

Title: Evaluating the use of corpus-based instruction in a language teacher education context: perspectives from the users

Author: Fiona Farr

Address: Department of Languages and Cultural Studies
University of Limerick
Limerick
Ireland

E-mail: Fiona.Farr@ul.ie

Short title: Corpus instruction in language teacher education

Words: 6600

Evaluating the use of corpus-based instruction in a language teacher education context: perspectives from the users

F. Farr, Forthcoming, *Language Awareness*

Abstract

A recent practice in the study of language on teacher education programmes has been the use of electronic corpora and we are therefore still at the initial stages of exploring key issues relating to their integration. Despite arguments for and against their adaptation, there is a dearth of evaluative research examining student teachers' perceptions of learning and teaching through corpus-based activities. This paper has two main foci. Firstly, it reports some of the ways in which corpora have been incorporated into a language systems module on an MA in English Language Teaching programme over a two-year period. More significantly, it outlines the findings from survey results, which uncover student teachers' perspectives on their experiences of using corpora. Additionally, it explores the potentials and problems foreseen by these practitioners in relation to using such an approach in their careers. The investigation of these pertinent issues with a participant group of 25 student teachers leads to the conclusion that there is generally a positive pre-disposition towards the use of corpora. These attitudes vary in relation to the projected adaptation in EL teaching and the results also show that the real teaching scenario often does not permit the ideal of full application.

Introduction: Language Teacher Education and the Place of Technology

Language teacher education programmes (LTE) available through both public and private sectors have been shaped in two critical ways, giving us a contemporary template for both their content and their mode of delivery. The first relates to changing educational philosophies and the second to the impact technology has had on all facets of our lives, including education, over the past thirty years. Each of these is relevant to present discussions and will be elaborated on in this introduction by way of contextualising the content of this paper, which evaluates the integration of corpus-based techniques in English LTE for language awareness raising purposes and the ultimate degree of diffusion this may have to the English language teaching (ELT) classroom.

Firstly, changing educational theories have inevitably impacted on practices in LTE. Many of us are now fully aware of how social constructivism and the work of Vygotsky (for example, Bruner, 1985; Vygotsky, 1978) have had some effect in western educational approaches, even in the last fifteen years. Traditional philosophies have changed somewhat, and Kozulin (1998) frames these changes in the comparison of what he and others (for example, Kinginger, 2002) discuss as retrospective versus prospective education (the terms progressive and regressive are sometimes used to refer to the same or similar concepts).

Traditional education is retrospective because teachers teach students how to reproduce already known answers to previously posed questions and how to use skills that have proved to be useful in the past. Under the dynamic conditions of modernity, the necessity of prospective education becomes obvious. Prospective education implies that students should be capable of approaching problems that do not exist at the moment of their learning. (Kozulin, 1998: 6-7)

The shift has been from acquiring information to cognitive development, from reproductive to productive knowledge, and this can be seen clearly in the changing modes of learning in educational institutions. Traditionally, knowledge has been delivered through transmitting pre-determined cultural norms and traditions that could be reproduced by students when solving problems. Prospective approaches imply that the body of knowledge should be creative so that students can deal with problems that do not already exist at the time of learning (Halliwell, 1993). Notions and practices of

co-construction and co-authoring become paramount within this developmental paradigm.

Under the retrospective model (also referred to as 'banking education' in the US tradition, see for example, Crookes, 1997: 74), material is a body of knowledge with fixed rules for its application, which are deemed necessary for an individual to become educated. The material is transmitted and the cognitive capabilities required of students for the effective processing of the information are assumed. The teacher has three roles: to deliver content, evaluate student performance, and serve as a model of an accomplished, educated individual. Consequently, students are required to learn the material and display knowledge on request. They are expected to possess normative cognitive characteristics required for these tasks, allowing them to fit into the institutionalised organisation of the school. These roles and relationships are reborn within a socio culturally shaped progressive educational philosophy. The focus moves to the on-going and future development of students' cognitive strategies, which are dependent on socio cultural forces. The learners' role is an active one emanating from two underlying theoretical assumptions (Kozulin, 1998: 157). Firstly, Piaget's recognition of a child's need for a stimulating, problem-solving environment so that natural curiosity and ability to discover can be optimised. And secondly, Vygotsky's assertions that the child as an independent learner is the outcome and not the starting point of the educational process (for a fuller relative discussion of Piagetian and Vygotskian views see Nyikos & Hashimoto, 1997: 506-507, and Kozulin, 1998: Chapter 2). The learner is imbued with responsibility to co-construct knowledge in an environment which is stimulating, interesting, playful, and interpersonal. This process is supported by the teacher, and others, in what has been termed the guided construction of knowledge (Mercer, 1995). Additionally, the diversity of learner profile is recognised as a product of differences in educational background, linguistic and cultural histories, and specific needs, abilities and preferences. This is evidenced by the already-existing, and growing body of literature on individual learner styles, needs and other variables (for example Oxford, 1996; Revell & Norman, 1997). Also, learning materials reflect the emerging learner roles and within the prospective framework are designed, not to contain specific, pre-determined, informational norms, but to support learning in the form of more creative, interactive, problem-based activities, which recognise the importance of cognitive development through

education. The notion of language awareness (as discussed below) as opposed to the more traditional language or content knowledge is very much in fitting with these generally philosophical and practical changes in LTE. Of course, the reality of contemporary education is most often a combination of the more traditional modes of delivery with the newer discovery and supportive models, and indeed it is very likely that such complementarity works well in most contexts.

Secondly, let us return to the issue of technology and its impact on modern day educational society. Many commentators (for example, Chapelle, 2001; Cummins, 2000) have, for some time, spoken of the digital space in which we live. We now find ourselves in a post-industrial era based on the management of information at a highly technical level. As I have commented elsewhere, ‘Literacy is no longer just about reading and writing. Society now demands “multi-literacies” (Warschauer, 2000), which include a high proficiency in digital and on-line competencies (see also Doering & Beach, 2002; Pennington, 2001 for related discussions)’ (O’Keeffe & Farr, 2003: 389). Transferring this notion of recalibration to language study, conventional consultation, assimilation, and production skills developed through paper-based trails should be expanded to encompass the use of electronically-based language resources. Consequently, teachers need to have the necessary technical capabilities and these should be acquired or honed during their formal LTE. It logically follows that language teacher educators have a fundamental obligation in this respect. In addition, it has been argued and documented that promoting critical attitudes and developing conceptual as well as practical frameworks as regards the potential role of technology for language learning is the key to meaningful future integration (Egbert *et al.*, 2002; Meskill *et al.*, 2002). This will help provide a strong cognitive basis for one of the most crucial roles of language teachers in today’s climate, that of life-long learner.

