

CLTK



P. S. Langeslag



pyenv + venv + Jupyter

```
$ pyenv install -v 3.9.10
$ pyenv shell 3.9.10
$ pyenv versions
  system
* 3.9.10 (set by PYENV_VERSION environment variable)
$ cd some/containing/folder
$ python -m venv nlp
$ source nlp/bin/activate
$ pip install cltk
$ python -m ipykernel install --user --name nlp
```

Integrated Pipeline

```
>>> from cltk import NLP
>>> nlp = NLP(language='ang')
>>> doc = 'him anum is wuldor and weorðmynd on worulda woruld'
>>> results = nlp.analyze(text=doc)
>>> results.tokens
['him', 'anum', 'is', 'wuldor', 'and', 'weorðmynd', 'on', 'worulda', 'woruld']
>>> results.lemmata
['he', 'an', 'is', 'bearm', 'and', 'weorðmynd', 'on', 'weorold', 'weorold']
```

Old English Lemmatizer Options

```
>>> from cltk.lemmatize.ang import OldEnglishDictionaryLemmatizer
>>> lem = OldEnglishDictionaryLemmatizer()
>>> lem.lemmatize_token('man', return_frequencies=True, best_guess=False)
[('mann', -6.829400539827225), ('man', -4.832846657953158)]
>>> lem.lemmatize_token('sindon', return_frequencies=True, best_guess=False)
[('wesan', -7.435536343397541), ('pesan', 0)]
```

LatinBackoffLemmatizer(), No Options

```
>>> from cltk.lemmatize.lat import LatinBackoffLemmatizer
>>> lem = LatinBackoffLemmatizer()
>>> doc = ['servum', 'cantare', 'iussit']
>>> lem.lemmatize(doc)
[('servum', 'servus1'), ('cantare', 'canto'), ('iussit', 'iubeo')]
```

More Hidden Old English Functionality

```
>>> from cltk.phonology.ang.transcription import Transcriber
>>> Transcriber.transcribe('habbað æfre ānrædne ġelēafan')
'hab:að æ:fre a:nræ:dne jelæ:avan'
>>> from cltk.phonology.ang.phonology import OldEnglishSyllabifier
>>> syll = OldEnglishSyllabifier()
>>> syll('monan')
['mo', 'nan']
>>> from cltk.phonology.ang.transliteration import Transliterate
>>> Transliterate.transliterate('RƿMƿƿƿNƿ ƿtM RMAƿƿƿNƿ ƿƿXƿt XIBRƿƿR')
'romwalus and reumwalus twoegen gibrothær'
```

Behind the Scenes

```
~/cltk_data/ang/model/ang_models_cltk/data/  
/usr/lib/python3.9/site-packages/cltk/stops/
```

Bibliography

Johnson, Kyle P. “CLTK Notebooks,” n.d.

<https://github.com/cltk/cltk/tree/master/notebooks>.

———. “CLTK Tutorials,” n.d. <https://github.com/cltk/tutorials>.

———. “The Classical Language Toolkit (CLTK): Legacy Docs,” 2019.

<https://legacy.cltk.org/en/latest/>.

Johnson, Kyle P., Patrick J. Burns, John Stewart, Todd Cook, Clément Besnier, and William J. B. Mattingly. “The Classical Language Toolkit: An NLP Framework for Pre-Modern Languages.” In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, 20–29. Association for Computational Linguistics, 2021.
<https://doi.org/10.18653/v1/2021.acl-demo.3>.