

ASTON UNIVERSITY

MSc TESOL – VIRTUAL SUMMER SCHOOL JULY 2008

RESEARCH METHODS 2: USING CORPORA [also suitable for the GLE module]

Ramesh Krishnamurthy

The materials are in 5 parts:

PART A: I think you should all read this first, as it gives you the background.

PART B: has 4 simple but engaging paper-based tasks, which can be fun to do and can be adapted for a variety of purposes. It also offers opportunities to use two publicly accessible free corpora (BNC and HarperCollins), which you may want to use later for your own purposes (if you get bitten by the corpus bug!).

PART C: has 8 quick online tasks using the ACORN corpus (which we think has a very user-friendly interface), and introduces the key corpus functions available in ACORN at the moment: WORD FREQUENCY lists, PHRASE FREQUENCY LISTS, and CONCORDANCES.

PART D: has links to corpus sites and other relevant links.

PART E: Suggested Reading, for those who want to know about Corpus Linguistics in greater depth.

PART A: INTRODUCTION

1. Why do we bother with a corpus?

- Corpus research has shown that many of our linguistic intuitions are not correct:
 - * “Language users cannot accurately report language usage, even their own” [Sinclair, J. (1987) Introduction, in the *Collins Cobuild English Language Dictionary*, London: Collins]
 - * “There are many facts about language that cannot be discovered by just thinking about it, or even reading and listening very intently” [Sinclair, J. (1995) Introduction, in the *Collins Cobuild English Dictionary*, London: HarperCollins]
 - * “Using a language is a skill that most people are not conscious of; they cannot examine it in detail, but simply use it to communicate” [Sinclair, J. (1995) Introduction, in the *Collins Cobuild English Dictionary*, London: HarperCollins]
- For hundreds of years, astronomers and sailors observed the stars in the sky with their eyes. Astronomers were able to construct satisfactory theories, and sailors were able to navigate over large distances successfully. However, once telescopes were invented, astronomers and sailors realized that many of their theories and assumptions were wrong, e.g. that the Earth moved round the Sun and not vice versa, and the Earth was round and not flat.

Similarly, corpus research has shown that although we may speak, read and communicate very well, our intuitions about language are in fact unreliable. What we *think* we have heard or read is often very different from what we have actually heard or read. (Indeed what we think we ourselves have said or written is often different from what we have actually said or written!)

2. Widening use of corpora

- From the 1960s to the 1990s, corpora were used mainly for research, in language description and lexicography.
- In the past decade, they are increasingly being used in other fields such as translation, forensic linguistics, and computational linguistics... and for language learning and teaching.

3. How do we learn and teach languages?

- We learn our mother tongue by experiencing thousands of examples and gradually working out any rules/patterns for ourselves [= 'inductive'] if we want/need to, in a wide range of texts and situations, over a long period of time.
- BUT we teach other languages through more abstract information (such as grammar rules and dictionary definitions) and our students must learn to apply these abstract rules and definitions in order to understand and generate instances [= 'deductive'], in speech or writing, on the basis of far fewer examples, limited in variety, over a shorter time period.
- Corpora may reduce this 'exposure deficit' by concentrated exposure to a variety of examples, and offer exciting new methods for language learning and teaching.

4. Changes in language learning and teaching

- goals: from intellectual and literary to practical and communicative (business, media, tourism)
- language content: from made-up examples and classroom dialogues to authentic, real-world communication
- learning and teaching methods: from passive to active learning; from teacher-focus to learner-focus and learner autonomy; use of computer technology

Corpora are well-suited to these changes!

5. How can corpora help language learning and teaching?

- flexible additional resources, with alternative methods (discovery procedures, quantitative approaches, etc), and a wide variety of texts.
- more examples than dictionaries.
- allow users to see common and typical patterns of language use.
- enable users to discover the different ways that words/phrases are used by different speakers/writers in different contexts, text genres, and registers.

6. How can corpora help students studying other subjects?

- Differentiate the concepts and terms of a particular discipline from general language use.
- Examine the opinions of experts in their discipline on the key concepts and terms.
- Analyse the way experts in their discipline discuss the key concepts and terms.
- Compare the different styles in which texts (especially academic texts) are written in different disciplines.

