

Appendix A

BOCHUM ENGLISH COUNTABILITY LEXICON (BECL) description:

The Bochumer English countability lexicon is a database for the lexical countability preference of English noun senses. Noun senses covered in the lexicon are taken from the Princeton WordNet for English.

The development of BECL has been directed by Prof. Tibor Kiss and Prof. Francis Jeffrey Pelletier and has been supported by the Anneliese Maier Research Award of the Alexander von Humboldt Foundation under grant 3.2 L KAN/1148876 and 3.2 L DEU/1148875.

The following researchers and research assistants (listed in alphabetic order) have contributed to the construction of BECL: Ron Hoffmann, Halima Husic, Tibor Kiss, Florian Kriete, Francis Jeffrey Pelletier, Johanna Marie Poppek, Roman Nino Simunic.

The annotation of the English noun senses was conducted by Mathieu Dovan, Meghan Jeffrey, Lisa Shorten, Tobias Stadtfeld and Fiona Wilson.

The BECL data consists of the following modules:

- BECL_2_1_master.csv
a Microsoft Excel spreadsheet containing the unanimously annotated English noun senses
- BECL_2_1_classes.xlsx and BECL_2_1_classes_no_pn.xlsx
two Microsoft Excel spreadsheets containing the unanimously annotated English noun senses sorted by their countability classes; one including proper names and one without.
- BECL_2_1_multiples.csv and BECL_2_1_multiples_no_pn.csv
two Microsoft Excel spreadsheets containing only multiples, i.e. noun senses with different countability classes for each sense; one including proper names and one without.
- BECL_2_1_multiples_xtabs.csv and BECL_2_1_multiples_no_pn_xtabs.csv
two Microsoft Excel spreadsheets containing crosstabs of multiples; one including proper names and one without.
- BECL_2_1_survey_classes.pdf and BECL_2_1_survey_classes_no_pn.pdf
two PDF documents containing summaries of the overall distribution of the annotated noun senses into countability classes; one including proper names and one without.

The BECL data file provides the following pieces of information for each noun:

- The noun form, its sense number in WordNet, and its sense description in WordNet
- The total number of senses in WordNet and the synset for the sense
- The set of similar synsets measured according to the LCH score
- An indicator whether all senses of a noun listed in WordNet have been annotated
- The frequency distribution of the lemma, the singular and the plural form in OpenANC
- The outcome of the annotations
- Whether the noun in question is likely to have a proper name as a homograph. Users can select whether they want to include proper names.
- Information about the annotation process.
- An indicator whether the noun occurs in more than one orthographic form (and is thus a duplicate)

apart from its orthography).

BECL is still under development, and its coverage is constantly being extended. The current size of version 2.1 is:

- 7,050 annotated lexical units in total
- 11,875 annotated noun senses in total