



HAL
open science

Beyond concordancing: Multiple affordances of corpora in university language degrees.

Alex Boulton

► **To cite this version:**

Alex Boulton. Beyond concordancing: Multiple affordances of corpora in university language degrees..
Procedia - Social and Behavioral Sciences, 2012, 34, pp.33-38. hal-00502633

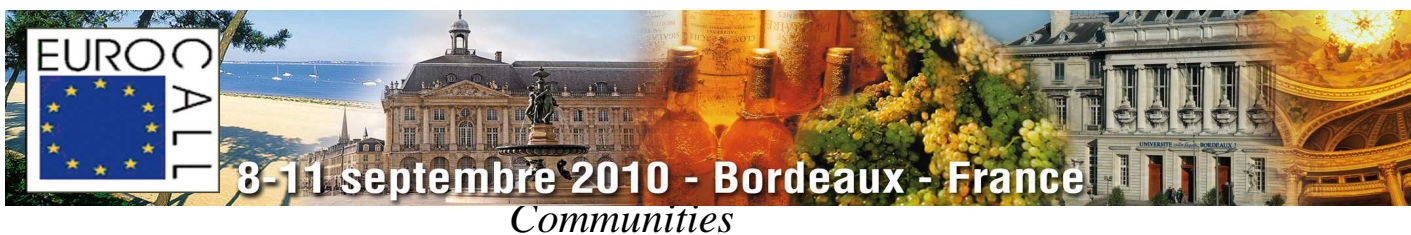
HAL Id: hal-00502633

<https://hal.science/hal-00502633>

Submitted on 21 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



*Université Bordeaux 2 - Victoire Campus
8-11 September 2010*

Beyond concordancing: multiple affordances of corpora in university language degrees

Alex Boulton*

CRAPEL-ATILF/CNRS, Nancy-Université, BP 3397, 54015 Nancy cedex, France

Abstract

Corpora have multiple uses for undergraduate students, including for translation. This paper explores how distance students cope with discovering corpora for this purpose for the first time. An analysis of their on-line exam papers reveals surprising complexity in their uses, leading to generally successful outcomes. Questionnaires show they lack confidence in using corpora, but see the relevance to this and other fields of study.

© 2011 EUROCALL All rights reserved

Keywords: corpora; translation; evaluation; data-driven learning; distance education; on-line

1. Introduction

Large electronic corpora have had considerable indirect influence in the field of language teaching and learning, especially in informing dictionaries, grammar books, usage manuals, syllabuses, test devices, even coursebooks. Their direct uses are largely associated with the work of Tim Johns, who coined the term data-driven learning or DDL in 1990 to introduce some of the things teachers and learners could do with corpora. However, it is frequently alleged (e.g. Turnbull & Burston 1998) that such use requires considerable training, which may be one factor explaining why DDL has caused much research interest but relatively little uptake in mainstream teaching around the world (Mukherjee 2006). As with any tool, the more uses that can be found for it, the more it becomes

* Contact author. Tel.: +33 (0) 354 505 109.
E-mail address: boulton@univ-nancy2.fr.

worthwhile investing in it (cf. Chambers 2007). In the case of corpora, much work has attempted to show the effectiveness of corpus use for different types of learners in different conditions for a range of language areas, from lexis to syntax to discourse (see Boulton 2010a for a survey of empirical studies). However, in the context of a university language degree, the possible uses go much further, potentially into all fields of study – not just language learning *per se*, but also linguistics, literature, civilisation and cultural studies, teacher training, and beyond (e.g. Boulton in press). If students can see the potential applications of language corpora and the tools and techniques associated with them to a number of areas of study, one might suppose they would be more open to spending time and energy in mastering them.

Corpora can also be used as a reference resource in translation, though the vast majority of research in this direction is concerned with post-graduate degrees for future translators (e.g. the papers in Zanettin *et al.* 2003; Beeby *et al.* 2009). Frankenberg-Garcia (2005), for example, found that her students used monolingual corpora more effectively than their preferred tools – bilingual dictionaries and term banks. Given that they had been using dictionaries for far longer, she concludes that:

The amount of training needed to use unmediated resources [mainly corpora] does not seem to be any greater than the amount of training needed to use dictionaries... In fact, had the students had as much exposure to unmediated resources as they had had to monolingual and bilingual dictionaries, it is possible that they would be much better at using the former (p. 352).