- Identify suitable quotes for use in their essays, assignments, etc

PART B: PAPER TASKS

Task 1: using sentence-length examples from a corpus

Note: <http://sara.natcorp.ox.ac.uk/lookup.html> - gives free but limited access to the British National Corpus (= BNC), a corpus of 100 million words of British English texts, written and spoken, from c. 1994. You can see a maximum of 50 (automatically and randomly selected) sentence-length examples per search.

Q1. I searched for a word in the BNC. Here is the first example I obtained. Can you guess which word I searched for?

HH3 2316 They report to work at 8.30am on an empty stomach.

Q2. Here is another example for the same word. Can you work out which word it is now?

G2Y 200 As English Heritage do not like us, it is with some glee that I report dissent in their own ranks.

Q3. Can you find the same word in this next example?

ADB 39 While the younger members of the family are treated by the British press and public as stock characters in some soap opera, when one of them speaks, on architecture, AIDS or literacy, the news media none the less report their comments seriously.

Q4. And in this final example?

B1E 308 Using data collected over a 2-year period from plots under a variety of management techniques, ranging from maize cultivation to a range of pasture grasses, they report that soil erosion and nutrient losses were greatest from maize (*Zea mays*) and guinea grass (*Panicum maximum* Jacq.) plots.

Note: It is often difficult to find the search word in long, sentence-length examples.

Q5. What do the codes HH3 2316, G2Y 200, ADB 39, and B1E 308 mean?

Go to <http://sara.natcorp.ox.ac.uk/lookup.html>, look up a word yourself, and find out.

Q6. What else can you search for at this BNC site, apart from single words?

Note: A language learner may come across/notice a particular word (or a particular use of that word; or a phrase) one day - but may not come across/notice the same word again for another week, a month, a year? By that time, they have often forgotten the first encounter. A corpus can show you many examples of the same word (or use of the word, or phrase) at the same time. Concentrated exposure = more impact?

Task 2: using KWIC (Key Word In Context) format examples from a corpus

Note: <http://www.collins.co.uk/Corpus/CorpusSearch.aspx> - the Collins' Corpus Concordance Sampler gives free but limited access to 56 million words of English texts from the mid-1990s. You can see a maximum of 40 (automatically and randomly selected) examples per search. But you can also search spoken and written texts separately, and British English and American English texts separately. The examples

are displayed in KWIC format – the keyword is in the middle of each example, with a few words of context on either side. The examples are therefore NOT always full sentences. But you can choose how long the examples are (up to a maximum of 250 characters). **NB you must ‘enable pop-ups’ in your browser in order to see the results of your query!**

This is a typical **CONCORDANCE**:

```
to make a sturdy coffee table with plenty of room for
others who sit round her table include the business men
MACPHERSON : Now on the table here we've got what looks
  a loveseat, a coffee table two large cardboard
suitcases lay open on a table. The silver had already
1818? [p] Answer: From Table C, we find the year 1816
  I lowered my tray table from the back of the seat
end of the dark shiny table and toasted each other in
the size of a snooker table. Another room upstairs is
```

Note: In KWIC format, the search word is completely obvious, but we now have the problem of having to deal with bits of text, text fragments, incomplete sentences. This may seem strange at first, but if you think about it, it is in fact quite natural. We actually learn a lot of our L1 not from complete sentences but from exactly such text fragments. Our attention is partial and sporadic. We overhear a bit of our parents’ conversation, or of a phone call, a radio or TV programme. We glance at a bit of text on a piece of paper or in a book. And even now as adults, we process text fragments all the time: we overhear conversations in the street, glimpse signs or adverts from a moving bus or train, see a bit of a torn poster on a wall, or a notice defaced by graffiti. We skim-read, we flick channels with the remote control, we surf the net. We probably process many more text fragments than full texts.

Q1. Take any KWIC concordance example, and try to guess the word before the beginning of the example, and the word after the end of the example. You can use one of the examples above, or go to the Collins website and choose one for yourself, or start with the example of ‘happen’ I have selected:

...?... upon these cans for food. This did not **happen** in counties of North Georgia, where the rivers ...?...

