

Scarrie Lexicon

1. BASIC INFORMATION

1.1 *Lexicon type (wordform, explanatory, terminological lexicon, wordnet, etc),*

Wordform

1.2 *Representation of the lexicon (flat files, database, markup)*

LMF

1.3 *Character encoding*

Utf-8

2. ADMINISTRATIVE INFORMATION

2.1 *Contact person (name, e-mail)*

Koenraad de Smedt

2.2 *Copyright statement and information on IPR*

META-SHARE Commons BY SA

The copyright holder is Koenraad De Smedt. Reproduction and use is allowed under share-alike and authorship credit conditions.

3. TECHNICAL INFORMATION

3.1 *Data structure of an entry*

Each LexicalEntry has an unspecified Lemma and a number of WordForm elements.

Normal WordForm elements have elements Feat with att: writtenForm, corrStyle, featureList, replacement and synCat.

Special entries, in particular prefixes and word forms occurring in abbreviations neither have feat with att: replacement nor with att: synCat.

The use of the attributes is explained in the SCARRIE deliverable 3.3.1 Tagset.

3.2 *Lexicon size (num. of lexical items)*

scarrie-abbrevwords-lmf.xml:23

scarrie-gramwords-lmf.xml:707

scarrie-gramwords2-lmf.xml:48

scarrie-idiomwords-lmf.xml:811

scarrie-main-lmf.xml:359684

scarrie-prefixes-lmf.xml:327

scarrie-suffixes-lmf.xml:562

4. CONTENT INFORMATION

4.1 *The natural language(s) of the lexicon*

Norwegian (nbo)

4.2 *Entry Type*

Entries are collections of word forms belonging to the same lemma.

4.3 *Attributes*

Elements of type feat have attributes att and val.

4.4 *Coverage of the lexicon*

The lexicon covers the vocabulary in Bokmålsordboka. It was constructed in 1999 and does not cover

vocabulary extensions and spelling changes after Jan. 1, 1999.

4.5 *Intended application of the lexicon*

In its original format, the lexicon was intended for use in the proofreading application developed in the SCARRIE project (LE3-4239), as described in the deliverables from this project. In the current LMF format it is intended to be reused in other applications.

4.6 *Reliability (automatically/manually constructed)*

The lexicon was mostly automatically constructed. It was partly manually inspected and was used in a series of tests of the proofreading application.

5. RELEVANT REFERENCES AND OTHER INFORMATION

<http://ling.uib.no/~desmedt/scarrie/>