

- We created, possibly the first, labelled idioms corpus with 10 classes, based on the BNC and the UKWaC corpora.
- The dataset may be extended to meet specific NLP needs by researchers.
- We make the dataset and the relevant codes for working with it publicly available.