Note: The focus here is not so much on getting the ‘right’ answer, as on the clues in the fragment we can see, which led you to guess that particular word or phrase. Indeed, it is arguable whether there is a ‘right’ answer. After all, we are looking at fragments of texts written or spoken by a variety of authors/speakers. The fragment we look at may reflect a highly individual/idiolectal wording. But in order for text – for language in general – to be a successful medium of communication, there must be a good degree of predictability involved as well. Highly creative texts (eg poetry) must depend on us being aware to some extent of the consensus norms against which it rebels.

I often play a similar game with my 12-year-old daughter, when she reads to me at bedtime. I stop her just before she turns the page of the book, and ask her to guess the word or phrase on the next page. She often guesses right – sometimes not exactly, but

this often leads to an interesting chat about her guess and the text, the author's choice of wording, etc.

Task 3: An easy-to-create corpus exercise: gapped concordances

Simply take a few examples from a KWIC concordance, and delete the key word.

Q1. Now look at one line at a time, slowly. What are the clues in the context which indicate the (deleted) key word?

Q2. How many lines did you need to examine before you were sure of the key word?

Q3. Which was the conclusive clue?

Q4. What is the key word?

to make a sturdy coffee	with plenty of room for
others who sit round her	include the business men
MACPHERSON : Now on the	here we've got what looks
a loveseat, a coffee	two large cardboard
suitcases lay open on a	The silver had already
1818? [p] Answer: From	C, we find the year 1816
I lowered my tray	from the back of the seat
end of the dark shiny	and toasted each other in
the size of a snooker	Another room upstairs is

Task 4: Another gapped concordance exercise

Note: Remember - look at one line at a time, slowly.

Q1. For each line, what are the clues in the context which indicate the (deleted) key word?

- so it's generally best to feed a small twice a day. Similarly, growing an

fascinate me. You can tell how a is feeling by his tail. If his tail is up,

of days later I'm watching TV and this food commercial comes on an it's the

Year. For what remains of the Year of the, the Tiger should remain committed

he was with a woman walking a black On
neither occasion did the man speak to

I'm prepared if a policeman sees me with my
off a lead in a public place

Note: If you want to continue experimenting with
<http://www.collins.co.uk/Corpus/CorpusSearch.aspx> - the Collins' Corpus
Concordance Sampler, it is worth taking two further steps:

1. At the top of the page you will find the rubric:

“The Collins Wordbanks *Online* English corpus is composed of 56 million words of contemporary written and spoken text. To get a flavour of the type of linguistic data that a corpus like this can provide, you can type in some simple queries here and get a display of concordance lines from the corpus. The [query syntax](#) allows you to specify word combinations, wildcards, part-of-speech tags, and so on.”

Look at the **Query Syntax**, and you will find that you can do some more detailed searches. You can specify the wordclass of the search word, or specify compounds, or look for all the inflected forms of a word together, etc.

2. On the same page, just below the Corpus Concordance Sampler, you will find the **Collocation Sampler**. Collocation is the linguistic phenomenon that certain pairs of words tend to be used together for reasons that seem to be above and beyond the requirements of syntax (grammar) and semantics (conventional ‘meaning’). This is an area of intense research by corpus linguists at the moment. You will see that you have a choice of Mutual Information or T-score to calculate the statistical ‘significance’ of the co-occurrence of two words. In this particular software (and generally accepted as a result of John Sinclair’s collocation experiments in the 1960s), co-occurrence is restricted to a distance (‘span’) of 4 words from the key word in a text. But remember that the collocation lists a) do NOT tell you which wordclass the collocating word (collocate) is being used in, and b) do NOT tell you which position the collocating word occurs in within that +/-5 context. For example, if we find that *hard* is in the collocate list of *work*, this may include examples of *hard work*; *work hard*; *the work was extremely hard*; and *it is hard to say which type of work I enjoy doing most*, etc.

PART C: ONLINE TASKS using ACORN (the Aston Corpus Network)

Logging in to ACORN:

1. Open an Internet Browser (preferably Mozilla Firefox (the “Edit – Find in This Page – Highlight all” function is very useful; but Internet Explorer is OK).
2. Delete current URL and type: <http://acorn.aston.ac.uk>
3. Use your normal Aston-id and password to login.
4. Read the next set of instructions before proceeding in ACORN.

