

Non-copyright encumbered corpora

SketchEngine

Sandra is contacting the owner of the SketchEngine Swahili data to see if we can get a license that allows us to release our annotated data.

Global voices corpus

- available in opus
- non-copyrighted

Unannotated version of helsinki corpus is under CC by 4

Revision #2

Created Wed, Feb 5, 2020 6:34 PM by [kenneth](#)

Updated Fri, Jan 8, 2021 1:20 AM by [kenneth](#)