Cesare Zanca

Developing translation strategies and cultural awareness using corpora and the web

Abstract

Corpora, as we all know, can be extremely useful to translators and in the classroom. Translators can explore them to find lexical items, collocative patterns, phraseology etc. Teachers can use wordlists, concordances, keyness and frequencies to create exciting activities. Students can develop awareness and self confidence, autonomy and accuracy in L2 and translate more effectively, but introducing corpora in the classroom is not an easy task: most of the experiences described by research articles, books and online publications illustrate either advanced and highly motivated students who are able to collect and sometimes annotate their own corpus, or a presentation to the classroom of the results of a piece of research previously conducted by the teacher. The aim of the paper is to suggest an approach leading to inductive discovery of language patterns and usage through activities based on selected online resources which are more familiar to students but can nonetheless gradually develop autonomy and methodological soundness.

Introduction

Many studies in the last decades have addressed the potentiality and importance of corpora and corpus related activities for language learning and teaching and for translators.[1]

Sabine Braun, for instance, maintains that:

The insights into language use which corpora offer seem to leave no doubt that they hold great potential both as a resource for the creation o

• realistic, showing language in real use;
• rich, providing more (and more diversified) information than dictionaries or reference grammars can;
• illustrative, providing actual patterns of use instead of abstract explanations;
• up-to-date, revealing trends in language use and evidence for short-term historical change (as shown, for example, by Mair, Hundt, Leech & Smith, 2002).

But she adds that:

However, a careful look around the many different places where languages are learnt and taught makes clear that corpora, while being the 'buzzword' in language research departments, are still far from being part of mainstream teaching practice, if not terra incognita altogether. This raises the question why the real uptake of potentially so valuable a resource has been rather limited to date. Braun S., 2005, p.48

It is undeniable that the vast majority of the corpus based activities described in research articles, books and online publications are based on presentations to the classroom of the results (concordances, collocative patterns, phraseology etc.) of research previously conducted by the teacher and only when learners are advanced and highly motivated are they involved in direct access to corpora[2] or are requested to collect and even annotate their own corpus.

Mc Enery and Xiao, discuss the possible advantages in using corpora in the classroom in a direct (students' direct access to corpora) and indirect way (examples taken from corpora by teachers and used in the classroom). They argue that:

> The use of corpora in language teaching and learning has been more indirect than direct. This is perhaps because the direct use of corpora in language pedagogy is restricted by a number of factors including, for example, the level and experience of learners, time constraints, curricular requirements, knowledge and skills required of teachers for corpus analysis and pedagogical mediation, and the access to resources such as computers, and appropriate software tools and corpora, or a combination of these"
> Mc Enery and Xiao (2010), p.2

While the indirect approach is more feasible for teachers, because it is easier to 'teach about' and exploit to teach', only a direct access to corpora and linguistic data – a 'teaching to exploit' modality - can help students develop a genuine autonomy and an exploratory capacity, in line with the Data Driven Learning (DDL) approach (Johns 1991)[3] and its later developments. But autonomy is a difficult goal to achieve with a limited L2 competence and restrictions of the kind described above.

The textual and discourse skills involved in accessing and interpreting a corpus are another important issue. Following Widdowson, Braun underlines the distinction between *genuineness of texts* and *authenticity of discourse* claiming that:

> Real-language texts, therefore, are only useful insofar as the learner is able to authenticate them, ie to create a relationship to the texts.

> Discourse authentication is seen as a key element for learner motivation and eventually for learning success. Isolation from the original discourse is likely to make authentication more difficult but not impossible – as long as learners are able to construct an appropriate context in their mind. Context construction heavily relies on subjective knowledge. Therefore, each learner will construct their own individual context, and the construction process will be greatly facilitated if the topic is familiar to, and interesting for, the learner.
> Braun (2005) p.53

