

# ***Lexical database for Danish - Available, includes documentation and toolset***

## **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](mailto: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)*  
237 873

## **4. CONTENT INFORMATION**

- 4.1 *The natural language(s) of the lexicon*  
Danish (dan)
- 4.2 *Entry Type*  
Lexicon items
- 4.3 *Attributes*  
POS, morphology, lemma, style, source
- 4.4 *Coverage of the lexicon*  
Frequency based 100K list, INSO- og SpeechDat materials
- 4.5 *Intended application of the lexicon*  
Speech technology
- 4.6 *Reliability (automatically/manually constructed)*  
There is no automatic inflector for Danish, all lexicon items have been transcribed manually.

## **5. RELEVANT REFERENCES AND OTHER INFORMATION**

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