Transition to teaching with corpora



I teach EAP to international PhD students at the University of Huddersfield



Transitions

A transition from using corpora as a tool mainly for teaching vocabulary and collocations

to teaching wider aspects academic writing through the use of corpora

"Why should L2 [second language] learners learn to use concordance information themselves? One reason is that no one can learn everything there is to know about a language from being told; at some point, the learning depends on observing how the language works and drawing one's own conclusions from observations — this is called data-driven learning."

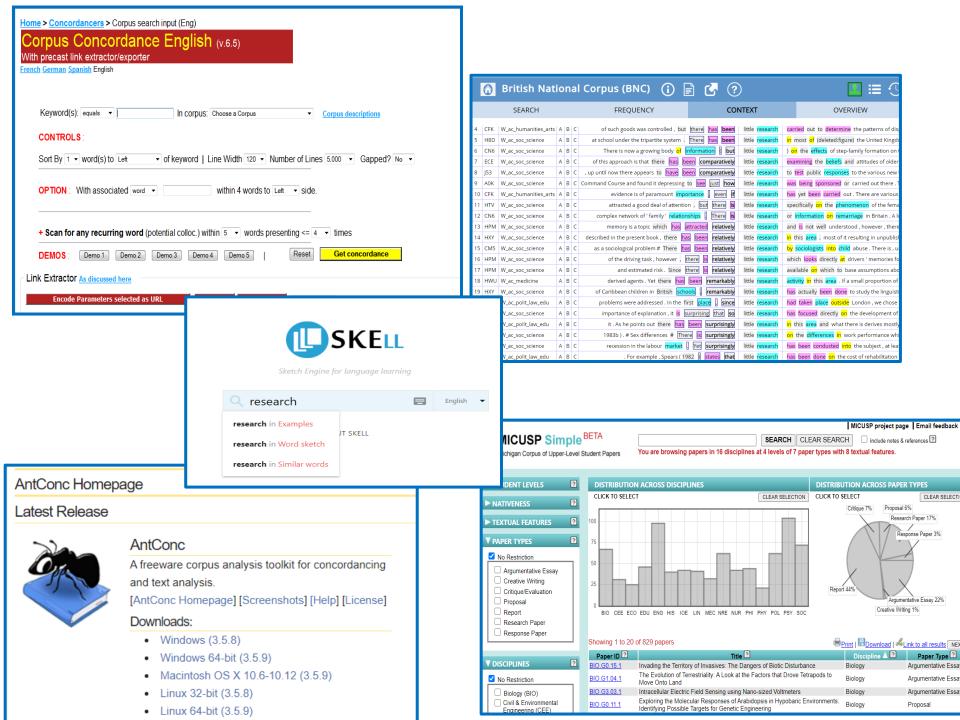
(Cobb, 2010 p. 16)

Beginning: word choice

Using some corpus materials

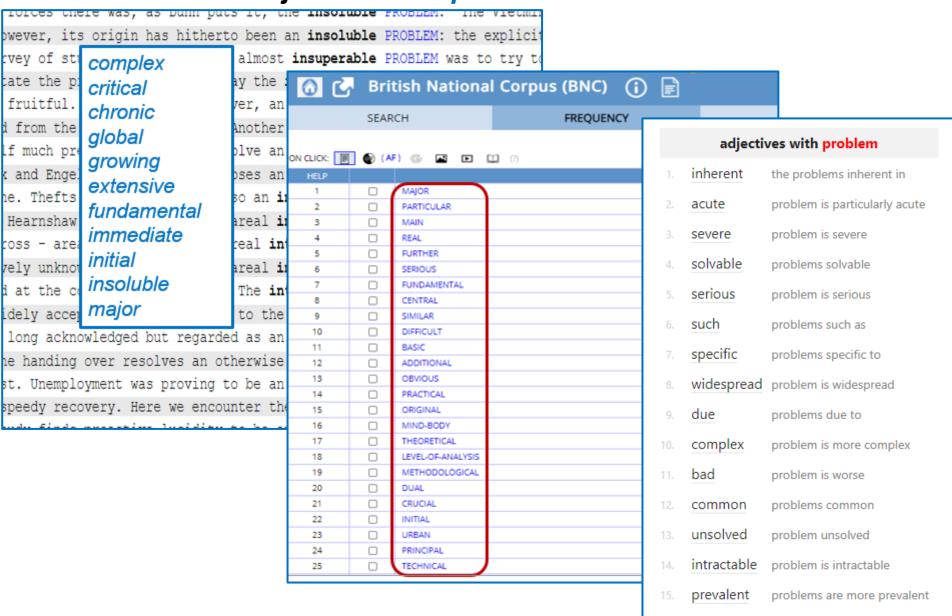
- in the course of Academic vocabulary
- in one-to-one tutorials





