

4th International Conference Sustainable Multilingualism 2017

THE INFORMAL SCIENTIFIC TALK: CORPUS-BASED FINDINGS AND APPLICATIONS FOR DATA-DRIVEN LEARNING

[Natalia Morgoun, Elena Kozharskaya](#)

Lomonosov Moscow State University, Russia

26-27 May, 2017

Vytautas Magnus University, Kaunas, Lithuania

outline

- Aims and motivation
- What one can do with this corpus? -
introduction to using corpora in teaching
practice
- Work in progress
- Conclusions

The notion of linguistic corpus

A collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research [Sinclair, 1991]

Course structure

- follow me style: Modal examples
- Formulate your own query to solve a given task
- Reveal your own problems independently and use them using a corpora to solve

Subjects and context

- for undergraduate students
(English<>Russian)
- 18 hours of contact time
- introduce junior students to basic corpus skills
and ready-made online corpora
- methodology: hand-on approach
- discussion stage

**report** noun *switch to [report](#) (verb)***Context** **verbs with report as subject**indicate follow state suggest conclude note recommend show say find highlight claim cite reveal describe**verbs with report as object**video-tape submit publish release issue file prepare write read commission receive deny compile produce present**adjectives with report**inaccurate accurate detailed critical false pursuant incorrect correct online available positive true consistent due
relevant**modifiers of report**annual news medium credit final recent police newspaper official progress press detailed initial interim intelligence**nouns modified by report**card writing yesterday finding state note recommendation today Monday Tuesday summary detail Apple Thursday
Wednesday

[View in Wordle](#)

natural conditions

replacing *natural* in 'natural condition'

[natural condition](#) (63)[normal condition](#) (91)[basic condition](#) (36)[environmental condition](#) (88)[atmospheric condition](#) (17)[prevailing condition](#) (54)[primitive condition](#) (18)[standard condition](#) (45)[natural condition](#) (63)[essential condition](#) (25)[general condition](#) (131)

replacing *condition* in 'natural condition'

[natural condition](#) (63)[natural state](#) (53)[natural order](#) (88)

natural conditions

[replacing *natural* in 'natural condition'](#), e.g.
normal condition

[replacing *condition* in 'natural condition'](#), e.g.
natural state

phrase (nn) nn is the frequency of usage

■ Good Word Combinations

■ Bad Word Combinations

— Similarity of meaning

.word - Means singular noun only

Corpus Structure

- 98 000 tokens
- 90 TED talks - oceanography, **entomology, ornithology, developmental biology, bioengineering, molecular biology**
- 2003 – 2015
- <http://www.antlab.sci.waseda.ac.jp/software.htm>

Typical functions

- Frequency lists
- collocational profiles of synonyms
- generic analysis
- study of concordance lines

Reading Concordances [Sinclair, 2004]

- Categorize or separate the word according to its part of speech (noun, adjective, verb, adverb, prepositions, verbals, etc...).
- Categorize or separate each part of speech according to its meaning.
How many meanings for the word did you find?
- Give a definition for each meaning you found in the data.
Go back to question two in order to make a list of definitions.
- Look at the language data again according to its part of speech. This time locate any recurring patterns in the data that you think are idiomatic phrases or collocations.
- Create one original sentence for each meaning of the word.

Decide where 'influence' is a verb (mark it as V) and where it is a noun (N)

The screenshot displays a concordance search interface. At the top, a navigation bar includes tabs for 'Concordance' (selected), 'Concordance Plot', 'File View', 'Clusters/N-Grams', 'Collocates', 'Word List', and 'Keyword List'. Below this, the 'Concordance Hits' section shows 310 results. The main display area is a table with columns 'Hit', 'KWIC', and 'File'. The 'Hit' column lists line numbers from 1 to 22. The 'KWIC' column shows the context of the word 'influence' in various sentences, with the word itself highlighted in blue. The 'File' column lists the source files, including LOB_G.txt, BROWN_J.txt, BROWN_H.txt, BROWN_J.txt, LOB_A.txt, LOB_B.txt, BROWN_C.txt, LOB_G.txt, BROWN_J.txt, LOB_E.txt, LOB_J.txt, BROWN_J.txt, LOB_B.txt, BROWN_E.txt, LOB_H.txt, BROWN_G.txt, LOB_F.txt, BROWN_G.txt, BROWN_C.txt, BROWN_G.txt, BROWN_J.txt, and BROWN_J.txt. At the bottom, there is a search bar with the term 'influence' and buttons for 'Start', 'Stop', and 'Sort'. Below the search bar, there are checkboxes for 'Words', 'Case', and 'Regex', and a 'Search Window Size' dropdown set to 50. There are also checkboxes for 'Level 1', 'Level 2', and 'Level 3', each with a '3R' dropdown. A 'Clone Results' button is located at the bottom right.

