

LICENSE INFORMATION ON THE FILES IN THE `aelius_data` FOLDER OF THE AELIUS DISTRIBUTION

The data in this folder is distributed solely for non-commercial, non-profit educational and research use under the terms of the Open Content License [<http://opencontent.org/opl.shtml>].

The files `AeliusBRUBT.pkl`, `AeliusRUBT.pkl` and `AeliusHunPos` contain statistical models of Portuguese for POS-tagging using NLTK or Aelius. These models were trained on a pruned, partly retagged and partly retokenized version of the *Tycho Brahe Parsed Corpus of Historical Portuguese* (TBCHP) [<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html>].

The file `freq_tycho_a.pkl` is a NLTK word frequency distribution of a sample of this corpus containing about 200000 alphabetical tokens.

Thus, all this material is a derivative work of the *Tycho Brahe Parsed Corpus of Historical Portuguese* (TBCHP), whose copyrights are held by the respective original authors. Any use of this material must be compliant with the TBCHP's license. For more information, see the English version of the corpus license page [<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/termos.html>] or the Portuguese version [<http://www.tycho.iel.unicamp.br/~tycho/corpus/termos.html>].

The file `example.txt` contains an excerpt with the first two paragraphs from the novel *Luzia-Homem*, published 1903 by the Brazilian writer Domingos Olímpio (born in Sobral, Ceará, in 1850, and died in Rio de Janeiro in 1906), extracted from Wikisource [<http://pt.wikisource.org/wiki/Luzia-Homem/I>]. The file `exemplo.nltk.gold.txt` is a manually revised version of the POS-tagged output from the AeliusBRUBT tagger. Both untagged and tagged versions of this excerpt were revised by Leonel F. de Alencar and constitute part of the CORPTXLIT Project [<http://www.leonel.profusehost.net/corptext.html>].

Publishing results based on any of these data requires that the author be acknowledged. To do so, please refer to the file's name (e.g. AeliusHunPos) and the package of which it is part, citing thusly:

Alencar, Leonel F. de. Aelius Brazilian Portuguese POS-Tagger and Corpus Annotation Tool. 2011.
<http://aelius.sourceforge.net/>

Since any work using the mentioned models is indirectly based on the *Tycho Brahe Parsed Corpus of Historical Portuguese*, you should also, according to its above-mentioned license, acknowledge the TBCHP, its creators and/or distributors and cite this corpus accordingly.