Despite such promising findings, there has so far been relatively little exploration of corpus use in translation at undergraduate level in language degrees where translation is just one activity among many. Translation in such contexts has had a bad press, but its ubiquity is not entirely irrational (cf. Zanettin 2009). Amongst other things, it is a common activity in professional life for many (IFOP 2009); it can promote learning, as well as increased language awareness of both L1 and L2; it is useful for language testing purposes, as it puts learners in a situation where they have to make language choices they might avoid in more general communicative contexts. Excessive use of translation may not be communicatively efficient, as the reaction against the Grammar-Translation Method suggests. However, simply attempting to ban all recourse to translation (as in the Direct Method and others) is also likely to be counter-productive: learners spontaneously resort to translation in many contexts, and depriving them of their L1 is to reject a major tool at their disposal. The choice is not between always-translate and never-translate; rather, it seems likely that some middle position may be most advantageous, encouraging intelligent uses of appropriate translation strategies and useful tools where they are most relevant and beneficial.

The students in the present study were enrolled for a distance degree in English in a French university; this includes translation in every year of study, and for logistical reasons the examination is no longer in a controlled environment: the students work at home, with the exam subject appearing on line at the appointed hour, and have a specified time limit to post their answers. Against this backdrop, a methodology section has been introduced for several reasons. Firstly, it limits the temptation of merely asking a competent bilingual or native speaker for help, as they would also need to have followed the course to complete the questions asked. Secondly, it seems useful for learners to think about what they are doing rather than just trusting to intuition, on the assumption that the metacognitive processes are likely to lead to increased autonomy and improved language skills (cf. Swain 2006 on “*languaging*”). Thirdly, it can be useful for the examiner not just to mark the translation product, but also to have some insight to the processes employed (Rodríguez Inés 2009). Finally, many of these students intend to go on to become teachers, and the new teaching qualification in France involves a methodology section as part of the translation exam.

In addition to helping with the immediate translation task, there are general theoretical reasons for thinking corpora can promote a variety of cognitive skills, and that using them within a constructivist problem-solving approach can promote autonomy and ultimately lead to life-long learning (Boulton 2010b). Given the practical limitations of the current teaching context, the introduction to corpus use is inevitably fairly light: a short theoretical background followed by demonstrations of particular functions, and activities accompanied by

feedback on translation-related questions; students were then encouraged to explore the corpora for their own questions, a practice which had been found fruitful for other work at the master's level in the same institution (Boulton in press). Against this background, the main research questions are thus:

- 1) What do these learners use corpora for?
- 2) Do these learners use corpora successfully?
- 3) How do these learners perceive corpus use?

2. Method

The data were collected from exam papers and questionnaires completed by a random sample of 12 students. The questionnaires were completed by email after the exam: all but one of the respondents was female; two did not have French as their L1; the average age was 35. Most other questions were in the form of statements on a 5-point Likert scale, from 1=strongly disagree to 5=strongly agree. The exam papers were those submitted after the second session of exams. Students had to translate a 250-word non-literary text, an interview with the French rock singer Renaud;* the corpus methodology questions asked them to choose any three short phrases and describe how they used a corpus to help in their translation of that segment.

The corpora chosen were the British National Corpus and the Corpus of Contemporary American English offered by Mark Davies at Brigham Young University (<http://corpus.byu.edu/>; see Davies 2009). These are relatively recent, general-purpose corpora, large enough to cover many text types but requiring sensitive querying. They have very similar interfaces (with the exception of PoS tags; CoCA also allows date queries), are relatively simple to use (use of drop-down windows; examples and demonstrations are also available on the site), and are available on-line (without downloading of corpus or concordancer) free of charge (registration is free). Unlike many commercial or copyright-protected corpora and concordancers, they are not limited to terminals on campus, so the students can consult them over the Internet from any computer around the world via a stable connection which is unlikely to crash, change interface, move site, or be removed from the web.