Despite such support and encouragement for the adaptation of ICT to language contexts it would be remiss not to show an awareness of ensuing complexities. Teachers have many concerns about the use of ICT: scarcity of computers; limited access to networked machines; inadequate levels of technical support in schools; time pressures; lack of resources and materials; dearth of direction and guidelines as well as inadequate training (for example, Murray, 1998). Conversely, some authors suggest that, despite the many incentives and opportunities afforded to teachers in more

privileged environments to pioneer the use of technology, a surprising degree of resistance remains (Cuban, 2001). In addition, there are many risks associated with blind adaptation and integration. Issues of access and credibility, knowledge control and moderation, criticality, misinformation and censorship, privacy, and community, may be relevant. However such complexities serve to strengthen the argument that critical technological awareness and skills development be included in initial teacher education programmes in order to ‘future-proof’ (Barnes & Murray, 1999: 178) the teachers of tomorrow. How can instructors astutely tackle such issues and dilemmas without first-hand appreciation and control of what they might be?

Through the inclusion of ICT in initial teacher education programmes we also help trainees advance their general computer abilities and extend their methodological skills. Potentially, there are also affective benefits ensuing from the renewed energy, positive attitude, increased confidence (Egbert *et al.*, 2002: 110) and teacher empowerment (Tammelin, 2001) experienced upon successful mastery. It has been concluded that ‘it is primarily through active participation with technology as opposed to receiving instruction about technology that pre-service teachers learn to recognize the value of technology tools’ (Doering & Beach, 2002: 128). Such training and application often aids the promotion of self and semi-autonomous learning in a freer learning environment (Benson, 2001). Though evident, it is pertinent to be reminded that there are advantages and disadvantages associated with all methods, all materials and pedagogic apparatus, and computer technology is no exception. This has not and will not detract from the need for its integration in language teacher education programmes, this means instruction in on-line resources, a range of CALL software, multi-media programmes, generic packages, and language corpora. It is against this double backdrop of progressive educational practices and the use of ICT for education that this paper discusses and evaluates how corpus technology has been integrated and evaluated as part of the language awareness component of an MA in ELT programme over a two-year period. Before proceeding to the empirical study and results, the next section will discuss the place that corpus linguistics has played in the study of language and in the promotion of language awareness in classroom contexts to date.

Corpus Linguistics and Language Awareness

We now have access to large banks of language data in the form of language corpora, which can be analysed automatically using the appropriate software (see for example, Leech, 1991; McEnery *et al.*, 2006). In an earlier article, we have outlined various definitions of corpus linguistics and some of the important considerations to take into account when considering their potential applications (O'Keeffe & Farr, 2003), and so I will not repeat these issues in detail here, but summarise and expand on those most relevant. Corpus linguistics is an approach (for a discussion of the methodological status of CL, see Tognini-Bonelli, 2001) which can be, and has been, used in many disciplines, for example, dialectology, lexicography, sociolinguistics, language materials development, language therapies, speech technology, forensic linguistics, literary studies, language change and evolution, and grammar research. Generally, there are three main stages when using this tool: extraction of data from texts, processing the output (reshaping according to your needs), and interpretation of output (asking the right questions). Corpus software has a number of advantages for users: it produces word lists and counts occurrences of individual search items, it allows for the presentation and (re)organisation of data in a way that facilitates the identification of patterns, it automatically produces cluster and collocation lists, and most software has a keyword tool, which allows a comparison of lexis between corpora to identify relatively significant items. In this way it becomes easier to comment on the collocation, colligation, semantic preference (Sinclair, 1996), and semantic prosody (Louw, 1993) of lexico-grammatical items. A useful distinction, for both teaching and research purposes, is between a 'corpus-driven' approach and a 'corpus-based' approach (Tognini-Bonelli, 2001, see also McCarthy, 1998). A corpus-driven approach implies no predispositions, hypotheses, or prejudices before searching the data, approaching it without any frameworks, categories or taxonomies, but allowing the corpus to dictate the formation of same, and is the approach favoured by Sinclair, Tognini-Bonelli and also those working in a purely inductive data-driven learning tradition (Johns, 1991). A corpus-based approach, on the other hand, 'treats the corpus as an adjunct to the development of theoretical positions' (Butler, 2002:2), 'the philosophy and ideas are taken

for granted beforehand, and the corpus is simply used to reinforce those ideas' (McCarthy, 1998:22). This is the stance taken by those who use corpora to test theoretical frameworks which had been formulated without data (Harris, 1995:120), and those who utilise theoretical frameworks to aid in the interpretation of the corpus data. This often, and incidentally, produces a modification of such frameworks or the production of entirely new ones.

Two broad approaches have developed within the field of corpus linguistics.

McCarthy *et al.*, (2002:70), exemplify as follows,

Broadly, corpus linguistics may be performed in two ways: quantitative and qualitative. The quantitative approach usually looks for the largest corpus possible [.....] from as wide a range of sources as possible. These data are then analysed computationally and the output comprises sets of figures that tell the discourse analyst about the frequency of occurrence of words, phrases, collocations or structures. These statistics are then used to produce dictionaries, grammars, and so on. But for the discourse analyst, statistical facts raise the question "Why?", and the answers can only be found by looking at the contexts of the texts in the corpus. Discourse analysts, therefore, work with corpora in a qualitative way.

In this way corpus researchers in the latter tradition use the statistics to formulate and check hypotheses and research questions to be explored more closely in a qualitative way (for example, Farr & O'Keeffe, 2004). At the other end of the continuum, linguists focus strongly on frequencies, patterns, recurrence (for example, Biber *et al.*, 1998 Hunston & Francis, 1998; Kennedy, 2002). This does not mean to suggest that one approach excludes the other, simply that one takes precedence over the other resulting from differences in interest and objectives. Both approaches are explored on the MA in ELT programme although the former is more frequently practised.