This is a perfect solution to a massive problem!

adjective + *problem*



adjective + solution

ngly tw acceptable recipit rom the alternative in othe: critical undaries convincing me after credible until perceived d by Jos definitive h to Arr desirable is was f the t: simple ia had suitable make ef: successful ce of ea temporary of the tentative effect theoretical method -algebr: unsatisfactory eady kno unexplored t attem workable art of

asses. (A satisfactory SOLUTION was not to be reached until the work of A with an 80% saturated SOLUTION of **f, dissolved in a small quantity of ri's pianos. Schantz's SOLUTION to the problem of the check in his grand g settings. The **second SOLUTION** is to regard linguistic knowledge or comp adopt an over - simple SOLUTION to the problem, always preferring one sou s formed: the simplest SOLUTION is to repeat the first phrase exactly as st. Drain off the soap SOLUTION of the suds cycle at such a time that the 1883). Planes of soap SOLUTION have the property that only three can int early anti - socialist SOLUTION". On 15 June, after further demonstration ich required a special SOLUTION. Still, the application of different prin l base and the **states**' SOLUTION was to modernize and diversify (see North apparently successful SOLUTION. Argentina's José Ingenieros, for exampl ew data sets. One **such SOLUTION** is Goodchild and Yang's (1989) scheme for infusion of a sucrose SOLUTION. Control subjects experienced the critical 88), but the suggested SOLUTION - the separation of "purchasers" and "cor entation to a suitable SOLUTION. A number of test areas have been investi o provide a systematic SOLUTION. The proposed methodology is elaborated by e dynamical system. The SOLUTION of linear systems is an ancient and inext he basis of their task SOLUTION of writing down English from sign users. information technology SOLUTION focused on executivels needs. To underst is is only a temporary SOLUTION and there have been many attempts to orga

Word choices

This is a perfect solution to a massive problem!

definitive

successful

effective

optimal

convincing

credible

serious

severe

fundamental

major

extensive

widespread

acute

'One criticism aimed at DDL is that it ignores the "big themes" of language such as tenses'.

(Alex Boulton 2007)

Big theme: Stance



Stance 'refers to the language features writers use to comment on their statements in order to convey their attitudes, opinions and degrees of commitment to what they say...'

(Hyland (2016).

It includes:

- hedges and boosters which increase or decrease the strength of what you are saying;
- evaluative language (or attitude markers) and
- self-mention (Hyland, 2016).



Big theme: strength of claim

Solution

Strong

Cautious

critical

convincing

credible

definitive

successful

effective

optimal

perfect?

acceptable

suitable

temporary

tentative

workable

possible



Big theme: strength of claim

Perfect solution?

