), the written/tagged ratio is 2.2 that is an average value
- user 3 made more "adjective spelling errors" in the earlier stage of her activity
- the most common error is the wrong spelling of *bientôt*, which is differently misspelled
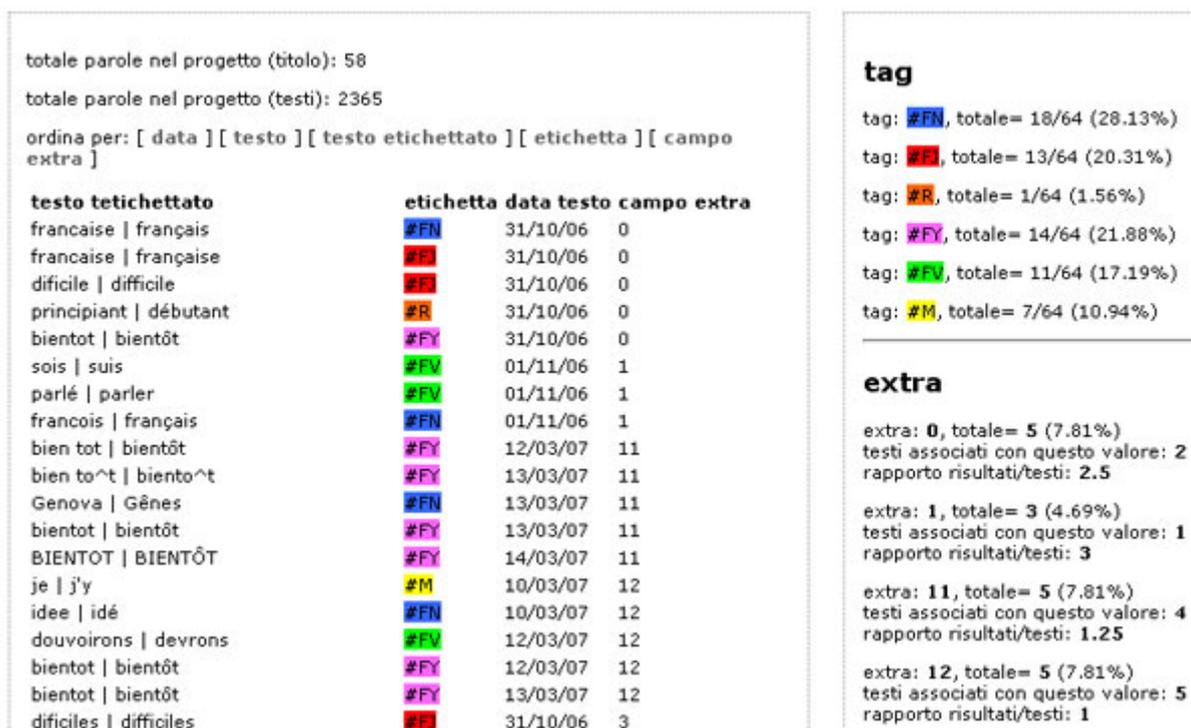
FIGURE 3 Blakkat: results

4.2. Exporting data to Wordsmith Tools

Learner Corpora are usually analysed through specific corpus software, such as Wordsmith Tools (Granger, 2004): this can handle larger amounts of data if compared to web-based applications, which suffer from severe limitations (see. Torsani, 2007), and offer a wide range of analysis tools. Another great advantage of Wordsmith is that it supports tagged corpora and, by consequence, allows user to upload tag lists to analyse their collections. This is especially the case with SLA research where, as noticed before, more time and resources are allocated to complete the task. We chose, therefore, to allow users to export the whole tagged corpus along with its own tag set in order be used inside such software as Wordsmith tools. Each corpus text is accompanied by a data description tag containing the text title, date and extra field, so that further complex extractions can take place, in the example an instance of error tagging can be seen:

<data text_title="msg_21" date="24/11/06" extra="9" />Bonjour a tout le monde, je m'appelle F.. Je viens de Arenzano, un petit village à coté de la mer. Je suis italienne et j'ai 23 ans. Je suis **<FN>etudiante | étudiante</FN>** et j'aime beaucoup pratiquer le ski et lire des livres. J'adore aussi les chats, j'ai 8 chats dans mon jardin. Et j'aime l'espagnol! Au revoir.

