Lexical Distance Project Sourcing and Methodology

**Sources:**

This section lists all source materials used during the research. Sources are for both documents analyzed, and the Python packages used.

**Documents analyzed:**

This section lists the documents used for analysis during the research.

* Title: UN Resolution 55/100
  + Source: https://opus.nlpl.eu/MultiUN.php
  + Resolution 55/100, [on the report of the Third Committee (A/55/602/Add.2 and Corr.1)]
  + Begins with:
    - “Reaffirming that the promotion and protection of all human rights and fundamental freedoms must be considered”
  + Ending with:
    - “… or acting in any other manner inconsistent with the purposes of the United Nations”
  + Citing:
    - [Alexandre Rafalovitch, Robert Dale. 2009. United Nations General Assembly Resolutions: A Six-Language Parallel Corpus. In Proceedings of the MT Summit XII, pages 292-299, Ottawa, Canada, August](http://uncorpora.org/Rafalovitch_Dale_MT_Summit_2009.pdf)
    - J. Tiedemann, 2012, [*Parallel Data, Tools and Interfaces in OPUS.*](http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf) In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)

**Python Packages:**

This section lists all of the Python packages used, along with official documentation if available.

*All packages used:*

* unidecode: https://pypi.org/project/Unidecode/
* re: https://docs.python.org/3/library/re.html
* ntlk: http://www.nltk.org/
  + stem: https://www.nltk.org/api/nltk.stem.html
    - snowball: https://www.nltk.org/\_modules/nltk/stem/snowball.html
      * SnowballStemmer
  + corpus: https://www.nltk.org/api/nltk.corpus.html
    - stopwords
  + translate: https://www.nltk.org/api/nltk.translate.html
    - Alignment
* numpy: https://numpy.org/
* pandas: https://pandas.pydata.org/
* seaborn: https://seaborn.pydata.org/
* matplotlib: https://matplotlib.org/
  + pyplot

*Packages by Python File:*

* Parser.py:
  + unidecode
  + re
  + ntlk
    - stem
      * snowball
        + SnowballStemmer
    - corpus
      * stopwords
* Ldistance.py:
  + re
  + matplotlib
  + numpy
  + ntlk
    - stem
      * snowball
        + SnowballStemmer
    - corpus
      * stopwords
* Parser.py:
  + matplotlib
    - pyplot
  + seaborn
  + numpy
  + pandas
* BagOfConsonants.py:
  + N/A

**Methodology:**

This section is designed to describe the methods by which the research was done. The section does not intend to provide high-level explanation of the methods from a linguistic standpoint, but does intend to explain certain choices that were made during research, and to explain how the code used in research was formulated.

***General information:***

*On the source material:*

a

*On the generation of documents:*

a

*On the use of Google Translate:*

A

***Parsing Information:***

*On the removal of common words:*

A

*On the removal of punctuation and numbers:*

A

*On the removal of vowels*

A

*On the use of the stemmer:*

A

***Calculation Information:***

*On the calculation of the Levenshtein Distance:*

A

*On the Alignment Process:*

A

*On the measuring of samples:*

A

**Use of Each Python Package:**

This section goes into further detail on how each Python package was used in the research.