Effectiveness in autonomous exploitation of the *authentic* data obtained from corpus queries is largely determined by the learners ability to create an appropriate discoursive context in their minds and this process is facilitated by L2 knowledge but also by personal interest and degree of acquaintance with the topic. It should be noticed that discourse authentication, to a lesser extent and in particular circumstances, is problematic also for native speakers when coping with unfamiliar subject areas. As M. Hoey explains, our fluency and capacity in interpreting and using language in context depends on our previous linguistic *encounters*:

> There is a way of looking at language that explains the existence of combinations like the ones above[4] and accounts for native-speaker fluency. This view of language assumes that language users store the words they know in the context in which they were heard or read. According to this view, every time speakers encounter a word or phrase, they store it along with all the words that accompanied it and with a note of the kind of context it was found in – spoken / written, colloquial / formal, friendly / hostile, and so on. Bit by bit, they begin to build up a collection of examples of the word or phrase in its contexts, and subconsciously start to notice that these contexts have some pattern to them.
> Hoey (2009)

L2 learners, of course, have had limited encounters and limited occasions to store linguistic patterns in their minds together with the appropriate socio-cultural, contextual and discoursive information. The meaning of single words can be easily reconstructed, perhaps with the support of dictionaries or other traditional lexical resources, but the combinatory and contextually bound nature of language is much more problematic.

When facing this kind of difficulty learners can rely neither on traditional linguistic resources nor on their internalized 'previous encounters'. Their possibility of recreating an appropriate discoursive context is very narrow. In such circumstances, access to contextualized examples from corpora or similar resources can prove very useful. As Braun suggests, learners' familiarity with the contextual discourse can be improved through pedagogical mediation, use of corpora that are coherent, relevant, restricted in size, in a multimedia format and pedagogically annotated, combining corpus techniques with a discourse based analysis of materials and processes and the integration of pedagogically relevant materials.

But, as the previous reference to the limited number of successful experiences in introducing direct use of corpora in the classroom suggests, Braun's recommendations might not be enough. Indeed, this is what I came across in my teaching experience: in the attempt to use corpora in my university classrooms. Technical, time, curriculum and awareness issues are crucial. Students realize that an autonomous use of corpora can be very useful, but if we ask them to use corpora as part of their work, they object that requirements in terms of technical skills, time and methodology are too demanding.

Introducing corpora in the translation classroom

When L2 learners are engaged in translating authentic texts, their lack of 'previous encounters' emerges in an even more evident way and the contextual and discoursive gap in identifying combinatory features of language, together with socio-cultural elements can be even more frustrating.

As we all know, becoming a professional translator requires a long period of 'subconscious storing of linguistic patterns' which doesn't derive from a degree in linguistics. Lanna Castellano's path to professional translation is probably too demanding:

> "Our profession is based on knowledge and experience. It has the longest apprenticeship of any profession. Not until thirty do you start to be useful as a translator, not until fifty do you start to be in your prime.
> The first stage of the career pyramid – the apprenticeship stage – is the time we devote to investing in ourselves by acquiring knowledge and experience of life. Let me propose a life path: grandparents of different nationalities, a good school education in which you learn to read, write, spell, construe and love your own language. Then roam the world, make friends, see life.
> Go back to education, but to take a technical or commercial degree, not a language degree. Spend the rest of your twenties and your early thirties in the countries whose languages you speak, working in industry or commerce but not directly in languages. Never marry into your own nationality. Have your children. Then back to a postgraduate translation course. A staff job as a translator, and then go freelance. By which time you are forty and ready to begin".
> Castellano L. (1988) in Baker M. (1992) p.3

but she points out the necessity for a good translator to go through intense linguistic and personal experiences in order to become familiar with the source and target languages. Unfortunately, or luckily, our students cannot be asked to go through the same process: they have to deal with translation problems because that's what we ask them to do as part of their school or university work. As any other L2 learner, they have to make up for years of experience and *encounters* with our help and, of course, the tolerance for mistakes granted by their status. But they also need to have access, possibly autonomous access, to examples of contextualized discourse and thus can benefit from corpora and similar resources,

One important feature of translation is that it is an asynchronous activity and gives time to search for information, consider several translation candidates and think about different strategies. Another peculiarity of translation is that learners are usually native speakers of one of the two languages involved and this gives them both the ability to deal with texts in L1 with more confidence and the awareness of their limitations in expressing the same linguistic features in L2. Moreover, it is an activity in which accuracy is at least as important as communication and the form is as important as the message. A focus on form and accuracy, expertise in L1, awareness of one's own linguistic, cultural and contextual limits in L2 and time availability are elements that

facilitate significantly the chance of success in introducing corpora in the classrooms and to foster the learners' awareness of their potential value. At least in theory.