Task 5: using a Frequency List

You are going to teach English to medical students. You want them to be able to write abstracts for their research papers. First, you want to check their knowledge of relevant vocabulary.

Q1. How would you usually find this vocabulary?

- **Click on English (Language selection)**
- **Click on the box next to 2. Brown Corpus (Corpus selection) and then click Continue>>**
- **Click on Word Frequencies**
- **Click on Top 50**

Q2. What do you notice about the first few words in the frequency list?

Q3. Make a list of the content/vocabulary words

The Brown Corpus is a corpus of GENERAL ENGLISH (written American texts from 1961), so grammar/function words are the most frequent words. (NB if you look at the Top 100 frequency list, you will find *said, time, new, two, man, years* lower down in the list). So the Brown corpus would not be a very useful starting point!

...but luckily for you, ACORN also has a corpus of medical abstracts! ☺

- **Click on English (Language selection)**
- **Click on the box next to 10. Medical Abstracts (Corpus selection)**
- **Click Continue>>**
- **Click on Word Frequencies**
- **Click on Top 50**

Q4. What do you notice about the first few words in the frequency list?

Q5. Make a list of the content/vocabulary words

This time, you should have a small but good selection of words to start your teaching with.

Q6. Can you group these words in any way?

Q7. If you click on Word Frequencies again, and this time on Top 100, can you assemble better groups?

Task 6: comparing domain-specific word frequency lists

- Click on Corpus Selection (top left, underlined)
- Click on the boxes next to 10, 11, and 16 (for Medical, Business, and Labour Party = Politics)
- Click Continue>>
- Click on Word Frequencies, then Top 50

Q1. You already have a list of content words for Medicine from the previous task. So now make a list of the content words in the Business and Politics corpora

Q2. Which words are unique to one domain?

Note: this is where the “Edit – Find in This Page – Highlight all” function in Mozilla Firefox is very useful.

Task 7: comparing phrase frequencies in domain-specific corpora

- Click on Phrase Frequencies (top, underlined)
- Click on 4-grams (= 4-word phrases)
- Click on Top 50

Q1. Which phrases occur only in one domain?

Q2. Which phrases occur in more than one domain?

Note: this is where the “Edit – Find in This Page – Highlight all” function in Mozilla Firefox is very useful.

Task 8: looking at concordances, and sorting

- Click on Corpus Selection
- Click on Brown Corpus
- Click Continue>>
- Click on Concordance
- Type in *same* and click Search
- Click on Display Results

Q1. What words you notice to the left and right of same?

- Click on Sort Results
- Click on ‘1st word before keyword’ in Main Sort line
- Click on the grey box Sort Results below

Q2. What do you notice now?

Q3. Which words in red are NOT the?

- Click Sort Results again and ‘1st word after keyword’ in Main Sort line

- **Click on the grey box Sort Results below**

Q4. What do you notice now?

- **Click on Corpus Selection**
- **Click on the box next to 10. Medical Abstracts (Corpus selection)**
- **Click Continue>>**
- **Click on Concordance**

Q5. Select one of the content words we extracted from the word frequency list, look at the concordances, sort them to left and right, and report back on what you found.

Q6. Which phrases/grammatical patterns would you teach your medical students?

PART D: ACCESS TO CORPORA , SOFTWARE, or BOTH:

1. You can use an existing corpus with its own software, by visiting a website, e.g. free online English corpus access is provided by:

HarperCollins:

<http://www.collins.co.uk/Corpus/CorpusSearch.aspx> [Collins - English only; search written, spoken, British, American separately; KWIC format; max 40 examples per search; can specify wordclass, and a few other features; collocation lists also available; 56 Million Words]

Oxford-BNC (British National Corpus):

<http://sara.natcorp.ox.ac.uk/lookup.html> [British English only; from c 1994; sentence-length examples only, not KWIC format; max 50 examples per search; 100 Million Words]

Brigham Young University-BNC:

<http://corpus.byu.edu/bnc/x.asp> [a better place to access BNC; KWIC format concordances, etc]

Corpus of American English:

<http://www.americancorpus.org/>

2. Or use several existing corpora using the same software:

ACORN (Aston Corpus Network):