Academic General

```
ands. Rome clearly never found a perfect SOLUTION for this problem: eventually practice." Of course there is no perfect SOLUTION. I had the easy choice of remaignation to America seemed like the perfect SOLUTION. Initially, many of those who was he HFEA 1990. Yet the latter is not a perfect SOLUTION since there are still aspects prothough "best practice" seems like the perfect SOLUTION, companies find it hard to implement though. None has found the perfect SOLUTION, but the debates on the '
```

BNC Medicine

Big theme: hedging, avoiding overstatements

red by many Afro-American literary scholars to be THE MOST glamorous and productive period for Afroroa, the ability to be social is considered to be THE MOST important and valued characteristic of hu s which, from Stevenson's point of view, would be THE MOST rational. It is essentially a matter of 1 sacrifice of an animal at the altar is said to be THE MOST fundamental element in religion. A sacrif selves, and then to respond to what appears to be THE MOST significant part of each complex sequence st common. Causation, conversely, is likely to be THE MOST interesting, but it is also the most diff t that it was seen as the basis of what was to be THE MOST dramatic change in the history of mankind tent - increasingly elsewhere, were proving to be THE MOST economically active. Public - sector hous terrible savage in the beginning, turns out to be THE MOST noble character in the novel. Many writer tion in terms of psychoanalysis. This seems to be THE MOST coherent position to hold. It does allow u valley, Nepal, population density appears to be THE MOST fundamental control on the chemistry of s ual engagement with these areas, seem to me to be THE MOST fruitful. Freedom and Manipulability In t wentieth - century distinctions may not always be THE MOST sensitive instruments for understanding t part of Shetland which has been expected to bear THE MOST immediate social and economic changes rel

The most is hedged in academic writing by the use of perhaps, even, among, probably, amongst, arguably, would have been + appears in quotes

Frequencies of *definitely* in different corpora

Frequencies per million	Academic abstracts		BNC Commerce	Academic General	BAWE	Learner student
definitely	0	4	7	10	22	140

How can you explain these differences in frequencies?



Frequencies of *definitely* in different genres





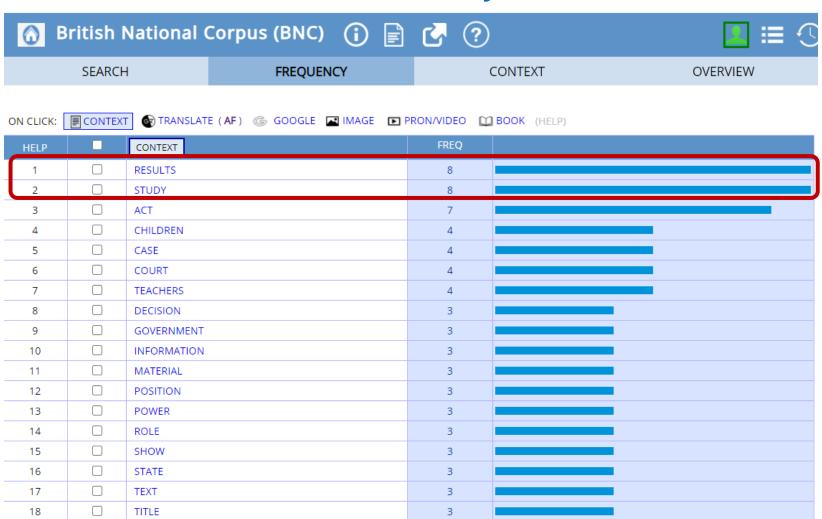
Boosters: *clearly*

```
t to ask questions, material must be presented CLEARLY to facilitate understanding.
onditions. However, rarely are such rationales CLEARLY worked out or attempts made
nose at risk of myco - bacterial reactivation. CLEARLY, HIV infected individuals co
ions which they might be expected to remember. CLEARLY they did take incontinence s
expression was remarkably induced. This result CLEARLY indicates that the enhanced
ing worse than the natural virus might result. CLEARLY, until these questions are a
more disorders than is the case. These results CLEARLY challenge any notion that wo
disease in unselected asthmatics. The results CLEARLY show a high prevalence of oe
specific radioimmunoassay systems. The results CLEARLY indicate that fasting gastri
worsens (~~ 2 =36.6; p<0.0001). These results CLEARLY show that hiatal hernia is a
med to be macrophages (CD68 + ). These results CLEARLY indicate that in untreated,
and malignant tumours. Exceptions to the rule CLEARLY exist, however, and Fearon h
atic cancers in the period between screenings. CLEARLY, only a percentage of lesion
sectioned by coloured tape, and each section CLEARLY headed (e.g. Introductory Co
stochemical analysis of frozen tissue sections CLEARLY shows the monoclonal nature
```

