

**RAMESH KRISHNAMURTHY**

**ELECTRONIC RESOURCES FOR LANGUAGE  
TEACHING AND LEARNING: CORNUCOPIA OR  
INFORMATION OVERLOAD?**

**Abstract** : The rapid increase in the amount of language teaching and learning materials available in electronic form, whether on CD-Rom or on the World Wide Web, now presents teachers and learners with problems of how to find them and how to evaluate them. This paper, based on the author's personal experiences and current research activities, describes the problems and suggests ways in which the situation may be improved in the future.

**1. Introduction**

As a teacher of courses in Corpus Lexicography and Linguistics since 1991, I probably encountered many of the problems associated with electronic resources earlier than many of my colleagues. But now that the Internet has exploded into the consciousness of every teacher and learner, and resources have increased at such an incredible rate, more and more teachers and students have become aware of them. In fact, I have been involved in teaching courses in Corpus Lexicography and Linguistics (especially in relation to the teaching of English as a Foreign Language) since 1984, to trainee lexicographers within the Cobuild project, but it was only in 1991 that I started introducing these topics to a wider audience.

**2. Courses**

Since 1991, I have delivered corpus-related courses to students at the University of Birmingham, then at other universities within the UK and abroad. I have also given many individual talks and lectures on aspects of these subjects at institutions of higher education (including the Esterhazy Karoly Teachers College at Eger) and other public venues all over Europe, to audiences of undergraduates, postgraduates, professionals, and interested

members of the public. Throughout this time, I have had the privilege of access to the large corpora of natural English texts collected by Cobuild at the University of Birmingham, initially as a member of the Cobuild staff and since 1997 as an Honorary Research Fellow of the University. In 1984 the corpus was about 7 million words in size, and by 1999 it has expanded to about 330 million words.

The Cobuild courses were extremely detailed and practically oriented towards specific publications, focussed substantially on in-house editorial policies, and made use of largely in-house resources, so they are not really of relevance to the topic under consideration. Here is an outline of some of the other courses I have taught on:

YEAR	PLACE	SHORT TITLE	AUDIENCE	DURATION
1991-3	Birmingham, UK	Corpus Lexicography	5-10 MA students (+ guests)	12 hours in 8 weeks
1992	Brighton, UK	Lexicography	50 Undergraduates	16 hours in 8 weeks
1995	Debrecen, Hungary	From Corpus to Dictionary	20 Undergraduates, Postgraduates, and Staff	39 hours in 6 days
1996	Budapest, Hungary	Computational Lexicography	20 Undergraduates, Postgraduates, and Staff	30 hours in 4 days
1997	Debrecen, Hungary	Computers and Text	20 Undergraduates, Postgraduates, and Staff	40 hours in 7 days
1997	Zagreb, Croatia	Dictionaries and Computers	30 Undergraduates, Postgraduates, and Staff	6 hours in 1 day
1998	Madrid, Spain	Corpus for Science and Technology	45 Academic Staff	16 hours in 4 days
1998	Sogndal, Norway	Corpora and Computer Text Analysis in the Classroom	20 Teachers and Teacher-Trainers	8 hours in 2 days

### 3. Course Contents

Course contents obviously varied according to the type of students and length of course. Inevitably, courses I myself designed were strongly influenced by my experience at Cobuild. For example, one of my early courses in Corpus Lexicography (Birmingham 1992) had the following components:

1. Types of dictionary, dictionary structures, dictionary contents
2. History of lexicography, use of intuition, citations, and corpus evidence
3. Corpus design criteria: data capture, coding, and storage systems
4. Corpus analysis: frequencies, concordances, collocations, part-of-speech tagging
5. Cobuild methodology: headword selection, definitions, examples
6. Cobuild products: dictionaries, grammars, usage books, guides
7. The Future: larger and different corpora, new software tools, electronic products

The rationale for the course design was to make students aware of (1) the wide range of dictionaries available for different purposes, the differences in the nature of the information provided, and the different ways in which the information can be presented (2) the historical changes in the philosophy and methodology of dictionary compiling, in particular the shift from prescriptive to descriptive goals, and the accompanying move from intuition and made-up examples to empirical analysis of data and authentic examples (3) the changes in dictionary-making technology from handwritten dictionary text and citations on index-cards filed in shoeboxes, to corpora on fiche (and later online) and analyses entered on printed forms and keyed and stored in electronic databases, and semi-automatic extraction of formatted dictionary files, to simultaneous online corpus analysis and keyboarding of dictionary entries by lexicographers using software templates (4) the impact that these changes in philosophy, methodology and technology have had on dictionary content (using Cobuild as the main example) and the creation of entirely new reference publications, with a speculative glance into the future.