<http://acorn.aston.ac.uk/> [only accessible by Aston staff and students at the moment; use your normal Aston-id and password; several sets of data for English, French, German and Spanish; also parallel texts – source texts and their translated versions; word and phrase frequencies, and KWIC format concordances; c 200m words altogether]

VLC, Hong Kong:

<http://www.edict.com.hk/concordance/> [English, French, Bilingual, Chinese, Japanese]

Sketch Engine:

<http://www.sketchengine.co.uk/> [free 30-day trial if you register; then modest annual payment required; access to several corpora via the same software]

Brigham Young University:

<http://corpus.byu.edu/>

The Compleat Lexical Tutor [very teacher and learner focussed

<http://132.208.224.131/>

3. There are other types of corpus-derived information at other websites:

Academic Word List:

<http://language.massey.ac.nz/staff/awl/awlinfo.shtml>

Frequency Lists of various kinds (written, spoken, lemmatised, by wordclass, etc) for BNC:

<http://www.comp.lancs.ac.uk/ucrel/bncfreq/flists.html>

4. You can even access large amounts of the Web as a corpus:

- WebCorp

<http://www.webcorp.org.uk>

- Web as Corpus

<http://webascorpus.org/searchwac.html>

- Basque Web as Corpus query system

<http://www.corpeus.org/>

- KWICFinder

<http://www.kwicfinder.com/KWiCFinder.html>

5. You can collect your own corpora (your students' writing; domain-specific, topic-specific, or genre-specific texts from websites; downloadable corpora at some websites; Gutenberg Project out-of-copyright literature in many languages, etc) **and download existing software to analyse them with (eg **AntConc** – free – for PCs; **Conc** – free, for Macs; **Wordsmith Tools** – for PCs – costs c £40 for single user)**

6. Sites with general information about corpora, including detailed descriptions and comprehensive lists with links to many websites giving corpus access, and to corpus software (also online courses, publications, journals, conferences) etc:

- David Lee's Corpora Bookmarks

<http://devoted.to/corpora>

- Developing Linguistic Corpora: a Guide to Good Practice

<http://ota.ox.ac.uk/documents/creating/dlc/index.htm>

- Gateway to Corpus Linguistics

<http://www.corpus-linguistics.de/>

- Corpus Linguistics

http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/introduction.html

- Corpus Linguistics

<http://www.lancs.ac.uk/fss/courses/ling/corpus/>

7. Tim Johns' website, with many exercises and useful links:

<http://www.eisu2.bham.ac.uk/johnstf/index.html>

PART E: SUGGESTED READING

1. Unit 2 of the GLE module.

2. From Reading List in Unit 2 of GLE:

Butterfield J & Krishnamurthy R. 2000. Beyond the Dictionary: On-line Learning in the Classroom. *TESOL Spain Newsletter*, Vol 23 (and 2001: *TESOL France Newsletter*, No 28)

Firth JR. 1957. Modes of Meaning. In *Papers in Linguistics 1934-51*. London: OUP.

Fox G. 1987. Chapter 7: The Case for Examples. In Sinclair JM (ed). *Looking Up*. London and Glasgow: Collins ELT

Hanks P. 2000. Do Word Meanings Exist? *Computers and the Humanities* 34: 205-15. (also available at <http://www.patrickhanks.com/papers/wordMeaning.pdf>)

Hunston S & Francis G. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam/Philadelphia: John Benjamins

Krishnamurthy R. 1996. Ethnic, Racial and Tribal: The Language of Racism? pp. 129-149 in Caldas-Coulthard CR & Coulthard M. *Texts and Practices*. London: Routledge

Krishnamurthy R. 1997. Keeping Good Company: collocation, corpus and dictionaries. pp. 31-56 in Cabre MT, Castellvi I & Lorente M (eds). *Cicle de conferencies 95-96: Lexic, corpus i diccionaris*. Barcelona: IULA

Krishnamurthy R. 2000. Collocation: from silly ass to lexical sets. In Heffer C & Sauntson H (eds). *Words in Context: A Tribute to John Sinclair on his Retirement*. Birmingham: Birmingham University

Krishnamurthy R. 2001. The Science and Technology of Corpus, and Corpus for Science and Technology. In De Cea GA & Escibano PD (eds). *La Investigacion en Lenguas Aplicadas: Enfoque Multidisciplinar*. Madrid: Fundacion Gomez Pardo
Sinclair JM, Jones S & Daley R. 1970. *English Lexical Studies*. Published (2004) as Krishnamurthy R (ed). *English Collocation Studies*. London and New York: Continuum.