Why is *clearly* consistently used with *results*?



Boosters: *clearly*



Big theme: Argumentation



Big theme: common patterns in introducing the research gap - tenses + word choice + evaluative language + prepositions

of the printed word : " a field where very	little research	had been done . " (Locke 1974 : " 121)
ms were addressed . In the first place , since	little research	had taken place outside London , we chose to investigate a
pean children in British <mark>schools</mark> , remarkably	little research	has actually been done to study the linguistic behaviour of
curriculum is to be based on research ? Very	little research	has been carried out or teaching methods and the effectiveness
mography of the developed world relatively	little research	has been carried out in relation to the corresponding (i.e.
ame group of pupils on different occasions .	little research	has been carried out or consistency of performance on similar
carried out at dispersed homes . However ,	little research	has been carried out on the quantities of wood required . An
vely small quantity of imported wares . Very	little research	has been carried out on domestic wares and it remains far from
that has attracted the most attention . Very	little research	has been carried out on the functions of domestic pottery ; there
r less easily . As she points out:Remarkably	little research	has been carried out on the question of how these filters operate
round for close on two <mark>hundred years</mark> , very	little research	has been carried out either to attempt to understand how it works
nore likely . # Improving co-operation # Very	little research	has been conducted to discover how the barriers to effective
elping children adjust to parental epilepsy #	Little research	has been conducted pn the needs of children who have a parent
ession in the labour market . Yet surprisingly	little research	has been conducted into the subject , at least in Britain .
n draw on this experience in their teaching ,	little research	has been done in the area . Using techniques developed during an
they would be best advised to consult . Very	little research	has been done on their
nen relapse after coming <mark>under stress</mark> . Very	little research	has been done into its cause among adolescents and adults . As
espite the Gulfs long history of oil pollution ,	little research	has been done on the long-term impact on marine life there .
. For example , Spears (1982) states that	little research	has been done on the cost of rehabilitation of severely eroded

Big theme: Patterns of introducing research problem

ect causal interaction between them. however, the PROBLEM is that, if this is true, there is nothing

friend, Claudianus Mamertus. Again, however, the PROBLEM is a matter of style as much as one of sub ve science" - in the current jargon. however, the PROBLEM at the cognisance level is to describe how al languages, however, were a special and limited PROBLEM. Except to some extent in the Habsburg ter ot universal and unchangeable. There is however a PROBLEM, if one is trying to do both these things ies by housing associations. however, the central PROBLEM remains - cities may have lost segments of initial hypotheses. There is, however, always the PROBLEM of knowing whether the fault in the mediat h such programmes there is, however, a persistent PROBLEM of renewal of connection with the classroo pes of birth control. however, the overpopulation PROBLEM remains today, with the United Nations pre To the anthropologist however the myth provides a PROBLEM of meaning because in the context from whi arded as problematic. It is, however, largely our PROBLEM, rather than that of many societies which ormance. however, it is in reading that the major PROBLEM of achievement lies. Reading and writing a from standard English"). There is, however, some PROBLEM in this last comparison, since family inco however, while this tells us how not to solve the PROBLEM, it is less clear what positive steps we a sm and the victory of individualism. however, the PROBLEM can by no means be so readily dismissed. A fruitful. There remains, however, an interesting PROBLEM of definition. In the case of processes, t s can have a merely cosmetic effect, however. The PROBLEM is that egalitarian feminist method, like ng them as a social problem. however, the way the PROBLEM has been constructed and made the subject many children. however, it does highlight a major PROBLEM that we have so far not considered - some iminal tendencies. Sometimes, however, there is a PROBLEM in distinguishing "learning" theories beca criminology. however, this hardly helps with the PROBLEM of retributive justice, of which Beccaria increasing importance. At this time, however, the PROBLEM was almost certainly one of finding outlet