Clearly, then as a methodology and research tool in many fields related to LTE and language teaching, corpus linguistics has been making its mark. However, it has not had an easy passage as evidenced in the following commentaries:

This cornucopia has not been welcomed with open arms [...] It has been kept waiting in the wings, and only in the last few years has any serious attention been paid to it by those who consider themselves to be applied linguists. For a quarter of a century, corpus evidence was ignored, spurned and talked out of relevance, until its importance became just too obvious for it to be kept out in the cold. (Sinclair, 2004b: 1)

Corpora seem to have entered the classroom from the backdoor. Whilst corpus data have long established themselves as the real language data [...] sweeping away

resistance as to their descriptive and, more controversially, pedagogic value, the actual use of corpora in language learning strategies has for a long time remained somewhat behind such momentous breakthroughs. This now seems less true. (Bernardini, 2004: 15)

Although the application of corpora in the compilation of dictionaries (for example, Gillard & Gadsby, 1998) and more recently grammar reference materials (for example, Carter & McCarthy, 2006) has become mainstream, according to Sinclair (2004: 2), two developments have led the language teaching profession to realise their potential for use in the classroom. Firstly, the teaching of lexical and phraseological patterns (for example, McCarthy & Carter, 2004; Schmitt, 2004) needed higher priority and these are best accessed through corpora, and secondly language variety, register, and genre have a determining influence on language choice and use (for example, Farr & O'Keeffe, 2004; Reppen *et al.*, 2002). Consequently, we now see evidence of corpus integration in various arenas including dedicated conferences, for example, TALC (Teaching and Language Corpora), IVACS (Inter-varietal Applied Corpus Studies), AAACL (American Association of Applied Corpus Linguistics), and ICAME (International Corpus of Archives of Modern English) and streams in major conferences, for example, EuroCALL. In addition corpus-based dictionary and grammar reference books have been produced, as well as pedagogically oriented books, for example, Aston, 2000; Burnard & McEnery, 2000; McEnery *et al.*, 2006; Sinclair, 2004a; Wichmann *et al.*, 1997. Dedicated journals have been published, for example, The International Journal of Corpus Linguistics, and ICAME, and special issues of international journals, for example, TESOL Quarterly (2003, 27/3), Language Learning and Technology (2001, 5/3). And inevitably many on-line corpus resources, activities, and discussion groups are now in circulation, as any quick www search will reveal.

While much debate continues on some of the issues and problems associated with corpus integration in the classroom, most of the application has happened by way of raising levels of language awareness for those engaged in its use. Most of this work has been done on how to exploit and evaluate corpus use in the language classroom (for example, Chambers, 2005; Tribble, 2000), and not in LTE environments, despite the fact that language awareness is also considered to be an essential element of a teacher's competencies. Andrews (2001) reports on how a teacher's language awareness affects their pedagogical competence, and later (Andrews, 2003), how it is

equally important as part of their subject-matter knowledge (see also Brumfit, 1991; Wright, 1991). As we have already discussed, technological competence is another important tool in a teacher's repertoire, and therefore it plausibly follows that as both are embodied in the use of corpus-based approaches, they might form a logical component of pre- or in-service programmes. However, despite calls for corpus assimilation at this level (Chapelle, 2001) there have been few reported accounts of its integration in teacher education contexts. There are some notable exceptions. Hunston (1995) argues for 'the use of computer-stored corpora in courses in grammar awareness for teachers of Mother-Tongue English' and exemplifies some of the ways in which this can be done and compares these with using a more traditional individual text-based approach. Subsequently, Renouf (1997), also reporting from the University of Birmingham, home of the Cobuild, and later Bank of English corpora, provides a complementary account of the corpus component of an LTE course for postgraduate students. In the same year, Coniam (1997) furnishes an equally practical account of some of the corpus-based tasks for language investigation in his teacher training programme. Finally, and most recently, we have broadened the applications of corpus-based techniques in LTE to their use for uncovering socio-cultural nuances as well as linguistic patterns (O'Keeffe & Farr, 2003). Despite these insightful accounts, none provide extensive evaluation of the type of integration they have proposed or have been practising. Therefore this paper aims to fill that gap and close the evaluation circle (Jamieson & Chapelle, 2004) so that the results may feed back into our professional reflection, practice, and development loop in relation to corpus use.

Context and Participants

The study reported in this paper is based at the University of Limerick, Ireland, and the participants were all student teachers (STs) on our year long MA in English Language Teaching programme between 2003 and 2005. Over the two-year period in question a total of 28 STs were admitted to the programme, 25 of whom responded to Questionnaire A reported below. They ranged in age from 22 to early 60s, and although most of the participants were Irish, a variety of other nationalities were represented including: American, Welsh, English, Polish, German, and Chinese.

Some had no previous teaching experience and others had up to 19 years, either in Ireland or in other international contexts. All had some general experience of using computers for personal and professional purposes, although a few were not confident in their abilities in this respect.

The corpus linguistics component of the MA programme is integrated into two language systems modules, one offered in each semester. All students take these content-based modules for a total of 4 hours per week in Semester 1, and 3 hours per week in Semester 2. Each semester comprises 12 teaching weeks. In general, the module content covers aspects of English grammar: formal, functional, discourse-based, phonology, morphology, and semantics. The use of corpora as an investigatory tool to explore various facets of this content is introduced in the first semester and is carried through to the end of the second semester exploring in further detail the relevant aspects of language as the modules progress. Approximately 15-20 hours of class time per semester is spent on corpus-based activities. STs had access to various corpus collections including: British National Corpus (BNC), American National Corpus (ANC), Corpus of Spoken Professional English, the ICAME collections, the Limerick Corpus of Irish English (L-CIE), and any available on-line collections or samplers, for example the Cobuild sampler. The on-campus CALL lab allows manipulation of the various corpora through the use of Wordsmith Tools, Monoconc, and Paraconc software packages. All activities are inductive and aimed at improving students awareness of language use in an independent learning environment. The lecturer is available in the capacity of facilitator or technical advisor during each of the hands-on sessions. The pedagogic application and potentials of using corpora for teaching language is not included in this module but in a Theory and Practice module which has a substantial section on the use of ICT for teaching purposes. The following paragraph gives a brief outline of the structure and content of the corpus component of the Language Systems modules.

In the first semester theoretical discussions address background issues, definitions, approaches, uses, and availability of corpora. These are accompanied by practical tutorials demonstrating the use of wordsmith tools (including wordlist, corpus statistics, concordance, and keyword tools). Students then complete guided and freer tasks corresponding with the content of the Language Systems lectures e.g. verb

patterns, main and auxiliary verbs, tense and aspect etc. There is also a hands-on seminar on the use of corpora for research, as some students will go on to use this for their research dissertations. In the second semester students complete an initial task, for example, “A corpus-based examination of conditionality in English”, during which they work in pairs using different and various corpora. The results of this task are compared in tutorial sessions and any problems addressed. Students are then set a more substantial assignment in the form of a corpus-based discourse analysis paper on topics of their own choice. Sample assignments submitted include: an analysis of convergence and divergence in some of the uses of *have* in Irish English, *Like* as a pragmatic marker in Southern Irish English and London Teenage English, *Okay* as a discourse marker in Irish English, an analysis of modal verb use in a Chinese learner corpus, and a discourse analysis of the use of vague language in conversational interaction. STs were given the opportunity to work on these assignments in supervised tutorial sessions for one to two hours a week over a period of several weeks. In addition a minimum of three guest lectures/workshops were organised over the two semesters in each year and delivered by external experts in the field of applied corpus linguistics.

