



University of HUDDERSFIELD

University of Huddersfield Repository

Karpenko-Seccombe, Tatyana

Integrating Corpus analysis into everyday classroom practice

Original Citation

Karpenko-Seccombe, Tatyana (2015) Integrating Corpus analysis into everyday classroom practice. *InForm*, 15. pp. 18-20.

This version is available at <http://eprints.hud.ac.uk/id/eprint/33166/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

First steps to integrating Corpus analysis into everyday classroom practice

Abstract

Language teaching is an exciting area for applying corpus linguistics. This article gives a glimpse into practical applications of Concordancers in academic writing and provides examples of Concordance exercises that can be used in the classroom to introduce students to working with corpora.

Dr Tatyana Karpenko-Seccombe

t.karpenko-seccombe@hud.ac.uk (work); tatyana@gogol.f9.co.uk (home)

15 Garden Terrace, Hebden Bridge, West Yorks, UK HX78BL

University of Huddersfield

Senior Lecturer in English for Academic Purposes

First steps to integrating Corpus analysis into everyday classroom practice

The aim of the article is to highlight the importance of data-driven learning for developing student autonomy by empowering and motivating students to do linguistic mini-researches that lead, in turn, to the discovery of important linguistic phenomena. The article presents several ways in which Concordancers and corpora¹ can be used in everyday classroom practice within the context of developing student autonomy.

The classroom use of Concordancers is largely based on the principle of Data Driven Learning (Johns 1994) and discovery or serendipity learning (Bernardini 2000), one of the main tenets of which is the idea that students are more likely to retain and recall information which they discover for themselves as a result of collocational analysis supported by grammatical and frequency analyses.

In language teaching these same methods have also started to take hold. The concept of autonomous language learning was famously pioneered by Holec, who defines autonomy as 'the ability to take charge of one's own learning' (1981, p. 3). Krashen in 1981 started promoting autonomy in language acquisition and learner control by searching for, and discovering, new material and resources as well as choosing the time, place, intensity and pace of learning.

The uses of Concordancers may range from very simple tasks like checking the correct use of prepositions or place of connectors, to more sophisticated tasks, for example, exploring word collocations, making decisions about correct word usage, analysing and using information about word frequencies.

In this article I will describe some easy ways in which Concordancers can be introduced into the EAP classroom. These ways include checking prepositions and common phrases, getting help with some grammatical issues, for example singular and plural forms, finding examples of authentic collocations.

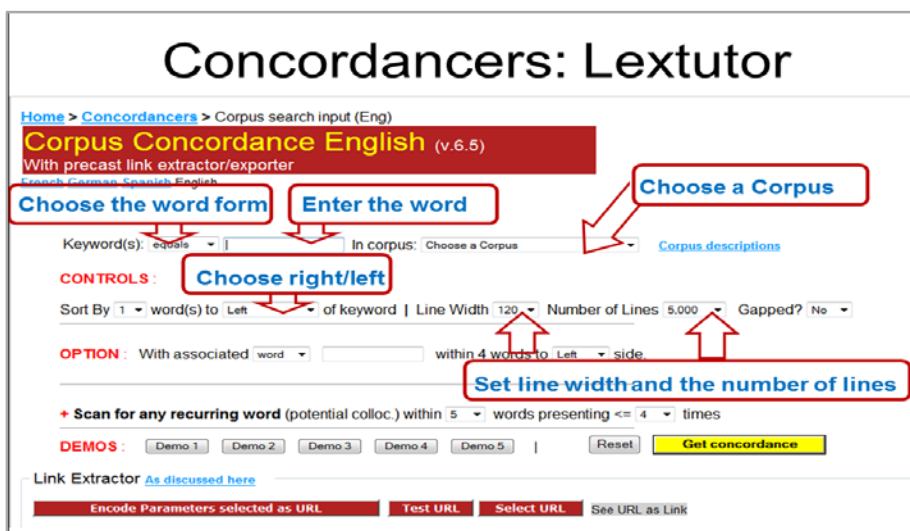
Corpora and Concordancers

Among a striking variety of Concordancers and corpora available now for teachers and learners, I find The Compleat Lexical Tutor <http://lextutor.ca/conc/eng/>, Sketch Engine open access <https://the.sketchengine.co.uk/open/> and COCA (Corpus of Contemporary American English, academic) <http://corpus.byu.edu/coca/> particularly useful for the purposes of my students because they are user-friendly, open access, easy to teach and to learn, and free.