Some of the course titles indicate the direction in which the courses have since developed: "From Corpus to Dictionary" is similar to the course above. But "Computers and Text: a practical course in using computers for language analysis" suggests a wider approach, still computationally-oriented but no longer solely corpus-oriented. "The Science and Technology of Corpus, and Corpus for Science and Technology" reflects the need for more specifically targeted corpora and techniques, and the interest in them by teachers of ESP. "The Use of Corpora and Computer Text Analysis in the Classroom" highlights the pedagogical applications of corpus and computational methodologies and CALL.

Several of these courses were conducted with co-tutors: several colleagues from Cobuild, Patrick Hanks (Chief Editor, Current English Dictionaries, Oxford University Press), Gregory Grefenstette (Project Leader, Rank Xerox Research Centre, Grenoble), Tamas Varadi (Hungarian Academy of Sciences), and Bela Hollosy (Senior Lecturer and Deputy Head of Department, Debrecen University). The addition of co-tutors can obviously increase the breadth and depth of the treatment of course topics. In the 1995 Debrecen course, Patrick Hanks dealt with the broader theoretical and philosophical aspects, as well as the publishing issues (Practical Issues in Dictionary Publishing), Gregory Grefenstette focussed on the computational methodology and technicalities, and I gave a more practical view of the lexicographer's task in trying to balance the demands of theory and the commercial publishing world against the wealth of linguistic description which corpus analysis can generate. The Budapest course in 1996 allowed me to take over some of the discussion of theory, with Gregory Grefenstette once more dealing with the programming side of corpus computational techniques, and Tamas Varadi giving a concentrated tutorial on the PERL programming language. The 1997 Debrecen course saw Bela Hollosy taking the tutoring role for computational methods, and a more thematic approach to the sessions.

In the 1998 Madrid course, I tried to focus on the use of corpora and other computational resources for research and teaching, with special reference to scientific and technological discourse. The 1998 Sogndal course included a session on computer text analysis (looking closely at newspaper articles, poetry, fiction, and dictionaries), and one on exploiting a corpus for classroom uses.

#### **4. Course Presentation techniques and problems: from OHP to computer cluster**

Initially, my course sessions were presented entirely on OHP transparencies, sometimes accompanied by some printed handouts, and sometimes making use of a blackboard/whiteboard. It has always seemed somewhat of a mockery to be illustrating the power of a huge computer corpus and sophisticated analytical software through static displays on overhead projector slides. However quickly I changed the slides to simulate the rapid display sequences of a computer screen, I always had to say 'at the press of a button/at a single keystroke, my computer would show you this...'.  
'

In many of the early courses, I could not get access to a computer at all. In some places, I had one computer whose screen display was projected onto a wall or white screen. In some sessions of Budapest 1996, and in all sessions of Debrecen 1997, Madrid 1998, and Sogndal 1998, every participant had a computer. As soon as it became feasible, I started to use a computer in my presentations, and demonstrated the corpus via an online connection, using `telnet' to login directly, or Netscape to access data via Cobuild's website (<http://www.cobuild.collins.co.uk/>). In the early courses, the problems that manifested themselves were fragile computer links, slow speeds of data transfer, and paucity of any other widely accessible resources.

I have recently seen with envy more and more of my colleagues presenting corpus-based papers and courses using Microsoft Powerpoint on a laptop and so on. But while these presentations are often visually entertaining, and informative, they still rely on pre-prepared (and therefore static) analyses. For example, if a member of the audience asks a question about a word or language pattern that the presenter has not prepared, the question simply cannot be resolved there and then. Only direct access to the corpus can supply the answer.

In principle, given the increasing power of laptop computers, and the increasing size of their hard disks, it would now be possible to take a fairly large sample of a corpus, with the retrieval software, on a laptop. But this would still mean that evidence for rarer words and patterns might not be found, and that word frequencies and collocational statistics and other corpus-size related displays would be scaled down and possibly skewed. Once you have worked with a large corpus, and got used to its scales and patterns, it is quite frustrating to work with smaller subsets. And of course one must not forget that the corpus sample would need to be re-indexed before transfer to the laptop, not necessarily a trivial task. One other technical point must be made here: until the arrival of Linux in recent years, corpora built and run on Unix systems could not be ported to a laptop PC running Windows.

In 1991, there were few other electronic resources available. More recently, I have started to take additional software (Microconcord, Wordsmith Tools, Multiconcord) on floppy disks with me (and with permission from the authors), in order to demonstrate the variety and range of products now on the market – and especially products that my audience could buy for themselves and use on their personal computers at home and at work, to look at their own data collections.