Student Teachers’ Perceptions and Attitudes to Corpus-based Study

Towards the end of Semester 2 a questionnaire aimed at uncovering their attitudes and gaining some feedback on the corpus component of the programme was distributed to all STs. The questionnaire was designed on the basis of a pilot study on the previous year’s cohort of STs, where they were asked to provide undirected feedback on this component of their programme. Before distributing the questionnaire, the researcher emailed the MA groups giving some information about the study being conducted. In this email several issues were addressed including, the option of giving their names or of excluding them and assurance that all names included would be kept confidential, information that the results would be used to form the basis of part of a conference paper and subsequent publication authored by the researcher. It was also made clear that there was no obligation on any of the students to complete the questionnaire. A week after the email had been sent the questionnaires were distributed during class hours and the STs were asked to complete them in their own time and return them to

the researcher either in person, by dropping them into a mailbox, or by email. A general email reminder was sent to the entire group two weeks later requesting them to submit any outstanding questionnaires.

25 questionnaires were returned by the two cohorts and the following results are based on these returns. The first question asked STs to rate a number of statements on a scale of 1 to 4 based on level of agreement, with 1 indicating strong agreement and 4 strong disagreement. The 1 to 4 scale meant that students had to chose to be either more positive of negative in their evaluations and was used intentionally for this reason, although, it does also contain the danger of promoting either positive or negative bias. If the STs were really unsure of their opinion on any of the statements, they were instructed to leave the scale blank. Each of the statements is reproduced below with the results graphed for ease of interpretation. Results for the other questions follow in appropriate presentation modes.

Question 1

Enjoyment and choice

A. I enjoy using language corpora in my studies.

B. I would choose to use corpus linguistics in my studies if I had the option.

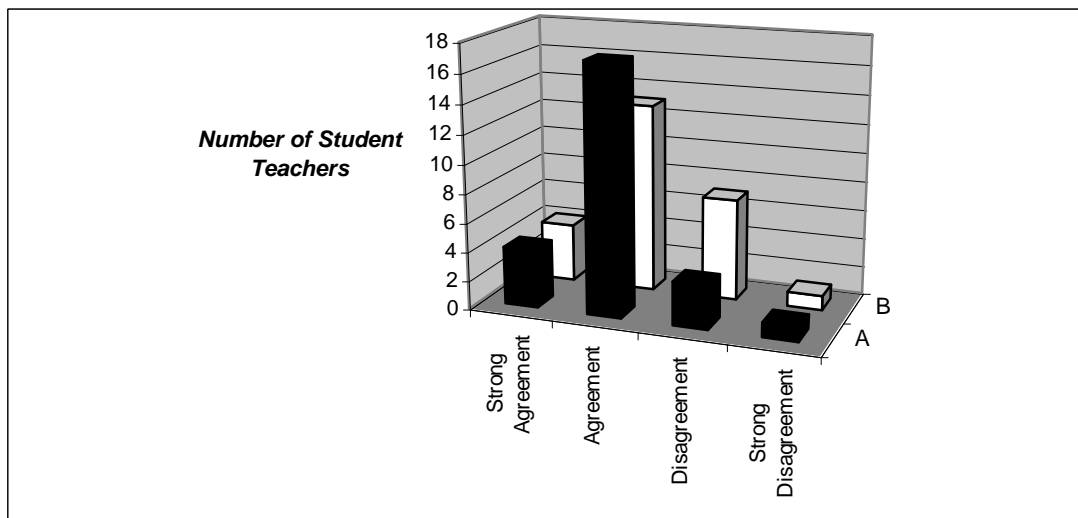


Figure 1 STs Enjoyment and Choice of Using Language Corpora

Figure 1 shows that the majority of students enjoy and would choose to use corpora in their studies. Interestingly, and seemingly contradictory, fewer would choose corpora than those who enjoy it. It may be that they would choose to use the time on other things or they would prefer to stick to more traditional instructional mode, which they may like equally or more, but this does not detract from the enjoyment of their experience with corpus linguistics.

Benefits

C. I feel my teacher education, in general, benefited from the integration of corpus-based learning.

D. I feel my understanding of language systems benefited from the integration of corpus-based learning.

E. I feel my spirit of enquiry and research benefited from the integration of corpus-based learning.

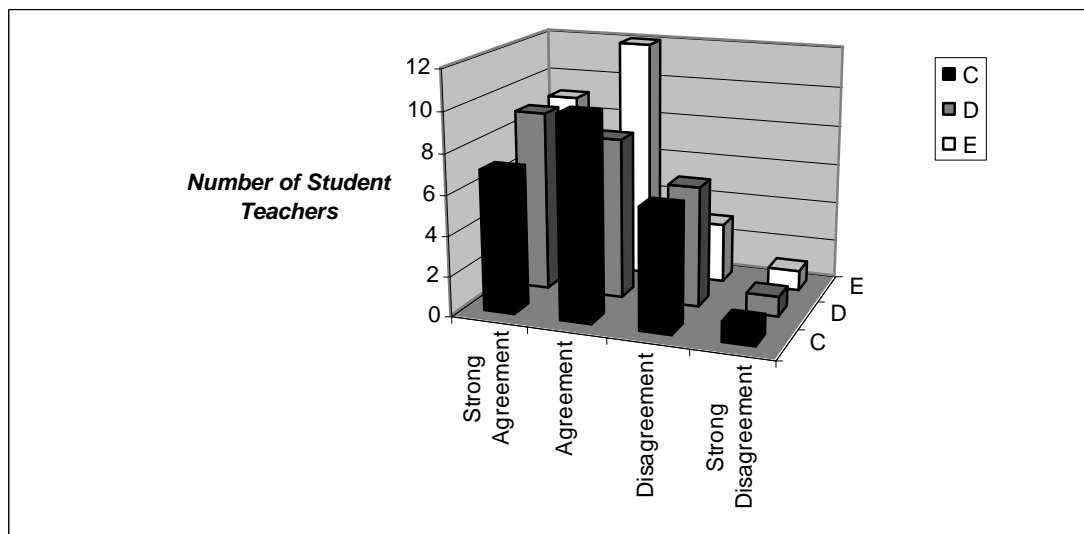


Figure 2 Perceived Benefits of Using Language Corpora

Figure 2 illustrates that although there is majority agreement that benefits exist at the three levels of general teacher education, understanding of language systems, and spirit of enquiry and research, the largest majority perceive the real advantage lies in

the latter of these. Already there is a hint that STs can see most of the pros at the level of using corpora for research type investigations.