A new perspective

In practice, the initial impact with corpus work, software and interpretation of results is still troublesome. It requires a lot of theoretical, technical and practical work that learners perceive as disconnected from their course. This makes it almost impossible to include it in a syllabus that doesn't focus primarily on corpus linguistics or translation studies.

After several years in which I tried to introduce corpus linguistics and DDL activities in my courses, with controversial results, I seem to have obtained better outcomes with a different approach that initially relies on more familiar web resources and gradually directs students to more reliable information they can get from online lexicographic tools and corpora. The idea is to get to the interrogation of online corpora as part of a discovery process that confirms or denies hypotheses about language and as a means to help learners solve specific linguistic and translation problems and not as an end.

Search engines such as Google, blogs and online dictionaries are examples of these more familiar web resources. Students use and already know most of these tools and, instead of insisting on the drawbacks and risks in getting results from the web, teachers might exploit the similarities and the added value of the heap of information available. Google in particular, can be used to introduce corpus queries. Learners type in words (keywords) and get lists of results with those words in context. They soon understand that they can get better results if they search for series of connected words in an exact or loose sequence: a rudimental investigation into phraseology and collocative patterns. The data base, the criteria in information retrieval and the presentation of the context words are, of course, very different from corpora and concordances, but the process is similar and in a learning context this facilitates the approach to 'proper' corpus linguistics tools.[5] The growing familiarity with other resources, such as online dictionaries, thesauri, Wikipedia or other encyclopaedic information, blogs, discussion, forums and social networks can be exploited to create a bridge towards corpus interrogation.

Another important aspect of my approach is that activities are based on an online learning environment. McEnery & Xiao, Braun and many others, underline the importance of *pedagogical mediation'* and of teachers' guidance, but also of learners direct access to linguistic data. I think that a fruitful integration of the two necessities can be obtained using the structured activities typical of e-learning[6] methodologies and an online collaborative learning environment that provides learners not only with learning materials and tasks, but also with links to relevant online resources forums and other collaborative opportunities.[7] Students carry out their activities online and find on the same web page links to resources that are often integrated in their tasks and that can be 'simple' tools like Google or an online dictionary, 'nice' as a virtual thesaurus, but also more sophisticated, as an online corpus or a text analyzer. Autonomous and direct access to resources and linguistic data is easier and more natural.

Of course, when dealing with information and data from the web, a good deal of guidance and pedagogical mediation should be devoted to conveying awareness about the importance of assessing quality, relevance, and reliability of results. It is essential that learners, who often lack the background and linguistic competence to recognize misleading findings, acquire adequate validity control routines.

They also need to be guided in asking the right questions: web resources, exactly like traditional corpora, are not necessarily

expected to provide the right answers … but constantly presents new challenges and stimulates new questions, renewing the user's curiosity and offering ample opportunity for researching aspects of language and culture
Bernardini S. (2002) p.166

The web is an immense deposit of data and we only get reliable answers if we ask the right questions, formulate queries properly, analyse results and check them against other sources. Even so, very often the first results are inadequate or inappropriate and abstract concepts such as frequency, statistical significance, discoursive contexts and domains, collocative patterns, contextual hints, semantic prosodies etc. can come into play and be recognized by learners as very useful discriminating devices.