¹Corpora – large sets of texts, e.g. British National Corpus which was created in the 1980/90s and comprises over 100 million words

Concordancer – a computer programme which allows users to browse corpora showing the key-word usage in multiple contexts

The essential first step is to introduce students to KWIC (key-word in context) search.

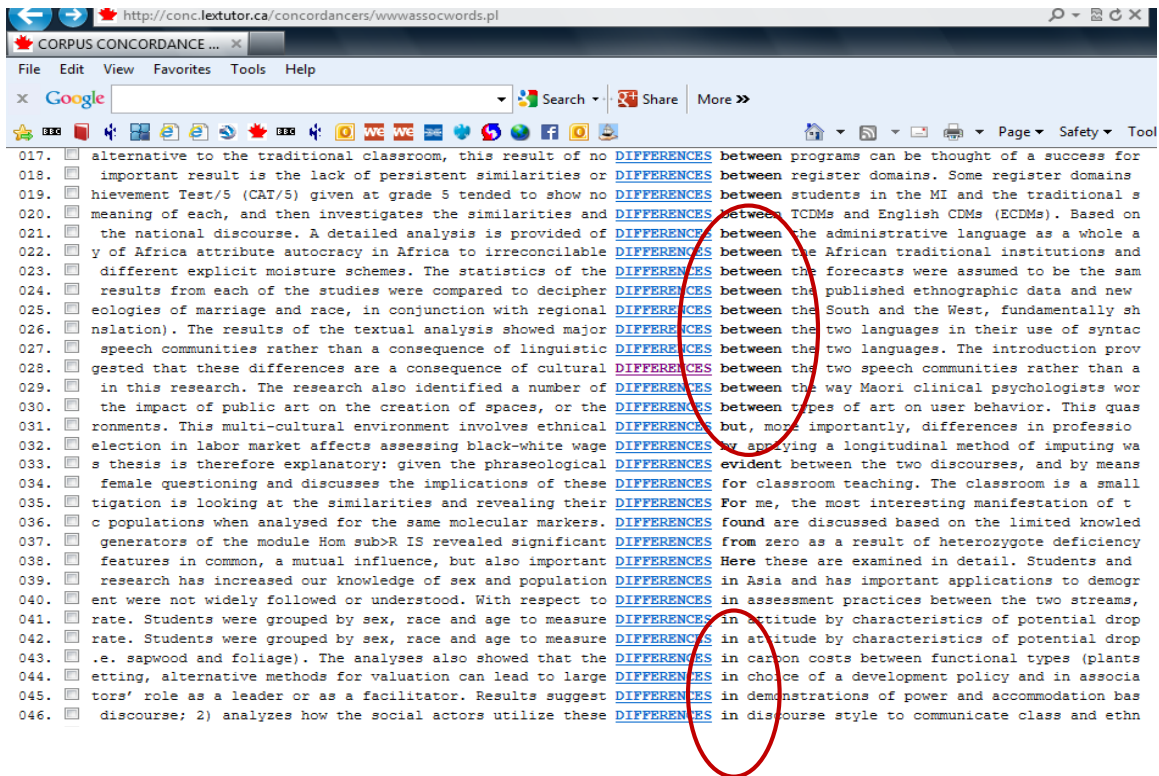


Sorting out simple language problems

Prepositions

Mistakes in the use of prepositions are a recurrent problem for students of English. Giving students a quick and effective reference tool usually helps in demonstrating the benefits of using this programme to help with writing. Students can be given an initial task:

Task: what prepositions are used after *difference* - ? and *different* - ?