Technical Difficulties

F. I had initial difficulties with the technical aspects of using corpora.

G. I still have difficulties with the technical aspects of using corpora.

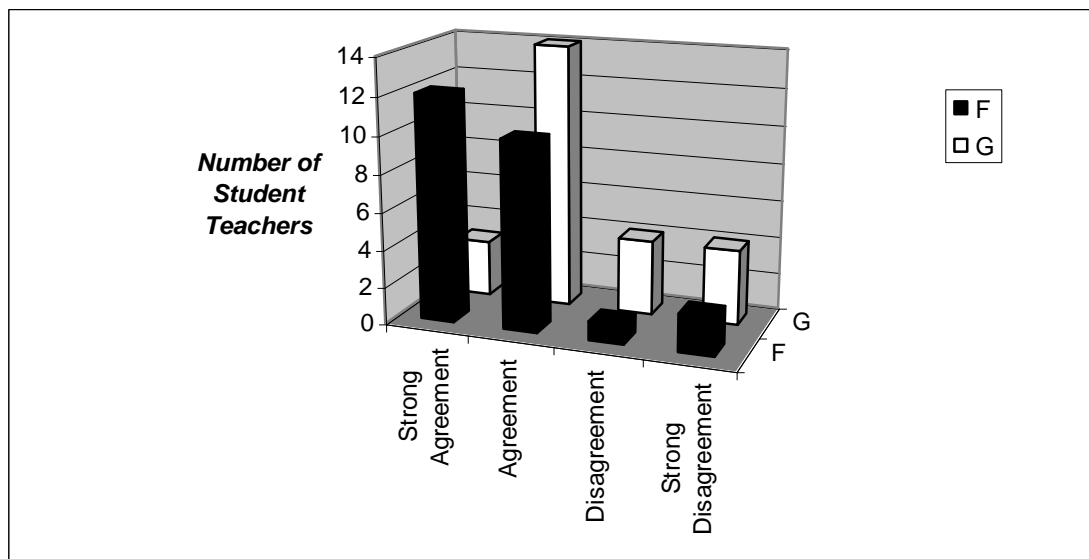


Figure 3 Technical Difficulties Using Language Corpora

Linguistic and Conceptual Difficulties

H. I had initial difficulties with the linguistic and conceptual aspects of using corpora.

I. I still have difficulties with the linguistic and conceptual aspects of using corpora.

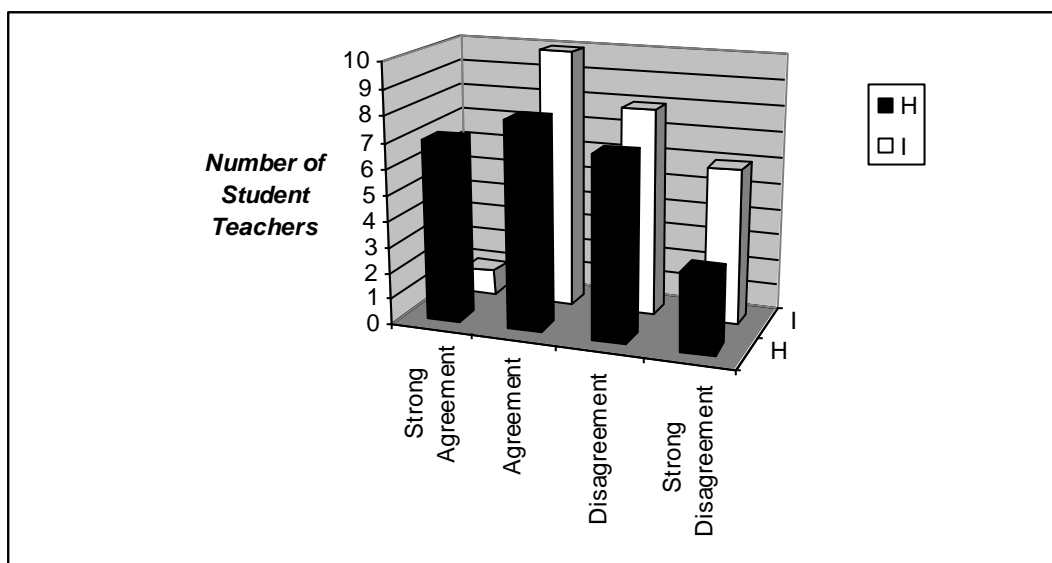


Figure 4 Linguistic and Conceptual Difficulties Using Language Corpora

Most students had both technical and conceptual problems at the initial stages of using language corpora, as seen in Figures 3 and 4. Technicalities caused over 80% of the STs some concern and this concern still persisted for over 60% by the end of the second semester when students had had much exposure and assistance in using the corpora and software. The linguistic side created fewer troubles and by the end of the programme the percentage in this respect had dropped to just above 40%. Clearly, then the complexity of the software was and continued to be an issue throughout.

Question 2

A. How much time, outside of class time, did you spend working with language corpora during the first and second semesters?

In Semester 1 there was a degree of difference in the results between the two cohorts. In the first cohort (03/04) only two STs spent any time outside of class time working with corpora and this amounted to only about 3 hours each. The second cohort spent an average of 7 hours, due mainly, it is strongly suspected to the fact that they were assigned an exam question based on corpus-investigation to prepare in advance of the end of semester assessment. In Semester 2 this average rose to 16 hours per ST for each group. This is attributable to the fact that students had to complete the compulsory corpus-based discourse project, as mentioned earlier, and also to the fact that the STs had already begun work on their MA dissertations, some of which were based on this methodology (see below).

B. Which corpora did you use and for what purposes?

Students mention many of the corpora found on the ICAME collection: COLT, Brown, WC, WSC and ACE. They also favoured L-CIE and CSPAE, as well as some on-line learner corpora they had sourced. No mention was made of the BNC or of the ANC, both of which were available to the STs in the university Language Resource Area. The BNC may have been avoided as it involves the use of its own in-built software, which was not demonstrated in tutorial sessions, and the ANC was a relatively new addition to our collection, arriving half way through the year and again may have been considered too much of an unknown to risk exploring.

Question 3

From your experiences what are the positive and negative aspects of learning through the use of corpora?

STs listed the following, many of which were reiterated among many members of the group. Comments are arranged in order of frequency of mention by the STs, with those identified most often appearing first. The number of students who made each of the points is indicated in brackets below.