A case study

An example taken from the Moodle web page of my translation course[8] can perhaps better illustrate how a simple translation task can foster an autonomous and direct access to resources that include corpora. The activity is addressed to university students of English as a foreign language with a B2 level of English and is based on the translation of a short text from Italian into English. In the text learners find the popular Italian proverb 'Hai voluto la bicicletta? …adesso pedala!' which literally means 'You asked for a bike? Now pedal' and is used to mean something like 'You have to accept the consequences of your decisions'. The students are asked to think about possible translations into English (and the reader of this paper is invited to do the same in a L2 language). Normally they do not come up with acceptable translations because their previous linguistic encounters, in Hoey's terms, do not include many English proverbs[9]. They are usually aware of the fact that the proverb forms a single translation unit and should be translated with an equivalent proverb or idiomatical expression in English, but they realize very quickly that bilingual dictionaries are not of much use and do not know what to look for in English monolingual dictionaries of idioms or online collections of proverbs. Attempts with words like 'bike' or 'pedal', obviously, fail.

In this case Google proves more useful than most traditional resources in reducing the linguistic gap. Some students typed the whole expression in the search engine and found a web page with the translation of the Italian proverb in several European languages:

> "Hai voluto la bicicletta? Ora pedala!"
> …
> In Germania si deve "finire a cucchiaiate la zuppa in cui si è sbriciolato il pane" ("die

> Suppe auslöffeln, die man sich eingebrockt hat"); in Polonia forse va un po' meglio perché si deve "bere la birra che ci si è preparati" ("Pic piwo, którego sie nawazylo"). In Francia, i tedeschi hanno fortuna nella sfortuna: chi arriva dalla nazione del pumpernickel (il famoso pane nero tedesco, ndr) e viene minacciato, come conseguenza delle sue azioni, di dover "mangiare del pane integrale" (manger du pain noir), In Spagna si devono "pagare i piatti rotti" (pagar los platos rotos) e **in Gran Bretagna "portare la latta" ("to carry the can").**[10](my emphasis)

The English translation suggested is 'to carry the can', but is it really appropriate? The web page looks well constructed and reliable, but the activity requires a validation routine. Students are asked to check their results using online resources. In this case they are asked to use Mark Davies' online corpora: the BNC (British National Corpus), the Times corpus and the COCA (Corpus of Contemporary American English).[11] The expression 'carry the can' is not present in the Times corpus, it appears three times in the COCA corpus, but occurs 26 times in the BNC (Table 1).

Table 1 '*Carry the can*' in the BNC

From the context, it is possible to infer that 'carry the can'[12] means something like 'taking somebody else's responsibility' which is very different from what we are looking for.

The 'carry the can' option is then rejected, and we need another translation candidate. Some students found it in a discussion forum for translators: an English native speaker member suggests the expression: 'As you make your bed, so you must lie in it', adding that it is an old expression and might be obsolete (Table 2)

Thread Tools ▾ | Search this Thread ▾

9th November 2007, 01:26 PM   #1

friedfrog ○
Member

Join Date: Dec 2006
Native language: Italian, Milan
Posts: 60

hai voluto la bicicletta? e allora pedala

Ciao a tutti,
mi stavo chiedendo se ci fosse una frase idiomatica corrispondente all'italiano:
"L'hai voluta la bicicletta? e allora pedala".
O comunque una frase per sottolineare a qualcuno di non lamentarsi per le conseguenze che derivano da una scelta fatta.
ciao

Quote

9th November 2007, 01:31 PM   #2

Einstein ○
Senior Member

Join Date: Dec 2006
Location: Milano, Italia
Native language: UK, English
Age: 59
Posts: 6,833

Re: hai voluto la bicicletta? e allora pedala

An old expression is:
As you make your bed, so you must lie in it.

I'm sure there's something more modern, but it doesn't come to mind...