Sinclair JM. 1996. The Search for Units of Meaning. *Textus* IX: 75-106

Stubbs M. 1995. Collocations and Semantic Profiles: On the cause of the trouble with quantitative studies. *Functions of Language* 2, 1

Willis D. 1990. *The Lexical Syllabus: A New Approach to Language Teaching*. London: HarperCollins

available via your Athens password:

Gavioli L & Aston G. 2001. Enriching reality: language corpora in language pedagogy. *ELT Journal* Vol 55/3 July

Mishan F. 2004. Authenticating corpora for language learning: a problem and its resolution. *ELT Journal* Vol 58/3 July

Further Reading:

Fall, L. 2004. *Corpus Linguistics (1): Meaning in Context* (available at <http://www.proz.com/howto/173>)

Fox G. 1998. Using corpus data in the classroom. In Tomlinson B (ed). *Materials Development in Language Teaching*. Cambridge: CUP

Hanks P. 2000. Do Word Meanings Exist? *Computers and the Humanities* 34: 205-15 (also available at <http://www.patrickhanks.com/papers/wordMeaning.pdf>)

Ranalli JM. 2003. *ELT coursebooks in the age of corpus linguistics: constraints and possibilities*. University of Birmingham. (available at www.cels.bham.ac.uk/resources/essays/Ranalli6.pdf)

- Sheehan R. 2005. Language as Topic: Learner-Teacher Investigation of Concordances. In Edwards C & Willis J (eds). *Teachers Exploring Tasks in English Language Teaching*. Basingstoke: Palgrave Macmillan.
- Willis J. 1998. Concordancing in the Classroom without a Computer. In Tomlinson B (ed). *Materials Development in Language Teaching*. Cambridge: CUP

3. From Aston MA Corpus Linguistics module

- Barnbrook G. 1996. *Language and Computers*. Edinburgh: Edinburgh University Press.
- Biber, D., Conrad, S. & Reppen, R. 1998. *Corpus Linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Kennedy, G. D. 1998. *An Introduction to Corpus Linguistics*. Harlow: Longman.
- Landau, S.I. 2001. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Meyer, C. F. 2002. *English corpus Linguistics: an introduction*. Cambridge: Cambridge University Press.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, M. 1996. *Text and Corpus Analysis*. Oxford: Blackwell.
- Teubert, W. & Cermakova, A. 2007. *Corpus linguistics : a short introduction*. London: Continuum.
- McEnery, T. & Wilson, A. 2001. *Corpus Linguistics : an introduction*. (2nd edition). Edinburgh: Edinburgh University Press.

4. Corpus For Language Teaching And Learning

- Aston, G. (1996). Corpora in language pedagogy: matching theory and practice. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of HG Widdowson* (pp. 257-270). Oxford, UK: Oxford University Press
- Aston, G. (1997). Enriching the learning environment: Corpora in ELT. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles, (Eds.), *Teaching and language corpora* (pp.51-64). London: Longman.
- Bernardini, S. (1998). Systematising serendipity: Proposals for large-corpora concordancing with language learners. *Proceedings of TALC98* (pp. 12-16). Oxford, UK: Seacourt Press.
- Braun, S., Kohn, K., and Mukherjee, J. (Eds). (2006). *Corpus Technology and Language Pedagogy: New Tools, New Resources, New Methods*. (English Corpus Linguistics, Vol. 3) Frankfurt/Main: Peter Lang
- Burnard, L., & McEnery, T. (Eds). (2000). *Rethinking language pedagogy from a corpus perspective: Papers from the third international conference on teaching and language corpora (Lodz Studies in Language)*. Hamburg, Germany: Peter Lang.