Analysis of the functions used in the students' papers showed that they generally used only one corpus, British or American according to taste, with only occasional comparison between them, though they sometimes compared uses in a particular sub-register (e.g. SPOKEN). They were primarily concerned with whether something existed, though frequency was one of the main concerns in justifying their choices. Frequency alone was not enough: they generally perused a selection of concordances to check usage in context, with one student for example finding that the only three occurrences were unconvincing for her particular query. They made extensive use of the asterisk as a wild card (for a single word or part of a word), and especially of the PART-OF-SPEECH function to look for particular types of items. They were largely interested in words that co-occur; in addition to the simple queries, they used the COLLOCATES function to check words within a certain span left or right, and occasionally the COMPARE function which highlights different collocates. The main features they did not use included: ordering concordance lines randomly; consulting the expanded context; using the lemma function; searching by synonyms; using the chronological data in CoCA. Of course, the fact that students did not report using these features does not mean they do not know how to use them, merely that they saw no reason to use them (or to report using them) in this short section of translation.

Given the space constraints here, two short examples will suffice.

* Available at:
http://www.lamontagne.fr/editions_locales/clermont_ferrand/entretien_avec_renaud_pour_la_sortie_de_son_nouvel_album_molly_malone_@CARGNjFdJSsHHhIACxg-.html

- 1) Source: Je suis paralysé entre le brûlot et la chanson d'amour...
- Translation: revolutionary songs
 - Explanation: I first thought of *pamphlet* for *brûlot*. I checked by typing *adjectives all + songs* and *nouns all + songs*. *Pamphlet* was not an option so I chose *revolutionary* instead.

Comment: With careless dictionary use, some students translated *brûlot* as *fireship*, *pamphlet* or *gnat*. Here it is a type of song that contrasts with love song, and Renaud is indeed known for his revolutionary songs / rebel songs / protest songs – all of which can be found in CoCA using this student's strategy.

- 2) Source: Tous les thèmes d'un prochain album, donc ?
- Translation: a forthcoming album
 - Explanation: In the LIST box I entered *album.[n*]* and in the COLLOCATES box I put *[aj*]*... For position I chose 1 and 0 because I was looking for an adjective that immediately preceded the noun. The most frequent results were *new*, *live* and *latest* which correspond to something else, followed in fourth place by *forthcoming*...

Comment: Many students translated word-for-word as *a next album*; while *a next [nn*]* does occur in both corpora, it is relatively infrequent and analysis of the concordance lines shows a quite specific usage (e.g. *There won't be a next time*). This student used the corpus to find an appropriate equivalent which is absent in many bilingual dictionaries.

In contrast with these encouraging results, in the questionnaire the students did not report finding the corpora easy to use ($M=2.1$) or think they could use them effectively (2.0). Nevertheless, they were sufficiently stimulated by the experience to want to use corpora in the future 'even if I don't have to' (3.7). They found them useful for translation (4.1), even more so for 'other aspects of learning English – vocabulary, grammar, style, etc.' (4.4). They were also able to see their uses for 'other courses – literature, civilisation, linguistics, etc.' (3.8), and even for fields outside their studies (3.9). Averaging the results like this conceals individual differences of opinion: insofar as these questions are indicative of students' receptiveness to corpus use as a whole, two were clearly "refusers" (averaging 2.0 and 2.3); two were keen "adopters" (4.4 and 4.6), to borrow Kaszubski's (2008: 174) terminology; the others scored between 2.9 and 4.0. An ANOVA test shows these differences to be statistically significant ($p<0.001$).

The students clearly would have liked further preparation in corpus use, in the form of more 'theoretical background' (3.2), more 'methodological explanation' (3.9), and especially 'demonstrations' (4.3) and 'concrete exercises to practise on my own' (4.2). That said, all of these featured in the course, and some of the open comments reveal that some students had not studied them very attentively if at all – explicitly requesting things which were already there. Further, the lack of training seems to reflect a lack of confidence rather than reality, inasmuch as the processes they described and the translations arrived at were generally successful: corpora can be of immediate use even with limited training (cf. Boulton 2009).