As is shown in Figure 3, Wordsmith tools allows users to examine the corpus tags and produce a KWIC[16] concordance, in this case noun spelling errors tags:

---

[16] A KWIC (Key Word In Context) is the very comon concordance output in which the key word is displayed in its context.

FIGURE 4 Exported tagged corpus analysed in Wordsmith tools

5. CONCLUSIONS

As with most software, it is not the tool itself which is important, but rather any possible implications and uses it may happen to have; this is especially true as regards linking with other software (as with Wordsmith tools) and, in the case of web based applications, as regards the integration inside learning environments. As we have seen Blakkat allows users to perform a number of tasks, but other software could do more: it is important, therefore, to create links with other applications in order to further widen the possible range of applications and deepen the level of analysis. But issues of language assessment are very important in the field of language teaching, especially as regards CASLA and NBLT. If such practices as those described above can take place inside the common SLA research framework, as noticed before, they seem hardly feasible in a teacher's perspective. Even if Blakkat is an open source and freely downloadable software, I would not expect a teacher collecting data and manually tagging large corpora of his students works to find out, for instance, that French beginners tend to misspell the word *français*, or mix up the French verb endings *–er,* eg. *parler* (inf.) > **parlé*. This is not a realistic perspective. But if we consider NBLT activities taking place inside a Virtual Learning Environment, where texts are produced as electronic text, and if we consider the possibility of transforming one such tool into a Moodle module, for instance, it could be effectively used as an assessment tool by teachers, especially as regards small individualised and localised text collections. As I have noticed elsewhere (Torsani, 2007), it is now time to force traditional web programming in order to transfer desktop programmes algorithms to web based applications, especially inside VLEs: the development of more than two hundred VLEs during the last few years (Devauchelle et Jarraud, 2006) illustrates how the need and demand for suitable tools is an important issue in Educational Technology, but illustrates also how much easy it is today, thanks to web server technologies, to build one's own "ideal tool".

Even if this practice does not fall in the mainstream of Computer-assisted assessment (CAT), defined as "testing practices requiring a computer to assist in construction, delivery,

response analysis and score reporting" (Chapelle, 2001, p. 38), it could prove extremely useful, especially as it could assess learners' output produced out of assessment-oriented practices, such as tests, and therefore more capable of viewing students "real" productions. Evidence collected in such a way, then, could influence syllabus design and provide evidence for global assessment and learning tracking. As it is shown in our case study, then, assessment issues of written language in VLEs, need not be conducted on large or even medium size corpora, thus making it a feasible task for teachers who chose online activities for their learners.

REFERENCES

1.   Adamson B. (2004). Fashions in Language Teaching Methodologies, in *The Handbook of Applied Linguistics*, 2004, ed. Davies & Elder, London: Blackwell Publishing pp. 604-619
2.   Calvani A., Fini A., Pettenati M., Sarti L. (2005), Teorie CSCL e piattaforme Open Source per l'e-learning: verso un'integrazione, in *Come costruire conoscenza in rete?*, 2005, eds. Delfino M., Manca S., Persico D., Sarti L., Ortona: Menabò, pp. 195-207
3.   Chapelle, C. (2001). *Computer Applications in Second Language Acquisition*, Cambridge: Cambridge U.P.
4.   Chiari, I. (2007). *Introduzione alla linguistica computazionale*, Bari: Laterza
5.   Chini, M. (2005). *Che cos'è la linguistica acquisizionale*, Roma: Carocci
6.   Chun, D., Plass J. (2000) Networked multimedia environments for second language acquisition, in *Network-based Language Teaching: Concepts and Practice*, 2000, eds. Warschauer M., Kern R., Cambridge: Cambridge U.P., pp- 151-170
7.   De Cock, S., Granger S. Petch-Tyson S. (2003): *The Louvain International Database of Spoken English Interlanguage – LINDSEI*, available at <http://www.lftr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Lindsei/lindsei.htm>,accessed 13 Jul 2003.
8.   Devauchelle B, Jarraud F. (2006). Les outils, situation actuelle et prospective, in *Comprendre les communautés virtuelles d'enseignants. Pratiques et recherches*, 2006, eds. Daele A., Charlier, B., Paris : L'Harmattan. pp. 203-209.
9.   Granger, S. (2002). A bird's eye view of learner corpus research, *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, ed. S. Granger, J. Hung & S. Petch-Tyson, 2002 . Amsterdam: John Benjamins. 3-33.
10.   Granger, S. (2004). Computer Learner Corpus Research: current status and future prospects, in *Applied Corpus Linguistics: A Multidimensional Perspective*, eds. Connor & Upton, Amsterdam: Rodopi pp. 123-145.
11.   Gruba P. (2004), Computer assisted language learning, in *The Handbook of Applied Linguistics*, 2004, eds. Davies & Elder, London: Blackwell Publishing pp. 623-642
12.   Hunston S. (2002). *Corpora in Applied Linguistics*, Cambridge: Cambridge U.P.
13.   Johns, T. (1997). Contexts: the Background, Development and Trialling of a Concordance-based CALL Program, in *Teaching and Language Corpora*, 1997, ed. Wichmann, Fliegelstone, McEnery, Knowles, Harlow: Longman, pp. 100-115
14.   Johns, T. (1991). "Should you be persuaded": two samples of data-driven learning materials, in *Classroom concordancing*, 1991, Birmingham: ERL Journal, University of Birmingham
15.   Leech, G. (1997). Teaching and Language Corpora: a Convergence, in *Teaching and Language Corpora*, 1997, ed. Wichmann, Fliegelstone, McEnery, Knowles, Harlow: Longman, pp. 1-23