Big theme: Supporting claims, giving reasons

stival as the basilinna, Apollodorus suggests the PROBLEM was not that she was an alien, but because the group was everybody's motivation to work. The PROBLEM seemed to occur because it could never be f the philosophy of science. I have discussed the PROBLEM of demarcation in some detail because I be e observed. because of this, Hume claims that the PROBLEM of justification is actually the problem o not seem to be any audio feedback. This can be a PROBLEM because if you set the times there is noth ble 11, we could conclude that Amersham does have PROBLEM in credit control procedures because the a ant for the study of women, because it solves the PROBLEM of centuries of ignoring women in research ocus-on-form, which can arise either because of a PROBLEM of communication or because of the occurre the latter circumstance is seen by many DCs as a PROBLEM because they in their opinion they tend to oning initiatives just because they are new. This PROBLEM is known as 'strategic drift' as shown by or Nietzsche, was an important topic because "the PROBLEM of the value of pity conjures up doubt, mi Guardian not behaving in this way would not be a PROBLEM for Plato, because he would be able to pro m of unrepresentativeness appears to be less of a PROBLEM in "Women Drug Users " because the lives o always clear what the root cause of a particular PROBLEM was. This was because of the fluctuating s ed meant that reform was slow-going. Maybe a main PROBLEM was never to be discovered, maybe because d enforced by all Member States." This is a large PROBLEM for the environment globally both because ges to ignore inductive inference. For Popper the PROBLEM of induction need not be solved because th should be a question as "do we need to solve the PROBLEM of induction?" because to ask whether indu

'PROFCORP, 95% of the tokens for problem (39 of 41) fall into a causal category whereas in STUCORP only 32% of the tokens (150 of 473) occur in a causal relation'. Flowerdew (2003: 499).



Big theme: Supporting claims, giving reasons

questions in their entirety. THIS is partly because, for reasons to be discusse ed by feminist psychologists. THIS is partly because, like most psychologists, t ore than 5 - 10 per cent, and THIS is apparently because of the much greater eff open about what he is doing. THIS is both because of the moral principle involv occurring rhizobial bacteria. THIS is chiefly because the soya bean and its natu ety in periods of transition. THIS is crucial because Marx was to go on to argue from an expression. Sometimes THIS is desirable because the expression seems awk the client though, at times, THIS is difficult because he has to approach someo eality is at best secondary). THIS is disturbing because it challenges the appea you are not a brain in a vat. THIS is false because if you were a brain in a vat ple enjoy what they want, and THIS is important because it encourages consumption ost the national economy, and THIS is important because it helps people enjoy wh nner areas was about 5 acres. THIS is important because specific mechanisms have ect and designed environment. THIS is important because the differences relate t op) would have been long. But THIS is incorrect because the" Although... process ction is, Who sent the virus? THIS is instructive because it shows how an explan triarchal relations. However, THIS is insufficient because by itself it does not gists and to achieve success. THIS is largely because of diversions in their car n has been much more limited. THIS is mainly because this account of the mechani ften called a "talking cure". THIS is misleading because counselling is more than n's particular point of view, THIS is mistaken because it ignores the role playe