However, I have also encountered problems with the students: there is much initial reluctance to engage in hands-on activities. Many participants on the courses are embarrassed at their poor keyboard skills, or their lack of familiarity with computer systems. In many cases, this is quite understandable: they are away from their own computers, being asked to use a strange machine and strange software, to do tasks which they have never before attempted to do.

So even now, whatever facilities are promised, I always take my notes and examples of corpus data with me in the form of OHP transparencies. You never know what technical problems may arise...

### **5. Currently available resources: problems**

Anyone who wishes to see what advances have been made in corpora in the past few years need only look at Michael Barlow's website (<http://www.ruf.rice.edu/~barlow/corpus.html>), or visit the site of one of the world's major centres for language engineering resources, such as ELRA (European Language Resources Association: <http://www.icp.grenet.fr/ELRA/home.html>) or LDC (The Linguistic Data Consortium: <http://www ldc.upenn.edu/>).

I am currently a member of the Language Learner's Workbench team of the European Commission-funded SELECT research project (Strategies for European LE-Enhanced Communication Training: EC Project LE4-8304) at the University of Wolverhampton (<http://www.wlv.ac.uk/select/>). A few months ago (just before the Sogndal course in October 1998) I collated and edited a review of existing language learning and language engineering resources and tools for the SELECT project. I was overwhelmed by the vast amounts of resources and tools now available, and the review eventually grew to 90-pages! For example, the review evaluated 12 CD-Rom products and 17 websites that catered for people learning Business English, and 14 CD-Rom products and 31 websites for students of general English. Language engineering resources included 16 speech corpora, 8 automatic translation systems, dozens of terminology banks, and so on.

As a simple illustration, here is a selection of webpages that offer help with English grammar:

**<http://www.ihes.com/Sresource/Sstudy/adverborder.html>**

Adverb Order: how to extend simple sentences by adding adverbials; where to put them and in what order.

**<http://www-personal.umich.edu/~cellis/antagonym.html>**

Common Errors in English – a page on the most common usage and spelling mistakes in English.

**<http://www.hiway.co.uk/~ei/intro.html>**

An Elementary Grammar – an entire grammar book online.

**<http://www.fairnet.org/agencies/lca/grammar2.html>**

ESL Grammar Notes: Articles – explanations and rules on using articles, countable and non-countable nouns, Explanations and rules on verb tenses.

**<http://www.pacificnet.net/~sperling/wwwboard2/wwwboard.html>**

ESL Help Center – twenty-four-hour help for ESL/EFL students from an international team of ESL/EFL teachers.

**<http://deil.lang.uiuc.edu/web.pages/grammarsafari.html>**

LinguaCenter's Grammar Safari. A great place for students to gather real grammar examples found on the World Wide Web.

**<http://www.edunet.com/english/grammar/toc.html>**

On-line English Grammar – an excellent grammar resource

**<http://www.ihes.com/Sresource/Sstudy/simplesentence.html>**

Sentence Structure: Simple Sentences: the parts of a simple sentence and how to put them together.

Anyone who has tried a simple search on a popular search engine such as Alta Vista will be familiar with the problem. For example, I have just searched for “English + grammar”, and I am told that “**688590 matches were found**”:

1. Business English grammar,vocabulary,listening and reading exercises
2. On-Line English Grammar
3. The Internet Grammar of English
4. Internet Grammar of English
5. Lydbury English Centre – Grammar page has moved
6. English Grammar
7. English Grammar Clinic – Links page
8. WORDbird: English grammar, editing, and writing
9. Welcome to Jonathan Revusky's Interactive English Grammar Pages
10. Basic English Grammar

How am I – or any teacher or student – supposed to cope with this inundation of information? One answer is that, of course, we do not have to use all of it! A visit to the first site listed may well give us the answer or the

material that we wanted! But why should we expect the first consultation to be perfectly successful? After all, when we go to the library, do we expect that the first book we find on our subject will be the ideal one? We are happy to chase up Index references, Bibliographic entries, and footnotes. Why should the Web be any easier?

But superabundance is not the only problem. Fortune magazine (March 1<sup>st</sup> 1999) did their own test of search engines and came up with several examples: for instance, searching for “hockey”, Lycos gave “SuperBowl.com: the official website of SuperBowl XXXIII” (for those who don’t know, SuperBowl is an American Football tournament) as its first hit! So inaccuracy is another problem.