Positive:

Can see real language use – language in context and cultural insights (19)

Variety, fun, interesting to explore (7)

Highlights frequencies and clusters making teaching decisions easier (3)

Uncovers functions and structures in language not found in grammar books (2)

Invaluable as a research tool. (2)

Changes perceptions of ‘correctness’ in language use (1)

Motivating to find answers without having to use grammar books (1)

Up-to date language (1)

Negative:

Technical aspects can be off-putting, corrupted data in some corpora (12)

With all the varieties that exist it becomes complicated if considering how to apply it to the classroom, classroom application is an issue (6)

Amount of time needed – can be frustrating (5)

Difficult to access and analyse the mega corpora e.g. BNC, ANC, and the appropriate software (3)

Not always confident of my interpretation (2)

Colloquial language use can be confusing, spoken language is messy (2)

From these listings, one can immediately identify the technical aspects, time, and language variety to be most problematic. Interestingly, in very recent research carried out by one of the MA students as part of her MA dissertation under my supervision, she found that the technical issue ceased to be a difficulty by the very end of the second semester as the STs had had enough exposure to be more familiar and confident in their usage of the corpora and software (Riordan, 2005: 52). Her research also found that the males in the group tended to be more positively predisposed to the use of corpora than the females. Time is often cited as a difficulty with the use of technology in education in general, and specifically in the use of corpora (see Chambers & O'Sullivan, 2004 for an account with undergraduate students of French, and Bernardini, 2004 for similar findings with undergraduate students of English language and literature). On the other hand, all participants could easily identify the advantages of corpus-use and their potential. The fact that corpora represent 'real language use in context' is viewed by many as an asset in their awareness of language process. The novelty, fun and interest aspect is also mentioned and the added dimension above what can be found in reference and grammar books. The obvious benefit of being able to ascertain information about frequency, clusters and collocations, not easily observable through other modes of language analysis is noted.

Question 4

Are you using corpora in your MA dissertation research? If so, for what purpose?

56% of the STs chose to use corpora in some way for their dissertation research.

Some took a direct approach and analysed language patterns or varieties, for example, features of Irish-English, while others used corpora indirectly to secure language examples for illustration or to elicit student opinions and evaluations. The percentage who have chosen to use corpora is relatively high given that they have complete open

choice of research topic and methodology, and also that only one member of their lecturing faculty and potential supervisors (of about seven) is involved in corpus-based research. This must be a sign that the STs either enjoy it or see it as a good and manageable research topic, or perhaps both.

Question 5

The final question relates to whether and, if so, how these STs might use corpora in their own future teaching careers. It was divided into four parts, each presented below. The results for the first three parts have been incorporated into Figure 5 below.

In relation to your own future teaching, please answer the following questions:

- a. Do you think you will use corpora as a resource in the preparation of your teaching and/or teaching materials?*
- b. Would you use computer-based corpus instruction in class if employed in an institution which already uses and has corpus resources and software?*
- c. Would you initiate the use of computer-based corpus instruction if employed in an institution which already had computer facilities for students, but not necessarily corpus software?*

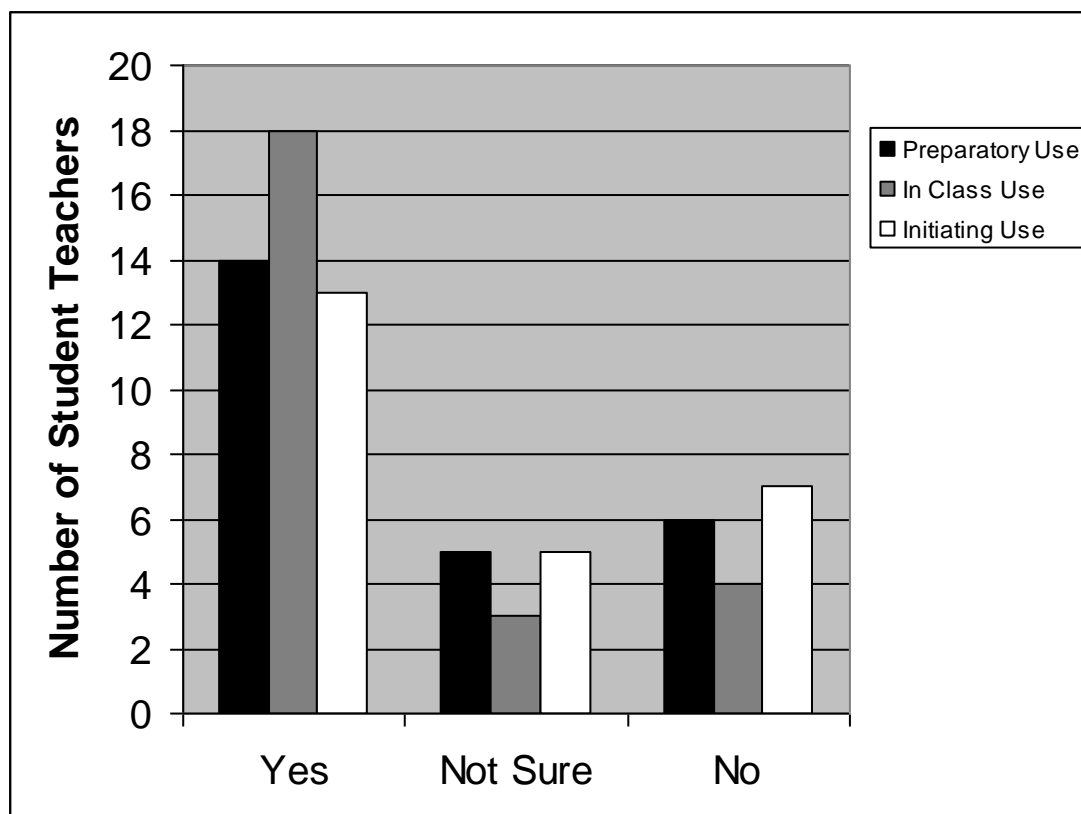


Figure 5 Future Integration of Corpora in the English Language Classroom

Interestingly, a clear majority indicate that they are very positively intentioned in relation to integrating corpora in their future teaching, with a maximum of 18 expressing a willingness towards in-class use. Understandably there is a relative reluctance to initiate corpus instruction where the specific facilities are not already installed and available and this is probably linked with the technical unease expressed in the previous question. The predicted uses listed below indicate that language awareness is seen as the predominant application by the STs.

If you have answered 'yes' to b. or c., please explain the uses you might make of corpora.

Gap-fills, idioms, phrasal verbs, semantic prosody, frequency, clusters.

To research words before teaching them.

To take real, authentic examples of use.

To teach Irish expressions, illustrate varieties of English.

To explain subtleties of language, e.g., semantic prosody, connotations, collocations.

To illustrate grammatical features not clear from the texts.

For areas of difficulty for students.

For student projects – for them to investigate corpora.

Hands-on lab sessions for student investigations, promote autonomy.

To show students the lexical and structural patterns to help them integrate these into their language use.

If you have answered 'no' to any of the above, please explain why not.