16. Levy, M. (1997). *Computer-assisted language learning: context and contextualisatio*n. Oxford: Oxford University Press.

17. Ligorio M (2005). "Alberi e fili": la costruzione di conoscenza nel forum, , in *Come costruire conoscenza in rete?*, 2005, eds. Delfino M., Manca S., Persico D., Sarti L., Ortona: Menabò, pp. 147-160

18. Littlewood, W.(2004), Second Language Learning, in *The Handbook of Applied Linguistics*, 2004, ed. Davies & Elder, London: Blackwell Publishing pp. 501-520

19. Mukherjee J. (2006). Corpus linguistics and language pedagogy: The state of the art – and beyond, in *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, eds. Sabine Braun, Kurt Kohn & Joybrato Mukherjee. Frankfurt am Main: Peter Lang, 5-24. available at < http://www.uni-giessen.de/anglistik/LING/Staff/mukherjee/pdfs/Corpus%20Linguistics%20and%20L anguage%20Pedagogy.pdf >

20. Mukherjee J., Rohrbach J. (2006). Rethinking Applied Corpus Linguistics from a Language-pedagogical Perspective: New Departures in Learner Corpus Research, in *Planing, Gluing and Painting Corpora: Inside the Applied Corpus Linguist's Workshop*, ed. Bernhard Kettemann & Georg Marko. Frankfurt am Main: Peter Lang, 205-232., available at < http://www.uni-giessen.de/anglistik/LING/Staff/mukherjee/pdfs/11%20Mukherjee%20and%20Rohrba ch.pdf >

21. Nicholls, D. (2003). The Cambridge Learner Corpus - error coding and analysis for lexicography and ELT. In: Archer, D. , P. Rayson, A. Wilson and T. McEnery (eds) *Proceedings of the Corpus Linguistics 2003 conference* (2003), pp.572 -581.

22. Ortega L., Iberri-Shea G. (2005). Longitudinal research in second language acquisition: recent trends and future directions, in *Annual Review of Applied Linguistics* (2005) 25, 26–45. Cambridge: Cambridge U.P.

23. Ranieri M. (2005), Instructional Design: modelli e teorie, available at <http://www.formare.erickson.it/archivio/dicembre_05/1_RANIERI_01.html>,accesse d 02 Jul 2007.

24. Stubbs, M. (2004), Language Corpora, in *The Handbook of Applied Linguistics*, 2004, eds. Davies & Elder, London: Blackwell Publishing pp. 106-132

25. Torsani S. (2007), DDL en réseau: un exemple d'utilisation des technologies ajax pour améliorer les outils internet d'apprentissage des langues, in ISDM n°29 - *TICE MEDITERRANEE 2007*, available at <http://isdm.univ-tln.fr/articles/num_archives.htm#isdm29>, accessed 02 Jul 2007

26. Warschauer M., Kern R., Theory and practice of network-based language teaching, in Network-based Language Teaching: Concepts and Practice, 2000, eds. Warschauer M., Kern R., Cambridge: Cambridge U.P.

27. Wilson, E. (1997), The Automatic Generation of CALL exercises from General Corpora, in *Teaching and Language Corpora*, 1997, ed. Wichmann, Fliegelstone, McEnery, Knowles, Harlow: Longman, pp. 1-23