NooJ is a freeware language-engineering development environment for formalizing and integrating nine levels of linguistic phenomena: orthography and typography, lexical, inflectional and derivational morphology, local, structural and transformational syntax, semantics. NooJ contains tools to help construct, test, debug, maintain and accumulate large sets of linguistic resources, as well as tools to process large texts and corpora. The system has been developed since 2002 and it has been used to build over 20 language modules, including language modules developed in the framework of the META-NET CESAR ICT-PSP project. This tutorial intends to help participants to master four basic NooJ functionalities: corpus processing, formalization of linguistic units, syntactic parsing.

The tutorial session consists of four labs and one presentation. Participants must come with their laptop with NooJ v3 (either .NET or MONO) pre-installed.

1. Corpus Processing (Kristina Vuckovic, Univ. of Zagreb)

Import a text in any file format. Need to define NooJ text units, which are independent areas inside which linguistic resources are applied. A corpus is a collection of text files that will share the same linguistic resources. Apply simple and complex queries to texts in order to build concordances. Perform statistical analyses on any query.

2. Linguistic Units and annotations (Max Silberztein, Univ. of Besançon)

NooJ’s basic objects are Atomic Linguistic Units (ALU). NooJ contains tools to formalize nine levels of linguistic phenomena. NooJ manages annotations (rather
than tags) to make its different parsers communicate and keep unresolved ambiguities in a Text Annotation Structure.

3. Local Syntax (Tamas Varadi, Budapest Academy of Sciences)

What is a local grammar? Developing and combining local grammars. Applying local grammars in cascade. How to maintain grammars with the help of NooJ debugger and contracts.

Automatic annotation of linguistic units and named entities. XML import and export.

4. Syntactic Analysis (Slim Mesfar, ENSI Tunis)

Syntactic parsing; structural vs derivational syntactic trees.

How to produce paraphrases automatically.
5. Conclusion : NooJ in the CESAR Project  (Tamas Varadi, Budapest Academy of Sciences)

The CESAR project (www.cesar-project.net) is an EU funded project launched in early 2011 which together with three other projects form META-NET (www.meta-net.eu). One of the central missions of META-NET is to develop an open linguistic infrastructure within the META-SHARE facility. NooJ is one of the flagship tools that the CESAR project contributes to META-SHARE. The NooJ community (www.nooj4nlp.net) is deeply embedded in META-NET. The current overhaul of NooJ is taking place as a collaborative effort within the CESAR project. The talk will briefly outline the CESAR project and what effect it may have on the NooJ community.

References


