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

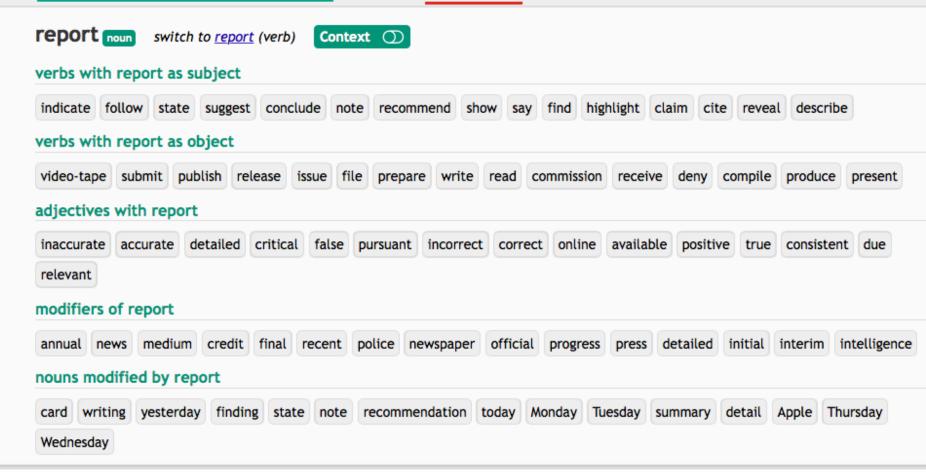
Subjects and context

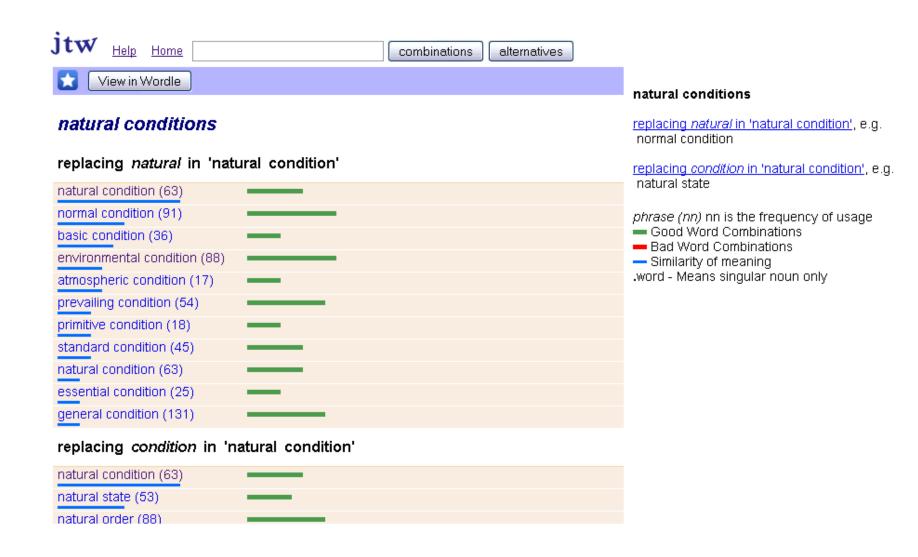
- for undergraduate students (English<>Russian)
- 18 hours of contact time
- introduce junior students to basic corpus skills and ready-made online corpora
- methodology: hand-on approach
- discussion stage





Examples Word sketch Similar words More features





Corpus Structure

98 000 tokens

 90 TED talks - oceanography, entomology, ornithology, developmental biology, bioengineering, molecular biology

- 2003 2015
- <u>http://www.antlab.sci.waseda.ac.jp/software.ht</u>
 <u>m</u>

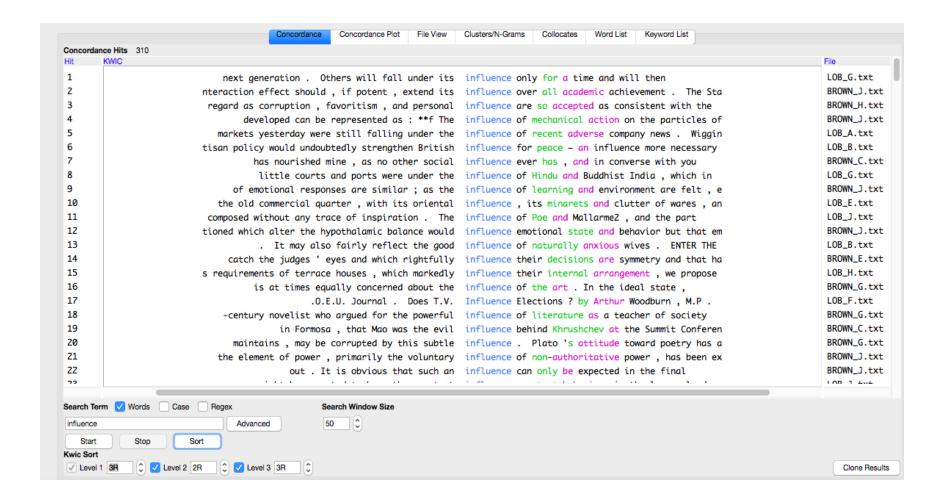
Typical functions

- 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)



REFERENCES

- Anthony, L. (2008). *AntConc (Version3.2.1.)*[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.americancorpus.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