- Flowerdew, J. (1996). Concordancing in language learning. In M. Pennington (Ed.), *The power of CALL* (pp. 97-113). Houston, TX: Athelstan.
- Granger, S. (Ed.). (1998). *Learner language on computer*. Harlow, UK: Longman.
- Johns, T. (1988). Whence and whither classroom concordancing. In T. Bongaerts, P. De Haan, S. Lobbe, & H. Wekker (Eds.), *Computer applications in language learning* (pp. 9-27). Dordrecht, The Netherlands: Foris.
- Johns, T. (1991a). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal*, 4, 1-16.
- Johns, T. (1991b). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal*, 4, 27-45.
- Mparutsa, C., Love, A., & Morrison, A. (1991). Bringing concord to the ESP classroom. *English Language Research Journal*, 4, 115-133.
- Römer, U. (due 2007). Corpora and Language Teaching. In: Lüdeling, A. & Kytö (Eds). *Korpuslinguistik - Corpuslinguistics. An International Handbook*. [HSK series] Berlin: Mouton de Gruyter.
- Stevens, V. (1991). Classroom concordancing: Vocabulary materials derived from relevant, authentic text. *English for Special Purposes Journal*, 10, 35-46.
- Tribble, C. (1991). Concordancing and an EAP writing program. *CAELL Journal*, 1(2), 10-15.
- Tribble, C. (1997). *Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching*. Paper presented at the first international conference "Practical Applications in Language Corpora," University of Lodz, Poland. Retrieved November 8, 2000, from the World Wide Web: http://ourworld.compuserve.com/homepages/Christopher_Tribble/PALC.html.
- Tribble, C. (2001). "Small corpora and teaching writing: towards a corpus-informed pedagogy of writing" in Rose, R. et al (eds.) *Small corpus studies and ELT*. Amsterdam: Benjamins.
- Tribble, C. (2001). "Corpora and corpus analysis: new windows on academic writing" in Flowerdew, J. (ed.) *Academic Discourse*. Harlow: Addison Wesley Longman.
- Tribble, C., & Jones, G. (1997). *Concordances in the classroom: Using corpora in language education*. Houston, TX: Athelstan.
- Tribble, C. & Barlow, M. (2001). Using Corpora in Language Teaching and Learning (Special Issue of) *Language Learning and Technology* 5,3; available at: <http://llt.msu.edu/vol5num3/default.html>

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

5. Recent publications available via Aston Library (up to 3/12/07):

O'Keeffe, A	2007	From corpus to classroom	CUP
Baker, P	2006	Using corpora in discourse analysis	Continuum
Brown S, Kohn K, Mukherjee J	2006	Corpus technology and language pedagogy	Peter Lang
Lemnitzer L, Zinsmeister H	2006	Korpuslinguistik: eine Einfuhrung	Gunter Narr
Hoey, M	2005	Lexical priming	Routledge
McCarthy M, Sampson G	2005	Corpus linguistics: readings in a widening discipline	Continuum
Mahlberg, M	2005	English General Nouns: a corpus theoretical approach	Benjamins
Mitkov R	2005	The Oxford handbook of computational linguistics	OUP
Siepmann, D	2005	Discourse markers across languages	Routledge
Sinclair, J	2004	Trust the text	Routledge
Sinclair J, Jones S, Daley R	2004	English Collocation Studies	Continuum
Schmitt, N	2004	Formulaic sequences: acquisition, processing and use	Benjamins
Halliday MAK, Cermakova A, Teubert W, Yallop C	2004	Lexicology and corpus linguistics	Continuum
Halliday MAK	2004	The language of science (vol 5: collected works)	Continuum
Sinclair, J	2003	Reading concordances: an introduction	Pearson/Longman
Barnbrook G, Danielsson P, Mahlberg M	2003	Meaningful texts	Continuum
Kuiper, K	2003	An introduction to English language	Palgrave Macmillan
Willis, D	2003	Rules, patterns and words: grammar and lexis in English language teaching	CUP
Simon- Vandenberg AM, Taverniers M, Ravelli L	2003	Grammatical metaphor: views from systemic functional linguistics	Benjamins
Granger S, Hung J, Petch-Tyson	2002	Computer learner corpora, second language acquisition and foreign language teaching	Benjamins

S			
Hunston, S	2002	Corpora in applied linguistics	CUP
Thornbury, S	2002	How to teach vocabulary	Longman
Landau, S	2001	Dictionaries: the art and craft of lexicography	CUP
Bronner M	2000	Bronner's rhyming phrases dictionary	New Edge