Too time consuming for students to use hands-on.

Expense.

Not enough confidence or expertise.

Difficult to change teaching ideologies in some institutions.

Prefer to use someone else's findings.

Material can be confusing without a framework for interpretation.

The STs postulate a wide range of predicted uses to which corpora can be put in their future teaching environments. It is encouraging to see that corpus techniques are seen as an aid to some of the traditional methods and to function in raising students' awareness of such features as language variety, difficulty aspects of language, prosody and patterning, and very importantly to support an independent learning environment where the EL students can function in an independent but supported way in which they discover and increase their own awareness. Such hypothesised uses of corpora for teaching must surely come from the STs' own experiences of how and how best they can be utilised in a language learning situation. On the other hand, four of the reasons quoted for not integrate are directly related to practical constraints: time, expense, not enough expertise, difficult to change institutional ideologies. Another, preferring to use others' findings, could be a reflection either of difficulties with time and expertise, or a dislike of the technology, again practical factors. And finally, problems with frameworks for interpretation is listed as the only conceptual or methodological issue, leading to the conclusion that given a favourable practical teaching scenario, the vast majority of STs show a balanced volition in relation to classroom integration. The interesting question, of course, is whether these good intentions are realised when the STs hit the reality of their first teaching position. In order to gain some insights in this respect, a follow-up email survey was conducted with the cohort of 2003/2004 in February 2005, and is planned for the 2004/2006

cohort some time in the coming year when they have begun working. This will be part of a three-year study to examine the more longitudinal effects. The perennial difficulty is of course getting responses to questionnaires issued at a distance. There was a return rate of just 6 from the group of 11 teachers, and of these, only 5 were employed in ELT positions. The results are therefore incomplete at this point and mention will be made only of the fact that 2 of the 5 are currently using corpus techniques and the other 3 still plan to but for practical reasons, such as availability of software and hardware or working to a strict methodology, they are not doing so at present. Longer term evaluation will be reported with this year's and subsequent cohorts of STs as the results become available.

Conclusion

At a point where corpus-based approaches to the study of language has finally reached the teacher education platform, primarily, and certainly in this case, for the promotion of a better conscious awareness of how language operates in multi-faceted ways, the findings from this study are most heartening. However, on-going evaluation and necessary moderation must always be enforced, as with any teaching tool or methodology in the good practice of a reflective model of on-going learning for teachers and teacher educators. There are a number of significant conclusions which can be drawn from the results discussed above and the most significant will be presented in brief bullet point format below.

- STs show an overwhelmingly positive disposition towards the use of corpora in terms of their enjoyment and also the perceived benefits.
- Although benefits are seen in general terms, they are mostly identified to be associated with the STs' appreciation of language systems and even more significantly with the promotion of their spirit of enquiry and research. This bodes well for their use directly in language awareness-raising.
- Most difficulties arose at the level of technicalities with the software, although some conceptual problems were mentioned. The latter were resolved in the short-term and the former over the course of the full academic year with quite a lot of practice.

- Time is pinpointed as a further constraining factor and this raises the question for instructors of the time to benefit return. Ideally this should be evaluated in an experimental research study involving different groups of STs.
- STs show a critical awareness as opposed to blind adaptation of this approach and this is to be encouraged and welcomed.
- A large majority of STs choose to use corpora in their own further research in which they had open choice of topic and methodology. Again this reinforces their positive perception of their use as research tools in the promotion of language awareness and independent investigation.
- Finally, circumstantial factors caused low rates of application in the one cohort of STs' first year of teaching which was evaluated. However, the now teachers continue to be committed to integrating them in the future given more conducive teaching environments. This conclusion is tentative and more longitudinal research is being conducted in this respect.

In general terms the results of this study are most encouraging and suggest that the continued integration of corpus-based instruction in the language content component of language teacher education programmes should be encouraged despite some identified difficulties. From the perspective of the tutor, I can reinforce the STs' voices in relation to the initial technical difficulties, especially in attempting to deal with various levels of technical competencies within the cohorts. Also, more time is needed than in traditional classroom mode language awareness activities but I feel strongly that the advantages and added dimension afforded to the participants in the MA programme far outweigh any issues or minor hitches and this is supported by present attitudinal findings. Therefore, I feel reasonably confident in saying that the use of corpora in LTE within the frameworks of progressive educational practices and the use of ICT in education is deemed to be largely successful in the context discussed in this paper.

References

- Andrews, S. (2001) The language awareness of the L2 teacher: its impact upon pedagogical practice. *Language Awareness* 10(2&3), 75-90.
- Andrews, S. (2003) Teacher language awareness and the professional knowledge base of the L2 teacher. *Language Awareness* 12(2), 81-95.
- Aston, G. (2000) Corpora and Language Teaching. In L. Burnard and T. McEnery (eds) *Rethinking Language Pedagogy from a Corpus Perspective. Papers from the Third International Conference on Teaching and Language Corpora* (pp. 1-18). Frankfurt: Peter Lang.
- Barnes, A. and Murray, L. (1999) Developing the pedagogical ICT competence of modern foreign languages teacher trainees. Situation: all change and *plus ça change*. *Journal of Information Technology in Teacher Education* 8(2), 165-180.
- Benson, P. (2001) *Teaching and Researching Autonomy in Language Learning*. Essex: Pearson Education Ltd.
- Bernardini, S. (2004) Corpora and the classroom: an overview and some reflections on future developments. In J. M. Sinclair (ed.) *How to Use Corpora in Language Teaching* (pp. 15-38). Amsterdam: John Benjamins.
- Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Brumfit, C. (1991) Language awareness in teacher education. In C. James and P. Garrett (eds) *Language Awareness in the Classroom* (pp. 24-39). London: Longman.
- Bruner, J. (1985) Vygotsky: a historical and conceptual perspective. In J. V. Wertsch (ed.) *Culture, Communication and Cognition. Vygotskian Perspectives* (pp. 21-34). New York: Cambridge University Press.
- Burnard, L. and McEnery, T. (eds) (2000). *Rethinking Language Pedagogy from a Corpus Perspective. Papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt: Peter Lang.
- Butler, C. S. (2002) 'Corpus linguistics as work'. A review. *System* 31(1), 128-132.
- Carter, R. and McCarthy, M. J. (2006) *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Chambers, A. (2005) Integrating corpus consultation in language studies. *Language Learning and Technology* 9(2), 111-125.
- Chambers, A. and O'Sullivan, I. (2004) Corpus consultation and advanced learners' writing skills in French. *ReCALL* 16(1), 158-172.
- Chapelle, C. A. (2001) CALL in the 21st century: looking back on research to look forward for practice. In *CALL in the 21st Century CD-ROM* (pp. -). Whitstable: IATEFL.
- Coniam, D. (1997) A practical introduction to corpora in a teacher training language awareness programme. *Language Awareness* 6(4), 199-207.
- Crookes, G. (1997) What influences what and how second and foreign language teachers teach? *The Modern Language Journal* 81(i), 67-79.
- Cuban, L. (2001) *Oversold and Underused: Computers in the Classroom*. Cambridge, MA: Harvard University Press.
- Cummins, J. (2000) Academic language learning, transformative pedagogy and information technology: towards a critical balance. *TESOL Quarterly* 34(3), 537-547.
- Doering, A. and Beach, R. (2002) Preservice English teachers acquiring literacy practices through technology tools. *Language Learning and Technology* 6(3), 127-146.