Quote

Table 2: a possible translation candidate found in a forum

The validating routine starts again, but the expression is not present in any of the corpora. The inexperienced user might be induced to reject it, but one of the students tries with only one part of the expression: 'lie in it'. The BNC offers only one result, not connected to the expression we are looking for, but in the COCA we find 33 occurrences (Table 3)

Table 3: 'lie in it' in the COCA corpus

The context indicates a correspondence of meaning and usage with the Italian proverb. At the same time students realize that the expression is more frequent in form which differs from the one suggested in the forum: 'you made your bed, now lie in it'. But why is the

expression absent in the British corpus?[13] Is it only used in American English? And is it really obsolete as suggested in the forum? In order to clear up these doubts students were suggested to go back to Google and type 'lie in it' in Google news. Limiting the query to the last month, students found 24 occurrences, all of them from American newspapers. The expression, then, is not obsolete, but might still belong only to American usage. Limiting the query to British newspapers ('lie in it' site:.uk in Google news) with no time restrictions, the students obtained many occurrences from national broadsheets and the BBC, suggesting that the expression is commonly used, even if probably less frequent in the UK than in the USA[14]. The translation candidate seems to be validated by online data driven queries that exploit corpora, but also other, more familiar and user-friendly resources. Learners were invited and guided to an autonomous exploration of these resources and to a systematic validation of possible inferences against other data. Among other findings, they also observed that by searching only one of the two parts of the English expression ('made your bed' or 'lie in it'), they are more likely to get results and to incur in unexpected and 'serendipitous'[15] findings typical of corpus-like investigations. For instance, they found different ways of using the expression, like 'let him lie in it', 'they're going to have to lie in it', 'it's his bed, let him lie in it' that sometimes were judged more suitable for the translation task. In one case, students discussed in the forum the puzzling expression 'You've buttered your bread, now lie in it'[16] and went back to search its components in corpora finding out that it was a mixture of 'You've buttered your bread, now eat it' and 'You've made your bed, now lie in it'. This process not only helps students understand a difficult linguistic item, but also offers a possible new candidate for the translation ('You've buttered your bread, now eat it') that, nevertheless, needs to go through the whole validation process.

Finally, the procedure described above has been only one of the possible paths followed by the students. Another resource some of them used is the parallel corpus OPUS (Opus Open Source Parallel Corpus) whose online interface offers access to several data bases, mainly collecting EU parliamentary acts and documents translated in many European languages[17]. The search for the Italian expression 'Hai voluto la bicicletta' produced no results, but the English candidate found through Google, 'You've made your bed', was present with one occurrence:

You have made your bed now you have to lie in it.

Italian Ve la siete voluta voi e adesso dovete subirne le conseguenze.
French Comme on fait son lit on se couche.
Spanish Usted se ha hecho su cama, y ahora tiene que acostarse en ella.
German Denn wie man sich bettet, so liegt man.
Swedish Som ni har bäddat, får ni nu ligga.
Portoguese Fizeram a cama? deitem-se nela agora.
Dutch U zult de gevolgen van uw daden onder ogen moeten zien.
Finnish Niin makaatte kuin petasitte.
Danish De har redt - nu må De ligge, som De har redt. Når min gruppe beslutter sig
- for enten hr.

I will not comment on the translation in other languages, but Italian students might be pleased in discovering that their translation work, based on a careful choice of information gathered online, produced a result that can be considered better than the solution offered by a professional translator, who opted for a plain description of the meaning and not for an equivalent idiomatic expression.

Conclusions

Most readers of this article probably agree on the fact that corpora and DDL activities have a great potential in L2 teaching and learning and that their impact is even more remarkable in translation studies and training. Access to contextualized examples of language in use helps learners to 'construct an appropriate context in their mind' thus reducing the lack of previous 'linguistic encounters' in L2. These examples can be introduced in classroom in an indirect way, to 'teach about' language, or in a direct way, as the result of learners explorations of corpora or corpus-like resources. The latter approach is probably more fruitful, fostering the learners' autonomy, inductive exploratory skills and awareness about L2, but its introduction and integration in ordinary teaching contexts is much more problematic and has produced a limited number of successful experiences. Teaching *about* corpus linguistics, corpus interrogation software, search options, methodology and analysis of results *before* students can even start using corpora on their own is probably too demanding. This article suggests a possible new perspective which inductively focuses on raising their awareness through experience rather than just increasing their knowledge by telling them what should be done. It draws on the fact that learners are nowadays more and more familiar with online resources and tools, such as search engines, forums, Wikipedia etc. which were not born for linguistic research but sometimes resemble the key-word-in-context query typical of corpus interrogation. These familiar resources might be used as springboards for more accurate and reliable queries, both on the web and in 'real' corpora that are becoming more and more available online. The suggested approach is an inductive, bottom-up, exploratory one that takes advantage of e-learning methodologies to engage L2 and translation learners, in structured activities based on an online learning space. This makes it easier for the students to find links to relevant resources and instructions on how to use them in a more effective way, to share their findings and to get individual feedback from the teacher and exchange opinions with fellow students.

