

Raising teachers' awareness to corpora

Ana Frankenberg-Garcia

Instituto Superior de Línguas e Administração, Lisboa - Portugal

Keywords: corpora, data-driven learning, user behaviour, teacher training, consciousness-raising

1. Introduction

The last decade and a half has seen a dramatic increase in corpus availability and a steady growth in the number of supporters of the use of corpora in language teaching. The very fact that TaLC is at its seventh edition only confirms that the applied use of corpora in the language classroom is here to stay. Yet there seems to be a long way to go before corpora can be understood and used by language teachers in general. This paper examines some of the problems inexperienced corpus users encounter on their first contact with corpora and proposes a task-based, consciousness-raising approach to help them overcome initial difficulties in selecting corpora appropriate to their needs, formulating corpus queries and interpreting corpus data. The methodology is not corpus-specific, and should ideally be tried out on different types of corpora.

2. Language teachers' use of corpora

Many teachers (especially in the EFL context) have been using corpus-based dictionaries, grammars and text books for some time now without actually knowing what a corpus is. Indeed, understanding corpora and how to use them is not strictly necessary when the language teaching community resorts to corpus-based materials mediated by experts. The picture changes when this public is encouraged to use corpora and concordances hands-on, in pedagogical activities such as those put forward in Johns and King (1991), Tribble and Jones (1997), Aston (2001) and Frankenberg-Garcia (2005a), to cite just a few. Despite the enthusiastic acceptance of a small number of language teachers who have managed to break the barrier of putting their hands on corpora, an e-mail survey carried out by Tribble (2001) suggests that very few have actually done so*. The main reasons why many of the respondents did not use corpora were that they didn't have access to the appropriate software, didn't have enough knowledge of the potential of corpora, didn't have enough time to develop corpus materials and were not sufficiently confident in using computers to analyse language. Most significantly, however, over 50% of the respondents who did not use corpora did not (or could not?) answer why. Although there are more and more people with access to computers and the Internet, and the number of free, online texts and corpora has seen a steady increase, a study carried out by Mukherjee (2004) in German secondary schools indicates that the gap between the direct use of corpora and the reality of language teaching is probably even wider than one might have thought.

* Tribble's (2001) survey was circulated on the Linguist List and on the Corpora List, and even though the readers of these lists are far more likely to know about corpora than the average language teacher, only 52.8% of the 89 respondents declared they had used corpora in their teaching.

3. Novice user behaviour

With a limited number of teachers using corpora, it comes as no surprise that there do not seem to be any studies of this kind user behaviour. Some of the difficulties novice corpus users encounter are however described in Bernardini (2000), Kennedy and Miceli (2001), Frankenberg-Garcia (2005b) and Santos & Frankenberg-Garcia (in preparation). Although these studies differ quite substantially among themselves, they all converge to suggest that corpus skills which come as second nature to experts are not obvious at all to the untrained. Apart from corpus-specific difficulties in handling different search interfaces and CQLs – and the human-computer interaction issue should not be overlooked – these studies bring to light a number of very basic problems that novice users encounter no matter which corpus they use. These include:

- a. not knowing how to choose between different types of corpora (e.g. using a fiction corpus to research highly specific terminology; not using sub-corpora)
- b. difficulties in formulating corpus queries (e.g. queries too general or too restrictive; absence of follow-up queries)
- c. a limited ability to interpret corpus data (e.g. not taking corpus size and composition into account; no experience of dealing with unedited language).

4. Developing corpus awareness

Findings such as the above suggest that language teachers who are new to corpora may find it difficult to grasp that corpora do not work in the same way as the familiar language learning resources – such as dictionaries, grammar books and text books – that they are accustomed to using. Although most corpora provide detailed information on their composition, non-experts cannot be expected to understand the full significance of this without being explicitly taught how to do so. To begin with, there is the question of size. While most people are aware of the differences between large and small dictionaries, the implications of using corpora with 100 thousand words and corpora with 100 million words are not obvious to everyone. Awareness that different corpora use different criteria for text selection is also important. Teachers are familiar with different types of dictionaries, but the implications of the differences between corpora of newspapers, fiction, business letters, parliamentary debates and general language may not be immediately visible to them. Likewise, the distinctions between corpora of spoken and written texts or between monolingual and multilingual corpora can be just as elusive.

The ability to devise reasonable corpus queries is also something that cannot be taken for granted. Language teachers who feel comfortable looking things up in dictionaries and are familiar with web browsing do not necessarily know that the principles underlying corpus searches are different. They may need help in developing the basic skills required to formulate meaningful and plausible queries.

Finally, there is the question of authenticity. Teachers who are used to dealing with language learning materials that have been carefully edited cannot be assumed to know how to handle data samples containing language mistakes and idiosyncrasies. As corpus output is very different from the polished materials normally used in the classroom, teachers might find it difficult to decode the results of their corpus searches.

I conclude by proposing a series of consciousness-raising exercises aimed at helping language teachers gauge different corpora and discern which ones are best suited to their purposes, develop basic corpus-searching strategies, and get used to interpreting corpus data. The exercises are task-based and, unlike most corpus tutorials available, they are not corpus-specific. The overall idea is not to train corpus linguists, but simply to encourage teachers to become more confident about using corpora in the classroom.

References

Aston, G. (ed.) (2001) *Learning with corpora*. Houston: Athelstan.

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

Frankenberg-Garcia, A. (2005a) Pedagogical Uses of Monolingual and Parallel Concordances. *ELT Journal*, June 2005, 59/3, 189-198.

Frankenberg-Garcia, A. (2005b) A Peek into What Today's Language Learners as Researchers Actually Do. *International Journal of Lexicography*, 18/3, 335-355.

Johns, T. & P. King (eds.). (1991) *Classroom Concordancing*. Birmingham: The University of Birmingham Centre for English Language Studies.

Kennedy, C. & T. Miceli (2001) An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5/3, 77-90.

Mukherjee, J. (2004) Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. Upton (eds) *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi, 239-250.

Santos, D. & A. Frankenberg-Garcia (in preparation) "The corpus, its users and their needs: a user-oriented evaluation of COMPARA.

Tribble (2001) Corpora and teaching: adjusting the gaze. Paper presented at the *ICAME 2001* conference, Louvain, Belgium.

Tribble, C. & G. Jones. (1997) *Concordancing in the classroom: a resource guide for teachers*. Houston: Athelstan.