- Egbert, J., Paulus, T. M. and Nakamichi, Y. (2002) The impact of CALL instruction on classroom computer use: a foundation for rethinking technology in teacher education. *Language Learning and Technology* 6(3), 108-126.
- Farr, F. and O'Keeffe, A. (eds) (2004). *Corpora, Varieties and the Language Classroom. Teanga 21*. Dublin: IRAAL.
- Gillard, P. and Gadsby, A. (1998) Using a learners' corpus in compiling ELT dictionaries. In S. Granger (ed.) *Learner English on Computer* (pp. 159-171). London/New York: Longman.
- Halliwell, S. (1993) Teacher creativity and teacher education. In D. Bridges and T. Kerry (eds) *Developing Teachers Professionally. Reflections for Initial and In-service Trainers* (pp. 67-79). London/New York: Routledge.
- Harris, S. (1995) Pragmatics and power. *Journal of Pragmatics* 23, 117-135.
- Hunston, S. (1995) Grammar in teacher education: the role of a corpus. *Language Awareness* 4(1), 15-31.
- Hunston, S. and Francis, G. (1998) Verbs observed: a corpus-driven pedagogic grammar. *Applied Linguistics* 19(1), 45-72.
- Johns, T. (1991) Should you be persuaded - two samples of data-driven learning materials. *English Language Research Journal* 4, 1-16.
- Kennedy, G. (2002) Variation in the distribution of modal verbs in the British National Corpus. In R. Reppen, S. Fitzmaurice and D. Biber (eds) *Using Corpora to Explore Linguistic Variation* (pp. 73-90). Amsterdam: John Benjamins.
- Kinginger, C. (2002) Defining the Zone of Proximal Development in US foreign language education. *Applied Linguistics* 23(2), 240-261.
- Kozulin, A. (1998) *Psychological Tools. A Sociocultural Approach to Education*. Cambridge, Massachusetts: Harvard University Press.
- Leech, G. (1991) The state of the art in corpus linguistics. In K. Aijmer and B. Altenberg (eds) *English Corpus Linguistics* (pp. 8-29). Essex: Longman.
- Louw, B. (1993) Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis and E. Tognini-Bonelli (eds) *Text and Technology: In Honour of John Sinclair* (pp. 157-176). Philadelphia: John Benjamins.
- McCarthy, M. J. (1998) *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. J. and Carter, R. (2004) *This that and the other: Multi-word clusters in spoken English as visible patterns of interaction. Teanga 21*, 30-52.
- McCarthy, M. J., Matthiessen, C. and Slade, D. (2002) Discourse analysis. In N. Schmitt (ed.) *An Introduction to Applied Linguistics* (pp. 55-73). London: Arnold.
- McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-Based Language Studies. An Advanced Resource Book*. London and New York: Routledge.
- Mercer, N. (1995) *The Guided Construction of Knowledge. Talk Amongst Teachers and Learners*. Philadelphia: Multilingual Matters.
- Meskill, C., Mossop, J., DiAngelo, S. and Pasquale, R. K. (2002) Expert and novice teachers talking technology: precepts, concepts, and misconcepts. *Language Learning and Technology* 6(3), 46-57.
- Murray, L. (1998) CALL and Web training with teacher self-empowerment: a departmental and long-term approach. *Computers and Education* 31(2/3), 17-23.
- Nyikos, M. and Hashimoto, R. (1997) Constructivist Theory applied to collaborative learning in teacher education: in search of ZPD. *The Modern Language Journal* 81(iv), 506-517.

- O'Keeffe, A. and Farr, F. (2003) Using language corpora in language teacher education: pedagogic, linguistic and cultural insights. *TESOL Quarterly* 37(3), 389-418.
- Oxford, R. (ed.) (1996). *Language Learning Strategies Around the World: Cross-Cultural Perspectives (National Foreign Language Center Technical Reports Series, No 13)*. Hawaii: University of Hawaii Press.
- Pennington, M. C. (2001) Writing minds and talking fingers: doing literacy in an electronic age. In *CALL in the 21st Century CD-ROM* (pp. Whitstable: IATEFL).
- Renouf, A. (1997) Teaching corpus linguistics to teachers of English. In A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (eds) *Teaching and Language Corpora* (pp. 255-266). New York: Longman.
- Reppen, R., Fitzmaurice, S. M. and Biber, D. (eds) (2002). *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins.
- Revell, J. and Norman, S. (1997) *In Your Hands. NLP in ELT*. London: Saffire.
- Schmitt, N. (ed.) (2004). *Formulaic Sequences*. Amsterdam: John Benjamins.
- Sinclair, J. M. (2004a) *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Sinclair, J. M. (2004b) Introduction. In J. M. Sinclair (ed.) *How to Use Corpora in Language Teaching* (pp. 1-14). Amsterdam: John Benjamins.
- Sinclair, J. M. (1996) The search for units of meaning. *Textus* IX, 75-106.
- Tammelin, M. (2001) Empowering the language teacher through ICT training and media education: Case HSEBA. In *CALL in the 21st century CD-ROM* (pp. Whitstable: IATEFL).
- Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tribble, C. (2000) Genres, keywords, teaching: towards a pedagogic account of the language of project proposals. In L. Burnard and T. McEnery (eds) *Rethinking Language Pedagogy from a Corpus Perspective. Papers from the Third International Conference on Teaching and Language Corpora* (pp. 75-90). Frankfurt: Peter Lang.
- Vygotsky, L. S. (1978) *Mind in Society. The Development of Higher Psychological Processes*. Cambridge MA: Harvard University Press.
- Warschauer, M. (2000) The changing global economy and the future of English teaching. *TESOL Quarterly* 34(3), 511-535.
- Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (eds) (1997). *Teaching and Language Corpora*. London: Longman.
- Wright, T. (1991) Language awareness in teacher education programmes for non-native speakers. In C. James and P. Garrett (eds) *Language Awareness in the Classroom* (pp. 62-77). London: Longman.