Big theme: Patterns of counter-argumentation

- Counter-argument: One could argue that
- Rebuttal of the counter-argument
- Restatement of own argument

```
ONE COULD ARGUE that at a time when the prospects
ONE COULD ARGUE that humanities scholars understan
ONE COULD ARGUE that I and Thou automatically impl
ONE COULD ARGUE that it is necessary for socioling
ONE COULD ARGUE that such a strong assumption of u
ONE COULD ARGUE that such complaints constitute so
ONE COULD ARGUE that the philosopher's thought for
ONE COULD ARGUE that the
ONE COULD ARGUE that work ind of rationale for a
ONE COULD ARGUE that work ind of rationale for a
one could argue that there
one could argue that such complaints constitute some are too many incomers in the
```

one could argue that such a strong assumption of underlying mechanism is only the normal situation in the sciences: one that allows experimental, as opposed to philosophical, work to proceed.

True enough but, as I shall argue below, AI is not an experimental science but an

engineering technique or, if

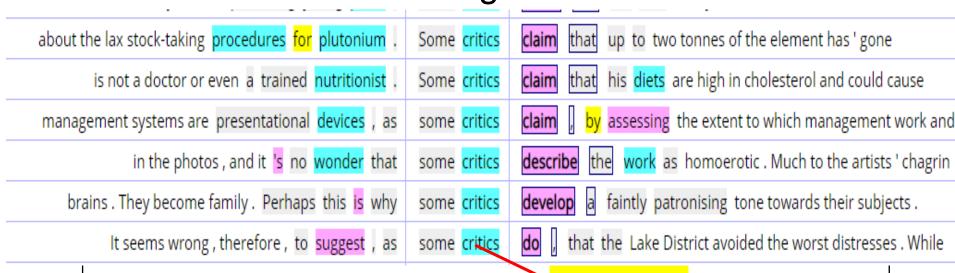
area.

There is no reason to refute
this argument: undoubtedly
this is a sentiment seemingly
shared by many Shetlanders.



Big theme: Patterns of counterargumentation

Counter-argument: Some critics claim that Rebuttal of the counter-argument Restatement of own argument



It seems wrong, therefore, to suggest, as **some critics** do, that the Lake District avoided the worst distresses. While enclosure came later to the Lake District than the southern parts of England, there is plenty of evidence of rural depopulation in the later years of the eighteenth century. This was because the system of land tenure was remarkably different from other areas ...



Big theme: Recurrent patterns in presenting results

In which examples authors:

- 1. highlight the significance of the results
- 2. compare their results with previous research
- 3. emphasise the novelty and contribution of their research
- 4. suggest treating results with caution

```
THE RESULTS agree with Bellugi's interpretations, thoug
THE RESULTS are as satisfactory as those obtained with
THE RESULTS are initial and tentative. Although concept
THE RESULTS are reliable, and representative of a much
THE RESULTS are robust to controlling for alternative g
THE RESULTS are still far from decisive. But it guickly
THE RESULTS are too varied to allow generalization. The
THE RESULTS are very impressive: there is obviously a r
THE RESULTS are very similar, using similar psychometri
THE RESULTS are widely divergent. Picasso's paintings a
THE RESULTS challenge accepted Australian approaches to
THE RESULTS challenge current concepts of Gelechioidea,
THE RESULTS confirm the hypothesis of this study that t
THE RESULTS confirm the validity of using stable isotop
THE RESULTS converge to the known values based on a New
THE RESULTS do not confirm their conclusion. Observers,
THE RESULTS encouraging, even with no differences in te
THE RESULTS should provide a better understanding of th
THE RESULTS show significant relationships between rele
THE RESULTS supported previous work conducted by Pate,
E RESULTS tend to agree with Mandler and Johnson (1977)
HE RESULTS were gratifying. The style of the élite was
THE RESULTS were highly productive and the basic laws o
```

My transition

From

using corpora as a tool to teach vocabulary

to

using corpus searches to raise students awareness of a 'bigger picture':

- Elements of academic style
- Stance, e.g., strength of claim, using hedging and boosters
- Joining the use of tenses and the purpose of writing
- Argumentation: supporting claims, introducing the research gap or problem, giving reasons, using counterargumentation, framing results