049. ergy. inus, in a comparative study of river basin planning, DIFFERENT environments can be contrasted, but similar functi
050. e and general impressions of Chemawa are categorized in the DIFFERENT eras and serve as the body of the thesis. Despite
051. development phase and depending on the level of uncertainty DIFFERENT evaluation methods are available to analyze the ec
052. ed using differences between model forecasts which utilized DIFFERENT explicit moisture schemes. The statistics of the c
053. with fiction, hard science and soft science represented in DIFFERENT factors. Strong patterns were also identified in c
054. nos, and Native Americans, can be attributed to a number of DIFFERENT factors. They include exclusionary pedagogical and
055. t tipped the equilibrium towards unchallenged collusion was DIFFERENT for each case. In the 19th century an exogenous c
056. at the driver of mergers and acquisitions activity are very DIFFERENT for each particular firm--still within the same ir
057. logical evidence for hunter-gatherer occupation of sites in DIFFERENT forest types within south-western Australia which
058. that specify the conditions under which war, democracy, and DIFFERENT forms of authoritarian order are self-enforcing. I
059. of learning through talking with other women, delving into DIFFERENT forms of knowledge, and coming to think in new way
060. Central Asian city developed into a distinct type, markedly DIFFERENT from cities further west and south. The examinatio
061. could not be concerned with a world which was significantly DIFFERENT from God himself. Aristotle's reasoning from the v
062. r the self-creation of a stratum which is not significantly DIFFERENT from God, or the self-destruction of humanity. In
063. derstanding will lead to improvements. Our goal is a little DIFFERENT from that of most algorithmic investigations. Sinc
064. as been defined as the retrieval of documents in a language DIFFERENT from that of the original query. To make this poss
065. tity which would be self-existent, and so not significantly DIFFERENT from the self-subsistent God. Each stage of such s
066. mental conditions in other ways may have been significantly DIFFERENT from those of the present- day do not necessarily
067. he level of frequency modulation, allows the application of DIFFERENT harmonics as high as the 75th harmonic; although s
068. nal banks but how is the role played by information systems DIFFERENT in and for Islamic banks. I studied the existing p
069. iding is ruling", this new vision groups problems seemingly DIFFERENT in nature, and provides, from their analysis, a ur
070. aviors of operational and resulting factors can be seen for DIFFERENT input, of the various initial factors. The model c
071. entative of draped architecture in order to demonstrate the DIFFERENT instances of its paradoxical representations. The
072. e on household decision-making makes no distinction between DIFFERENT institutional processes of household formation, ne
073. nd contrasting faculty perceptions of academic freedom from DIFFERENT institutions, ranks and disciplines, this research
074. ereating manifestation of this exchange between similar and DIFFERENT is in the designed object and designed environment
075. spaces of illusion, artifice and play has to be based in a DIFFERENT kind of experience and spatial practice, one that
076. MI program. Yet, considering that the MI program emphasizes DIFFERENT kinds of activities and more diverse ways of learr
077. the notion of coherence and the feature of having access to DIFFERENT kinds of information which distinguishes our theor
078. ion and computing homology. In several areas of Mathematics DIFFERENT kinds of resolutions have proved useful. For exam
079. tigation included ethnographic studies, which addressed the DIFFERENT land, riverine and marine food groups consumed by
080. basic assumes that translation will leave similar process in DIFFERENT languages. Major corpus based studies have recent

This search usually leads students to the following conclusion:

Difference noun

- *between* (e.g. *differences between two communities*)
- *in* (e.g. *differences in attitude*)

Different adjective

- *From* (e.g. ... *language different from original query*)

It is useful to ask students to run checks on most commonly confused prepositions. Here is another example of a task:

Task: what preposition is used after

insight -?

draw -?

discuss -?

Try several corpora; use *equals* search and *sort to the right*

Record your findings

word	Preposition/s	Examples
insight (n)		
draw (v)		
discuss(v)		

Set phrases

Checking the correctness of commonly confused phrases is also a productive method of introducing students to working with Concordancers. Below is a task which I give students at the initial stages.