Luckily for us, solutions are being developed. The Guardian newspaper recently reported on a website (<http://www.teem.org.uk>) called “Teachers Evaluating Educational Multimedia”, which contains reviews of software by teachers. Another issue of the same paper refers to the Virtual Teaching Centre on the National Grid for Learning website (<http://vtc.ngfl.gov.uk>), where teachers can dip into additional resources set up by local education authorities in the UK, and the Learning Resource Index (<http://www.ngfl.gov.uk>), which is a directory of educational resources, products and services. Some of these sites may be restricted to UK members, but apparently even Bill Gates is trying to help us: Microsoft is investing heavily in “Adaptive Probabilistic Concept Modelling”, software which identifies the concepts or ideas behind a text, remembers sequences of texts that you have looked at in previous searches, and tries to filter incoming data accordingly! Another recent newspaper article tells us about the increasing number of educational software retail outlets where members of the public can browse the electronic products and evaluate them before deciding whether to purchase them or not.

## **6. Proposed Temporary Solution**

Meanwhile, is there nothing we ourselves can do? I would like to propose a temporary solution. Each academic institution should build up an evaluated list of websites, to which all members of the institution would add the results of their own experiences, especially students. Indeed, as our students are now often more comfortable with computers than the staff, we should utilise their enthusiasm, experience, and ingenuity. Just as students are shown the library and how to find books in it, we should show them how to use the Internet and ask them to record sites of academic or pedagogic worth. And



then we can share the information with similar-minded institutions, and also share the task of verifying and evaluating the websites.

## **7. Postscript**

I realise that I may – unintentionally and inadvertently – have put off some colleagues who have become interested in using electronic resources for their teaching and learning, by focussing on the problems involved. For those colleagues, who may be benumbed by the awesome advances in Internet technology, and feel like a rabbit trapped in the headlights of an onrushing car, there are a few simple points which may help to ease their anxieties. I summarised them as follows in my recent Sogndal course:

1. Computer technology is here, so why not make use of it? Computers have become part of our daily lives in the past decade, in our homes, schools, shops, and offices. Many of us use computers to write letters, to email friends and colleagues, to search Websites for information, and perhaps even to do our accounts, to produce course notes, or school timetables. Why not also use them in our teaching?

2. The pace of change may itself be one of the problems. Computer technology continues to progress at an incredible rate. From mainframes to desktop machines, laptops, palmtops, and notebooks. Processor speeds have vastly increased. New formats and media: from floppy disks to CD-Roms, zip-drives, writable CDs and DVD (Digital Video Disk). We may be worried that things we learn about today may be obsolete tomorrow. But our students will often be more comfortable with computers than we are. We can utilise their enthusiasm, experience, and ingenuity.

3. There are two main approaches to using computers in the classroom. In computer-assisted language learning (CALL) systems, the computer is actually used as a surrogate teacher. In the data-driven learning (DDL) technique, the computer acts as an informant. The teacher's role is more like a research supervisor.

4. The use of corpora in language learning is increasing. A corpus is a structured collection of language texts, and it can be used for various purposes: providing examples, checking existing reference materials (dictionaries, grammars, etc), generating exercises, raising language awareness, etc.

5. In general, students seem to have a positive learning experience with corpora. The impact of seeing language data on a computer screen is more immediate and the practice of discerning patterns of language use oneself seems to have a deeper and more long-lasting effect than the traditional methods of learning rules and trying to understand abstract explanations. Students respond well to the inductive method, moving from observation of the data to classification and generalization.

6. Another advantage of using corpus data is that lexis, grammar and other linguistic features are presented together, not as isolated entities (as in traditional coursebooks, dictionaries and grammars). This is a more accurate and more holistic view of language.

7. Here is a brief summary of the reasons for using computers/corpora for language studies:

a) accuracy – printed books have to be brief, so often leave many questions unanswered

b) comprehensiveness – especially for non-native teachers, access to a wider range of language

c) speed – no need to look up several books separately (e.g. dictionary, grammar, coursebook) or look in several different places within a book (using contents page, cross-references, or index)

d) repetition – tasks can be repeated instantly, so checking and validation are easier

e) access – many people can use the same data at the same time

8. When analysing texts, computers can do most of the tasks you can do manually, but can do them more quickly and more accurately. But computers can also enable you to do types of analysis that you wouldn't have thought of doing before.

So may I encourage any diffident colleagues to try out some of the techniques and strategies I have suggested, and I am confident that within a few weeks they will begin to realize that we can – and must – harness the power of the Internet, and the growing abundance of electronic services, and use them to enhance and expand our range of teaching and learning opportunities!