Starting from practical and simple tasks, such as translating an idiomatical expression, finding out about the correct preposition, checking whether a word or a phrase is actually used with a particular meaning and in a particular context, learners gradually acquire confidence in their exploratory skills and might feel the need to learn more about corpus interrogation and corpus linguistics methodology.

Bibliography

Aston, Guy (1995) «Corpora in language pedagogy: matching theory and practice» in Cook G. and Seidlhofer B. (eds.) *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*, Oxford: Oxford University Press, pp. 257-270.

Aston, Guy (2000). «Learning English with the British National Corpus» In Battaner, M.P. C. López (eds.) *VI jornada de corpus lingüístics Barcelona*, Institut universitari de lingüística aplicada, Universitat Pompeu Fabra, pp.15-40.

Aston, Guy (ed.) (2001) «Learning with Corpora» Houston, TX: Athelstan.

Aston, Guy Bernardini, Silvia and Stewart, Dominique (eds.) (2004) «Corpora and Language Learners» Amsterdam: John Benjamins.

Bernardini, Silvia (2000). «Systematising serendipity: proposals for concordancing large corpora with language learners» in Burnard, L. and McEnery T. (eds.) *Rethinking language pedagogy from a corpus perspective*, Frankfurt: Lang, pp. 225-234.

Bernardini Silvia, Baroni, Marco and Evert, Stefan (2006) «A WaCky introduction» in Baroni M. Bernardini S. (eds.) *Wacky! Working papers on the Web as Corpus*, Bologna, Gedit, pp. 9-40.

Bernardini, Silvia (2000) «Systematising serendipity: Proposals for concordancing large corpora with language learners» in Burnard L. McEnery T. (eds.) *Rethinking language pedagogy from a corpus perspective,* Frankfurt am Main, Peter Lang, pp. 225-234

Braun, Sabine (2005), « From pedagogically relevant corpora to authentic language learning contents» in *ReCALL 17 (1)*, Cambridge University Press, pp. 47–64

Council of Europe (2001) «Common European Framework for Languages: Learning Teaching Assessment» Cambridge, Cambridge University Press. Also available online at: http://www.coe.int/t/dg4/linguistic/cadre_en.asp

Gavioli, Laura (2006) «Exploring Corpora for ESP Learning» Amsterdam, John Benjamins.

Gilquin, Gaëtanelle and Granger, Sylviane (in press). «How can DDL be used in language teaching?» in O'Keeffe A. and McCarthy M. (eds.) *The Routledge Handbook of Corpus Linguistics*, London, Routledge.

Hoey, Michael (2009), «Lexical Priming» Online publication available at: http://www.macmillandictionaries.com/med-magazine/January2009/52-LA-LexicalPriming.htm

Hundt, Marianne Nesselhauf, Nadja and Biewer, Carolin (2007) «*Corpus Linguistics and the Web*»Amsterdam, Rodopi BV.

Johns, Tim and King, Philip (eds.) (1991) «Classroom Concordancing» Birmingham, University of Birmingham.

Johns, T. (1991) «Should you be persuaded: two examples of data-driven learning. In *Classroom concordancing*»*ELR Journal*, 4 (special issue) Johns T., King P. (eds) 1-16.

