

XV UNL School, 21-25 July 2014, Geneva

Exercise #2 – N-Grammar

- Goal: To prepare the N-grammar for the training corpus
- Deliverables:
 - *ngrammar_<ID>.txt*
 - *corpus_seg_<ID>.txt*
- Activities:
 1. Read the instructions available at www.unlweb.net/wiki/N-rule
 2. Analyze the English N-grammar available at UNLWEB>UNLARIUM>GRAMMAR>ENGLISH>GRAMMARS>NORMALIZATION>N.1
 3. Create a N-grammar (*ngrammar_<ID>.txt*) necessary to
 - a) Normalize the abbreviations and contractions appearing in your training corpus
 - b) Segment your training corpus in sentences
 4. Upload the N-grammar to the UTK (UNLWEB>UNLDEV>UTK>CORPUS>N-GRAMMAR>ADD)
 5. Segment your training corpus:
 - a) UNLWEB>UNLDEV>UTK>CORPUS>CORPUS>LOAD [YOUR CORPUS]
 - b) UNLWEB>UNLDEV>UTK>CORPUS>N-GRAMMAR>LOAD [YOUR N-GRAMMAR]
 - c) UNLWEB>UNLDEV>UTK>CORPUS>PROCESS>SEGMENT
 6. Check the results and do the necessary changes until the normalization and segmentation are correct
 7. Export the resulted segmented text as *corpus_seg_<ID>.txt*