

Hands-on session #3 – Cesax: Working with databases – annotation

Lesson content:

1) CorpusStudio: database preparation	(09:00 – 09:15)
2) CorpusStudio: database creation	(09:15 – 09:25)
3) Cesax: opening a database	(09:25 – 09:35)
4) Cesax: checking database results	(09:35 – 09:55)
5) Cesax: create your own databases!!	(09:55 – 10:20)
6) Cesax: annotating database features	(10:20 – 10:40)
7) Cesax: database output for statistics	(10:40 – 11:00)

1 CorpusStudio: database preparation

The information for the database is within the CorpusStudio results, and needs to be extracted from there, and saved in a separate database *xml* file.

- 1) Opening the corpus project of yesterday
 - a) Open CorpusStudio, and load yesterday’s LongdaleTest project
 - i) Use **File/Recent**
 - b) Load the *data* calculated in this project using Tools/LoadResults (or Ctrl+F12)
- 2) Preparing database extraction
 - a) Each query pipeline element can have its own database, so go to the “constructor editor”
 - b) Click on “Features for each Dbase result”
 - i) The standard three features will appear: TextOfpreNP, TextOfrc, TextOfcompl
 - c) Add all the new features we added yesterday (use “Add”)
 - i) PreDet - *we added that yesterday*
 - ii) RCsize - *we added that yesterday*
 - iii) RCdepth
 - iv) NPfunction
 - d) NOTE:
The names of the features used here, in the Constructor Editor, will be names of the features in the database. They do not need to match the names you have used in your main query or in the definitions file for your project
 - e) Save the project (Ctrl+S)

2 CorpusStudio: database creation

- 1) Creating a database
 - a) Remain in the “constructor editor”
 - b) Click on “**Create result database**”
 - i) Wait...
 - c) Note the *location* of the results database
 - i) Copy and paste it to a document, for instance
 - ii) Or just copy and paste it to have it ready
- 2) Moving it into safety
 - a) Copy the database to a *different* location
 - i) Otherwise it might get overwritten the next time you press “Create database”...
 - b) Change the name of the database, if needed, but keep the extension to “.xml”

3 Cesax: opening a database

- 1) Accessing Cesax
 - a) Start up Cesax again
- 2) Opening a database
 - a) Choose “Corpus/Load”
 - b) Navigate to the location where you have saved the *copy* of the database
 - c) Make sure to load it (press OK)
 - i) If the database is large, loading can take some time...
 - ii) Wait until the status becomes “Corpus has been loaded”

4 Cesax: checking database results

Verification of the results you have obtained can be very illustrating and lead you to adapt your query.

- 1) Leaf through results
 - a) Use the listbox on the left side of the screen to select individual records (hits) in your database
 - b) Once selected, the hit will show:
 - i) The text line belonging to this hit
 - ii) The amount of preceding and following context (defined in your Corpus Research Project)
 - iii) All the adaptable features that have been calculated
 - iv) The “Category” for this hit (the one that has been used for sub-categorization)
 - v) The syntax of your hit (in bracketed labelling)
- 2) Look at the syntax
 - a) The bracketed labelling may not be too informative, unless you are familiar with it...
- 3) Double click to selected line in the listbox on the left side
 - a) The program jumps to the original line of text this hit comes from
 - b) You can check the “Tree” view of this line here, for instance
- 4) Go back to the database
 - a) Select the “CorpusResults” tab
- 5) Check the features
 - a) See if the feature values are what you had expected them to be
- 6) Change the “selected feature” (a combobox)
 - a) Try “RCsize” (if that is how you have called it – otherwise try a different one)
 - b) Sort on RCsize
 - c) Find examples of the biggest RCsize – are they what we wanted??

5 Cesax: Create your own database

Turn to your own corpus research project:

- 1) Do database preparation in CorpusStudio
- 2) Create the database from within CorpusStudio
- 3) Open the database in Cesax
- 4) Check the results

6 Cesax: annotating database features

You have your database, but you may want to edit it.

You may not like some of the features, for instance

- 1) Feature changing
 - a) Features can be changed manually per field
- 2) Find and replace features
 - a) The “ComplSize” feature gives “-1” for zero complements. We will change it to “0”
 - b) Use Corpus/FindAndReplace
- 3) Status tracking (using F2)
- 4) Add “Notes” per record

7 Cesax: database output for statistics

Preparation for statistics depends on the package you are going to use.

- 1) String values
 - a) If the package can handle ‘string’ values, then do Corpus/Report
 - b) Open the resulting *.txt* file
- 2) SPSS preparation
 - a) Do Corpus/PrepareForSPSS
 - b) Position the textual features in “do not include in statistics”
 - c) Three files are created:
 - i) ...Dbase_corpus.htm The *html* file you see
 - ii) ...Dbase.sps SPSS “syntax” file (links names and values)
 - iii) ...Dbase_spss.txt Data file to be opened by SPSS