The open questions asked them what they liked and disliked most about using corpora. Favourable comments included discovering for themselves, the number of 'examples' available, the specificity of the queries, the quick and easy on-line consultation, usage of individual words and expressions, collocates, register use, and so on. Causes of concern focused on the complexity of the interface (the functions and the query syntax) and the time it took to conduct some queries. As one student (neither an adopter or a refuser) put it:

It might seem stupid and lazy (I probably should have persevered more at the beginning), but I find they're really complicated tools, the explanations are too theoretical and it's difficult to learn on your own on line. Ideally I'd like a separate course devoted to corpora, which are really wonderful when you know how to use them.

3. Conclusions

Perhaps the most striking feature of the students' queries is the sheer diversity: they are certainly doing more than just concordancing, combining features to produce some highly complex and at times quite sophisticated queries. Despite (or because of) the limited training and experience, many show considerable insight and creativity in some ingenious searches to make the most of what they are able to do. Furthermore, they are largely successful in their outcomes, with the occasional problems usually due to stopping their queries too soon. While not all are excessively enthusiastic, the perceptions are generally positive, with only 2 students consistently rating them negatively. They particularly appreciate the large number of examples and the fact they can tailor them to closely mirror the specific case at hand; the negative aspects include the time required and the complexity of the interface. They see other uses and would appreciate more training, but most say they intend to use corpora in the future on their own initiative.

The small sample in the present study needs extending, though results so far are promising in terms of what learners do with corpora for translation, the successes they have, and their perceptions overall. Inevitably further training would be desirable, especially if learners are to extend corpus use to other areas of study where they do see the potential applications. Yet however much training or help is provided, there will inevitably remain the occasional dissenter: it seems unlikely that any tool, technique or approach will appeal to every student, but this is clearly not a case for abandoning all innovation, including the use of corpora.

4. References

- Beeby, A., Rodríguez Inés, P. & Sánchez-Gijón, P. (Eds.). (2009). *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. Amsterdam: John Benjamins
- Boulton, A. (2009). Testing the limits of data-driven learning: language proficiency and training. *ReCALL*, 21/1, 37–51.
- Boulton, A. (2010a). Learning outcomes from corpus consultation. In M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez (Eds.) *Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching*. London: Equinox.
- Boulton, A. (2010b). Data-driven learning: taking the computer out of the equation. *Language Learning*, 60/3, 534–572.
- Boulton, A. (in press). Bringing corpora to the masses: free and easy tools for language learning. In N. Kübler (Ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice*. Bern: Peter Lang.
- Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda & J. Santana (Eds.), *Corpora in the Foreign Language Classroom* (pp. 3–16). Amsterdam: Rodopi.
- Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990-2008+): design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14/2, 159–188.
- Frankenberg-Garcia, A. (2005). A peek into what today's language learners as researchers actually do. *International Journal of Lexicography*, 18/3, 335–355.
- Institut Français d'Opinion Public. (2009). *Le Multilinguisme dans l'Entreprise*. <http://www.ifop.com/europe/docs/multilinguismeentreprise.pdf>

Johns, T. (1990). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14–34.

Kaszubski, P. (2008). A guided collaboration tool for online concordancing with EFL EAP learners. In A. Frankenberg-Garcia (Ed.), *Proceedings of the 8th Teaching and Language Corpora Conference* (pp. 167–175). Lisbon: ISLA-Lisboa.

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: the state of the art – and beyond. In S. Braun, K. Kohn & J. Mukherjee (Eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods* (pp. 5–24). Frankfurt: Peter Lang.

Rodríguez Inés, P. (2009). Evaluating the process and not just the product when using corpora in translator education. In A. Beeby, P. Rodríguez Inés & P. Sánchez-Gijón (Eds.), *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate* (pp. 129–149). Amsterdam: John Benjamins.

Swain, M. (2006). Linguaging, agency and collaboration in advanced language proficiency. In H. Byrnes (Ed.), *Advanced Language Learning: The Contribution of Halliday and Vygotsky* (pp. 95–108). London: Continuum.

Turnbull, J. & Burston, J. (1998). Towards independent concordance work for students: lessons from a case study. *ON-CALL*, 12/2, 10–21.

Zanettin, F. (2009). Corpus-based translation activities for language learners. *Interpreter and Translator Trainer*, 3/2, 209–224.

Zanettin, F., Bernardini, S. & Stewart, D. (Eds.) (2003). *Corpora in Translator Education*. Manchester: St Jerome.