Corpus 作成・使用について

Using corpus linguistics 3Cs

 Corpus = a database of texts of the target genre

 Concordance = software to help view text features [*AntConc*](file:///C%3A%5CUsers%5CElwood%5CDropbox%5CGRIPS%20Common%5CCourses%20%26%20Curricula%20Resources%5COsaInformatics08%5COsaInfoPres090508%5Cantconc3.2.1w.exe)([Laurence Anthony](file:///C%3A%5CUsers%5CElwood%5CDropbox%5CGRIPS%20Common%5CCourses%20%26%20Curricula%20Resources%5COsaInformatics08%5COsaInfoPres090508%5CLaurence%20Anthony%27s%20software.htm))

 [*http://www.antlab.sci.waseda.ac.jp/software.html*](http://www.antlab.sci.waseda.ac.jp/software.html)

 *(Japanese explanation:* [*Shin Ishikawa*](file:///C%3A%5CUsers%5CElwood%5CDropbox%5CGRIPS%20Common%5CCourses%20%26%20Curricula%20Resources%5COsaInformatics08%5COsaInfoPres090508%5C%E7%A5%9E%E6%88%B8%E5%A4%A7%E5%AD%A6%E7%9F%B3%E5%B7%9D%E6%85%8E%E4%B8%80%E9%83%8E%E7%A0%94%E7%A9%B6%E5%AE%A4.htm)

 <http://language.sakura.ne.jp/s/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.

**Action**

**Substance**

**Form**

1. First, save this file into your USB memory
2. Building a corpus
	1. Create a folder named “CorpusFamilyName,” where “Family” is your family name and “Name” your given name, in your MyDocuments directory or USB memory.
	2. Access for example, Sciencedirect: <http://www.sciencedirect.com/> or online journals in your university library.
	3. Select top-ranking journals in your field.
	4. Download as many papers as possible (only articles or full papers, do not download review articles or short communications), **as html** (if available)**, pdf and text files**, into the corpus directory created in step 2.1. Do NOT do automatic downloading!! Hand pick the articles you wish to include in your corpus and manually download each one. Try to select recent articles.
	5. Name all files with the same name (the file type extensions will be different so this is possible) as follows: Journal name\_Year\_Vol\_ Issue\_Pages\_Key word

 Here is an example:

 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

 Corpus file name:

 EnvModelSoft\_2006\_21\_4\_566\_576\_prediction ozone levels.htm

 EnvModelSoft\_2006\_21\_4\_566\_576\_prediction ozone levels.pdf

 EnvModelSoft\_2006\_21\_4\_566\_576\_prediction ozone levels.txt

1. Using AntConc
	1. From <http://www.antlab.sci.waseda.ac.jp/software.html>, download AntConc and its Readme file into your folder (USB memory).
	2. Execute AntConc, choose “Open Dir” and select the folder which contains your corpus.
	3. In the Search window below, input for example “such” and push “Start.”
	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.
	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
	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”).
	7. You can also compare results of different searches by pushing the “Save Conc Window” button.
	8. You can also save the search results by placing the cursor in the window and [ALT + A] to select all and then [CTRL + C] to copy. Open an Excel spreadsheet and with the cursor in the A1 cell, push [CTRL + V] to paste the search results. Next, choose all the cells and change the font to Courier New to align the results.