

ESLLI 2006

Introduction to Corpus Resources, Annotation and Access

Exercise: The SALTO Annotation Tool

Sabine Schulte im Walde and Heike Zinsmeister

August 3, 2006

Login: `esslli`

1 Task Description

You are provided a corpus with five sentences that are pre-annotated with syntactic trees. The task is to use the *SALTO Annotation Tool* to add a layer of annotation: For each verb, the constituents that represent thematic roles for that verb, are selected and labeled. You can find the inventory of thematic roles that is used (and a brief description) on slides 21 and 22 of the *Semantic Annotation* session: AGENT, PATIENT, EXPERIENCER, THEME, RECIPIENT, BENEFACTIVE, GOAL, LOCATION, SOURCE, INSTRUMENT.

2 Preparation

1. Define the environment variable `$SALTO`. To be provided by the lecturers.
2. Access the SALTO directory: `cd $SALTO`
3. Start SALTO: `$SALTO/Salsa.bat` (Windows) or `$SALTO/Salsa.sh` (Unix/Linux)
4. Login with your login name. DO NOT register as administrator.
5. Left mouse double click on the directory symbol for `in` in `User`.
6. Right mouse click on the file symbol for the corpus `esslli` and select `move to working directory`;
alternatively: select the file symbol for `esslli`, left mouse click on the right-hand symbol below the `User` box, and select `move to working directory`.
7. Left mouse double click on the directory symbol for `work` in `User`.
8. Left mouse double click on the file symbol for `esslli`.
9. Ignore the error message (that file could not load): click `OK`

You have opened the corpus file `esslli` with its syntactic annotation and are ready for adding the second layer of annotation.

3 Annotation

TASK: Annotate the thematic roles that are evoked by each verb within the five sentences in `esslli`.

Note that (a) not all constituents necessarily represent a thematic role in our inventory, and (b) some constituents might be thematic roles of other than verbal elements. Annotate only those constituents that are thematic roles of verbs.

1. For each verb:
 - (a) Choose the verb and make a right mouse click.
 - (b) Choose `Invoke frame` → `ROLES`.
 - (c) For each constituent in the sentence that represents a thematic role of the verb:
 - i. Right mouse click on the verb's `ROLE` box.
 - ii. Choose `Add element` and the thematic role of the element.
 - iii. Drag the new box with the thematic role to the constituent in the tree that represents the thematic role.

If you want to correct a verb or role choice, perform a right mouse click on the respective box and choose `Delete`.

2. Use the arrows at the bottom of the sentence to jump to other sentences.
3. Save your changes: `File` → `Close corpus` → `Save`.
4. When you have finished your annotation, right mouse click on the file symbol for the corpus `esslli` and select `move file to out directory`.

You can annotate as many sentences as you like (between one and five) and either compare your annotation with the lecturers' solutions or continue with the adjudication.

4 *Extra Task: Adjudication*

TASK: The adjudication assumes that two annotators have finished annotating the same corpus parts. You should let the lecturers know that you have finished the annotation and would like to adjudicate. The lecturer prepares a merged version of the two annotations and provides this merged file for adjudication. The task is to go through the disagreements within the two annotations and decide *together*, i.e., with the two annotators, which parts of the annotations should be kept and which parts should be deleted.

1. Login and open the corpus file with two merged annotations in the `in` directory, as described in section 2–*Preparation*.
2. Adjudicate the corpus sentences:
 - (a) Use `CTRL-D` (`Apple-D` for Macs) to jump to the first/next disagreement in the merged annotations.
 - (b) Decide whether
 - i. you accept the annotation as correct: do nothing **OR**
 - ii. you decide to delete the annotation: right mouse click and select `Delete`.
 - (c) Finish the adjudication when you receive the message `No further differences found`.
3. Save your changes: `File` → `Close corpus` → `Save`.

You have reached an annotation of the corpus on which two annotators agree.