Kettemann, Bernhard (1996) «On the use of concordancing in ELT» Online publication available at: http://cc.shu.edu.tw/~yjchen/Concordancing%20in%20ELT.doc

Kilgarriff, Adam and Grefenstette, Gregory (2003)«Introduction to the Special Issue on Web as Corpus» in *Computational Linguistics 29 (3)*, pp. 1-15. Also available online at: http://www.kilgarriff.co.uk/Publications/2003-KilgGrefenstette-WACIntro.pdf

McEnery, Anthony & Xiao, Zhonghua (2010) «What corpora can offer in language teaching and learning» in Hinkel E. (ed.) *Handbook of Research in Second Language Teaching and Learning (Vol. 2)*. London / New York, Routledge. Also available online at:
http://www.lancs.ac.uk/postgrad/xiaoz/papers/Corpora%20and%20language%20teaching v7.rtf

Tognini-Bonelli, Elena (2001) «Corpus Linguistics at Work» Amsterdam, John Benjamins.

Partington, Alan (1998) «Patterns and Meanings: Using Corpora for English Language Research and Teaching» Amsterdam, John Benjamins.

Scott, Mike and Tribble, Christopher (2006) «Textual Patterns: Key Words and Corpus Analysis in Language Education» Amsterdam, John Benjamins.

Zanca, Cesare Mesh, Linda (2005) «WebHelp: structured web resources and a blended approach for successful language learning»in Taylor C., Gotti M. (eds) *Atti del 3° Convegno Nazionale AICLU, "I Centri Linguistici Italiani: Approcci, progetti e strumenti per l'apprendimento e la valutazione" Vol. 2*, Trieste, Centro Stampa del Dipartimento di Scienze del Linguaggio, Interpretazione e Traduzione, Università di Trieste, pp. 603 – 617.

Zanca Cesare and Mesh, Linda (2006) «La formazione degli insegnanti online: favorire il cambiamento nei confronti dell'utilizzo delle tecnologie dentro e fuori la classe» in Cagliero R. (ed) *La funzione dei Centri Linguistici di Ateneo nel rapporto Scuola/Università*, Pescara, LUE, pp. 111 – 131.

Notes

[1] See among many others : Aston (1995, 2000, 2001, 2004), Gavioli, (2006), Johns (1991), Johns & King (1991), Kettemann (1996), Partington (1998), Scott & Tribble (2006), Tognini-Bonelli (2001).

² This also applies to studies on using the Web as a corpus, see for example Bernardini, Baroni & Evert (2006), Kilgarriff and Grefenstette (2003), Hundt, Nesselhauf and Biewer (2007)

³ See also Gilquin and Granger (in press).

⁴ Hoey refers to the fact that words «work together in predictable combinations», like the word *crazy* in *be crazy to do something,crazy about somebody*, *crazy about something*, *drive somebody crazy, go crazy, like crazy.*

⁵ Other google features can be explored with students: the number of results, the limitation of queries to specific domains (e.g. ac.uk, .edu, google news or google books), the visual information available through google images and even more sophisticated tools like google sets and google Ngram viewer.

⁶ See for example Zanca C. and Mesh L. (2005) (2006).

⁷ My courses include online materials activities and collaborative areas based on a Moodle learning environment.

⁸ It is a university undergraduate course for second year students of English Language anf Literature. My module consists of just 25 hours of contact time.

⁹ Proverbs and fixed phrases are an important part of a sociolinguistic competence, as recognized by the Common European framework of reference for languages (2001), pp. 118-120. Infact, fixed phrases make up a large part of texts.

¹⁰ http://www.cafebabel.it/article/25779/hai-voluto-la-bicicletta.html

¹¹ http://davies-linguistics.byu.edu/personal/

¹² Of course the meaning of 'to carry the can' could be found on a dictionary of idioms or a good monolingual dictionary, but corpora offer a better view of how the expression is used in context.

¹³ In fact, further queries reveal that the expression is present 5 times in the BNC with the preposition 'on' instead of 'in'.

¹⁴ An interesting way to check diachronic frequency of words and expressions in Books printed in British, American and other languages is google Ngram viewer: http://ngrams.googlelabs.com/

¹⁵ See Bernardini 2000

¹⁶ The expression is used in Disney's movie *Pinocchio*, but isn't really widespread