Hit	KWIC	File
1	next generation . Others will fall under its influence only for a time and will then	LOB_G.txt
2	interaction effect should , if potent , extend its influence over all academic achievement . The Sta	BROWN_J.txt
3	regard as corruption , favoritism , and personal influence are so accepted as consistent with the	BROWN_H.txt
4	developed can be represented as : **f The influence of mechanical action on the particles of	BROWN_J.txt
5	markets yesterday were still falling under the influence of recent adverse company news . Wiggin	LOB_A.txt
6	tisan policy would undoubtedly strengthen British influence for peace - an influence more necessary	LOB_B.txt
7	has nourished mine , as no other social influence ever has , and in converse with you	BROWN_C.txt
8	little courts and ports were under the influence of Hindu and Buddhist India , which in	LOB_G.txt
9	of emotional responses are similar ; as the influence of learning and environment are felt , e	BROWN_J.txt
10	the old commercial quarter , with its oriental influence , its minarets and clutter of wares , an	LOB_E.txt
11	composed without any trace of inspiration . The influence of Poe and Mallarme2 , and the part	LOB_J.txt
12	tioned which alter the hypothalamic balance would influence emotional state and behavior but that em	BROWN_J.txt
13	. It may also fairly reflect the good influence of naturally anxious wives . ENTER THE	LOB_B.txt
14	catch the judges ' eyes and which rightfully influence their decisions are symmetry and that ha	BROWN_E.txt
15	s requirements of terrace houses , which markedly influence their internal arrangement , we propose	LOB_H.txt
16	is at times equally concerned about the influence of the art . In the ideal state ,	BROWN_G.txt
17	.O.E.U. Journal . Does T.V. Influence Elections ? by Arthur Woodburn , M.P .	LOB_F.txt
18	-century novelist who argued for the powerful influence of literature as a teacher of society	BROWN_G.txt
19	in Formosa , that Mao was the evil influence behind Khrushchev at the Summit Conferen	BROWN_C.txt
20	maintains , may be corrupted by this subtle influence . Plato 's attitude toward poetry has a	BROWN_G.txt
21	the element of power , primarily the voluntary influence of non-authoritative power , has been ex	BROWN_J.txt
22	out . It is obvious that such an influence can only be expected in the final	BROWN_J.txt

REFERENCES

- Anthony, L. (2008). *AntConc (Version 3.2.1.)* [computer software]. Retrieved November 1, 2008, from <http://www.antlab.sci.waseda.ac.jp/software.html>.
- 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. (2002). Exploring new directions for discovery learning. In B. Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Boulton, A. (2008). DDL: Reaching the parts other teaching can't reach? In Frankenburg- Brown, D.F. (1974). Advanced vocabulary teaching: The problem of collocation. *RELC Journal*, 5, 1-11.
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9, 111-125.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing?. *System*, 25, 301-315.
- Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 393-409). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
- Davies, M. (2008-). The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present. Available online at <http://www.americanacorporus.org>.
- Fan, M. (2009). An exploratory study of collocational use by ESL students-A task based approach. *System*, 37, 110-123.
- Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, 21, 231-244. Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, 41 (2), 339-359.
- Gavioli, L. (1997). Exploring texts through the concordancer: Guiding the learner. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 83-99). London: Longman
- Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive on-line database. *Language Learning & Language*, 9, 90-110.
- Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 107-117).
- Kettemann, B. (1995). On the use of concordancing in ELT. *Arbeiten aus Anglistik und Amerikanistik*, 20, 29-41.
- Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1-23). London: Longman.

THANK YOU!

morgounn@gmail.com