

Corpus 作成・使用について
「理系英語のライティング」 p. 21-24, 154-157

Using corpus linguistics 3Cs

Corpus = a database of texts of the target genre

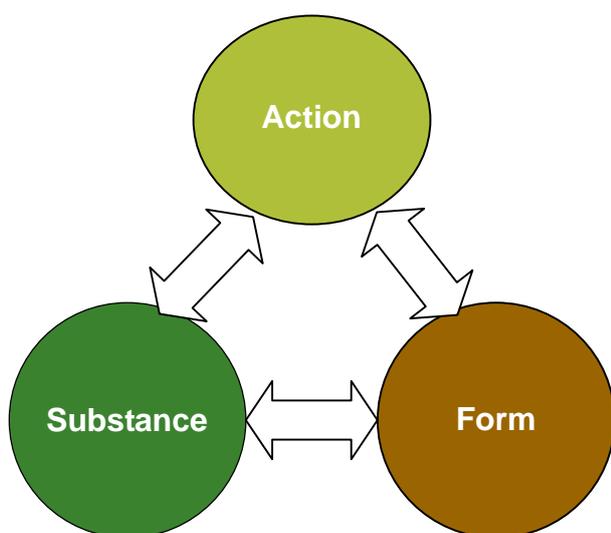
Concordance = software to help view text features [AntConc \(Laurence Anthony\)](#)

<http://www.antlab.sci.waseda.ac.jp/software.html>

(Japanese explanation: [Shin Ishikawa](#))

<http://www11.ocn.ne.jp/~iskwshin/antconc.html>

Collocations = words that frequently appear together in this genre being examined



A **corpus** is a database of scientific papers on the subject of your interest. By analyzing your personal corpus, you can learn very efficiently the structure, the characteristics, the most used expressions or, for short, the **form** of a paper in your field. With this corpus-based knowledge, you should become able to write papers much faster than if you just tried to grasp the **form** of papers by reading many of them.

1. First, save this file into your USB memory

2. Building a corpus

2.1 Create a folder named "FamilyName," where "Family" is your family name and "Name" your given name, in your MyDocuments directory or USB memory.

2.2 Access for example, Scencedirect: <http://www.sciencedirect.com/> or University Library online journals.

2.3 Choose "Journals" and the journal of interest. Select top-ranking journals in your field.

2.4 Download as many papers as possible (only articles or full papers, do not download review articles or short communications), **as html, pdf and text files**, into the corpus directory created in step 2.1. Do NOT do automatic downloading!! There is a warning about this on the library page. You need to hand pick the articles you wish to include in your corpus and manually download each one. Try to select recent articles.

2.5 Name all files with the same name (the file type extensions will be different so this is possible) as follows: Journal name_Vol_ Issue_Pages_Key word

例 :

Paper: Prediction of ozone levels in London using the MM5-CMAQ
modelling system

Journal: Environmental Modelling & Software

Volume 21, Issue 4 , April 2006, Pages 566-576

の場合は、ファイル名は次のようにする。

Corpus file name: EnvModelSoft21_4_566_576_prediction ozone levels.htm

EnvModelSoft21_4_566_576_prediction ozone levels.pdf

EnvModelSoft21_4_566_576_prediction ozone levels.txt

3. Using AntConc

3.1 From <http://www.antlab.sci.waseda.ac.jp/software.html>, download AntConc3.2.1w.exe and its Readme file into your folder (USB memory).

3.2 Execute AntConc, choose "Open Dir" and select the folder which contains your corpus.

3.3 In the Search window below, input for example "such" and push "Start."

3.4 In the "Level 1" for sorting, set "1R" and push the "Sort" button. You will see, then, various uses of "such," and will be able to see which usage is more frequent.

3.5 Select the tag "Word Clusters," set "Minimum size of Clusters" to "2" and "Maximum size of clusters" to "4" and push "Start." This will confirm that "such as" is very frequent

3.6 You can verify which verbs come often before "from" for example, by doing a search for "from" and repeat the steps 3.3 to 3.5 (in step 3.4, change the "1R" to "1L").

3.7 You can also compare results of different searches by pushing the "Save Conc Window" button.