

Lexical database for Norwegian

1. BASIC INFORMATION

- 1.1 *Lexicon type (wordform, explanatory, terminological lexicon, wordnet, etc)*
Wordform
- 1.2 *Representation of the lexicon (flat files, database, markup)*
Database
- 1.3 *Character encoding*
ISO-8859-1
Speech Assessment Methods Phonetic Alphabet SAMPA

2. ADMINISTRATIVE INFORMATION

- 2.1 *Contact person (name, e-mail)*
Arne Martinus Lindstad, ArneMartinus.Lindstad@nb.no
- 2.2 *Copyright statement and information on IPR*
Open Source

3. TECHNICAL INFORMATION

- 3.1 *Data structure of an entry*
A lexicon entry is associated with information about 51 different features. Each entry covers one line in a text file that comprises the whole database.
- 3.2 *Lexicon size (num. of lexical items)*
784 240 items

4. CONTENT INFORMATION

- 4.1 *The natural language(s) of the lexicon*
Norwegian (nbo)
- 4.2 *Entry Type*
Lexicon item
- 4.3 *Attributes*
Pronunciation, POS, morphology, lemma, style, source and other information. 51 features in all.
- 4.4 *Coverage of the lexicon*
All words in *Bokmålsordboka* (the reference dictionary for bokmål)
All members of the 100k list (list of the 100 000 most frequent word forms)
All words in the *SpeechDat* database (telephone recordings)
Subsets of place names, person names, etc.
- 4.5 *Intended application of the lexicon*
Speech technology
- 4.5 *Reliability (automatically/manually constructed)*
Automatically constructed

1. RELEVANT REFERENCES AND OTHER INFORMATION

- <http://www.nb.no/spraakbanken/tilgjengelege-ressursar/leksikalske-databasar>