My transition

From

using corpus materials in a course of Academic vocabulary and occasionally in tutorials to

- teaching a course Academic writing with corpora (Beginners and Advanced)
- and using corpus materials in

Literature reviews

Writing introduction

Writing abstracts

Writing about methods

Describing visual data

Tenses and passive voice in academic writing

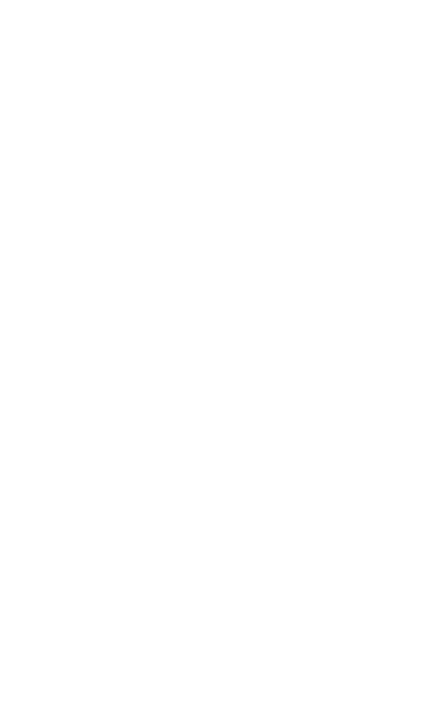


Academic Writing with Corpora

A resource book for data-driven learning

Tatyana Karpenko-Seccombe





Thank you Have you got any questions?

Tatyana Karpenko-Seccombe **Huddersfield University**





References, corpus tools and corpora

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International journal of corpus linguistics*, 14(3), 275-311.

- Boulton, A. (2007). DDL is in the details... and in the big themes. In Corpus Linguistics Conference: CL2007 (p. XX).
- Carter, R. & McCarthy, M. 1995. "Grammar and the spoken language". Applied Linguistics, 16 (2), 141–158.
- Chan, T. and Hsien-Chin Liou, H-Ch. Effects of Web-based Concordancing Instruction on EFL Students' Learning of Verb Noun Collocations
- Cobb, T. & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (eds), *Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, p. 478-497. DOI: 10.1017/CBO9781139764377.027
- Cobb, T. Lextutor concordancer. Available at lextutor.ca/conc/eng/
- Conrad, S. M., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*.
- Davies, Mark. (2004-) British National Corpus (from Oxford University Press). Available online at https://www.english-corpora.org/bnc/
- Flowerdew, L. (2003). A combined corpus and systemic-functional analysis of the Problem-Solution pattern in a student and professional corpus of technical writing. *TESOL Quarterly*, 37 (3), 489–511.
- Lee, D. &. Swales J. M. (2006) A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self compiled corpora. *English for Specific Purposes*,25 (1), 56–75.
- Nesi, Gardner, Thompson & Wickens, BAWE (used in Lextutor) British Academic Written English corpus, developed at the Universities of Warwick, Reading and Oxford Brookes <u>coventry.ac.uk/bawe</u>
- MICUSP Michigan Corpus of Upper-level Student Papers (2009). Ann Arbor, MI: The Regents of the University of Michigan. Available at micusp.elicorpora.info/
- Römer, U. (2005). *Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics* (Vol. 18). John Benjamins Publishing. Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, 7(1), 140-162.
- SkELL (Sketch Engine for Language Learning) Baisa, Vít a Vít Suchomel. SkELL Web Interface for English Language Learning. In Eighth Workshop on Recent Advances in Slavonic Natural Language Processing. Brno: Tribun EU, 2014, pp. 63-70. ISSN 2336-4289. Available at https://skell.sketchengine.co.uk/run.cgi/skell
- Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. Journal of Second Language Writing, 32, 17-35.