Task: choose the correct form of each phrase using Concordancer. Try *families* searches and *equals* options to search separate phrases in several corpora, e.g. Humanities and Medical:

a) *The structure of the first chapter is as following / following / as follows / as follow*

b) *As regard/ regards / as regarding / with regards /as regards/ with regard to housing, employers' policies vary*

c) *As far as the new rent is concerned/ As far as concerning new rent / Concerned new rent/ concerning new rent options fall into four categories*

Make a note of your results:

Correct form	Example

Clarifying some grammatical issues

Plurals

Some common grammatical mistakes in student writing can be addressed by Concordancers searches. One of the common questions students ask is whether *data* is singular or plural. A quick search of Lextutor Concordancer shows approximately similar numbers in the Academic General Corpus and in BNC Humanities, however, the BNC Medical Corpus shows definite preference for plural usage. This can be a good way of introducing linguistic differences in academic writing in different subject areas. Differences in the usage of *data* were highlighted by Tim Jones in his Kibbitzers: "The choice of plural or singular ... also correlates with the **meaning** of the word. The (traditional) meaning "*evidence used in experimental procedures*" is most often plural, while the (more recent) meaning "*digital information stored or manipulated by a computer*" is most often singular." (<http://lexically.net/TimJohns/Kibbitzer>). As a result, singular usage can be expected to be more common in computing texts.

Another common mistake is using the noun *literature* in plural. A search of singular and plural forms on Lextutor gives the following results. In the Academic General Corpus – there are 1239

hits (206 instances per million) for *literature* and 13 hits for *literatures* (2 per million). The BNC and COCA give similar results: the BNC has 49 instances of the plural as opposed to 5246 for the singular; COCA has 305 against 28,055 for the singular. I then draw students' attention to the usage of the plural form of *literature*, where it mostly means different kinds of literature, or literature of different peoples, but not sources in literature reviews.

Collocations

Collocations are one of the most persistent difficulties of L2 students because of the formulaic nature of the English language (see, for example, Wray 2002). Students can be given tasks to run specific searches for collocations sorting the results to the right or to the left of the key word.

For example, the search for prepositions used with *difference*, as described earlier, can also be used to learn about collocations. By sorting the results of the search to the left, students can find out what adjectives commonly collocate with *difference*:

Adj + *difference* *basic, big, clear, considerable, crucial, cumulative, essential, factual, fundamental, further, genuine, great, important*

A variety of material provided by such searches could result in follow-up tasks, in this case centred round using *adjective + difference* collocations in the students' own writing.

In the same vein, the search on the word *literature* can also reveal its collocations:

Adj + *literature*: *Existing, extensive, previous, recent, relevant, contemporary, early, earlier, analytic, expert, limited and even unexpurgated*

***Literature* + verb**: *reveals, deals, emerges, shows, tells, recycles old material, has been accumulated, has been aimed, has been developed*

Word combinations: *Body of literature, subset of literature, recent strand of literature, broad range of literature, contribution to the literature, theories available in the literature, common perception in the literature, has often been neglected in the literature, adapted from the literature, existing theoretical literature*

To sum up – this article has attempted to give teachers some initial suggestions for using Concordancers in the classroom. The possibilities offered by these programmes are much deeper and richer: they can offer substantial insights into grammar, vocabulary and academic writing style.

Bernardini, S. (2000). Systematising serendipity: proposals for concordancing large corpora with language learners. In: Burnard, Lou & McEnery, Tony (eds), 225-234.

Johns, T. (1994). From printout to handout. Grammar and vocabulary teaching in the context of data-driven learning. In: Odlin, Terence (ed), 293-313.

Johns, T. Kibbitzer <http://lexically.net/TimJohns/Kibbitzer>, accessed September 2015.

Wray, A (2002). Formulaic Language and the Lexicon. Cambridge: Cambridge University Press.

