Exercise #5 – D-grammar

- Goal: To prepare the d-grammar for the training corpus
- Deliverable:
    - *dgrammar_<ID>.txt*
- Activities:

1. Check the annotation of your corpus:
    a. UNLWEB>UNLDEV>UTK>CORPUS>CORPUS>LOAD <corpus_org>
    b. UNLWEB>UNLDEV>UTK>CORPUS>DICTIONARY>LOAD <DEFAULT DICTIONARY>
    c. UNLWEB>UNLDEV>UTK>CORPUS>DICTIONARY>LOAD <dic.txt>
    d. UNLWEB>UNLDEV>UTK>CORPUS>PROCESS>ANNOTATE (ATTRIBUTE = ALL, FORMAT=VALUE ONLY, DETERMINISTIC=TRUE)
2. In case of incorrect categorization, provide the d-rules that are necessary to solve the lexical ambiguities (see the English example below).
3. Upload the dgrammar to the UTK (UNLWEB>UNLDEV>UTK>CORPUS>D-GRAMMAR>ADD)
4. Load the dgrammar, recheck the annotation of your corpus and do the necessary changes until the annotation is correct.

ENGLISH EXAMPLE

... whether it be independent …
(whether,C,SCJ)(BLK)(it,R,PPR,NOM,3PS,@3)(BLK)(be,I,AUX,INF)(BLK)(independent,J,ADJ)
"be" should have been annotated as a copula (be,V,COP) instead of an auxiliary (be,I,AUX)
In order to correct this, there are two options:
EITHER:
To block "be" as an auxiliary in that context (preferable), e.g.:
    (AUX)(BLK)(ADJ)=0;
 OR:
 To induce "be"as a copula in that context (more expensive in terms of processing), e.g.:
    (COP)(BLK)(ADJ)=1;

48, Route de Chancy. CH-1213, Petit-Lancy. Geneva, Switzerland. Tel. +41 22 879 8090 Fax +41 22 879 8091

info@undlfoundation.org www.undlfoundation.org