Sofie Treebank

1. BASIC INFORMATION

1.1 Corpus composition

The novel "Sofies verden" and translations

1.2 Representation of the corpora (flat files, database, markup)

Flat files and markup

2 ADMINISTRATIVE INFORMATION

2.1 Contact person (name, e-mail)

Victoria Rosén, victoria@uib.no

2.2 Copyright statement and information on IPR

Open Source

3 TECHNICAL INFORMATION

3.1 Data structure of an entry

An LFG grammar assigns two representations to each reading of a sentence, a c-structure (a phrase structure tree) and an f-structure (an attribute-value graph).

Prolog file format

3.2 Corpora size (num. of tokens)

200 sentences

4 CONTENT INFORMATION

4.1 Type of the corpus (monolingual/multilingual, parallel/comparable, raw/annotated)

Monolingual, annotated

4.2 The natural language(s) of the corpus

Norwegian (nbo)

4.3 Domain(s)/register(s) of the corpus

Fiction

4.4 Annotations in the corpus (if an annotated corpus)

4.4.1 Types of annotations (paragraph mark-up, sentence mark-up, lexical mark-up, syntactic mark-up, semantic mark-up, discourse mark-up)

Syntactic mark-up (LFG)

4.5 Intended application of the corpus

Linguistic research

4.6 Reliability of the annotations (automatically/manually assigned) – if any

Automatically annotated, manually checked

4 RELEVANT REFERENCES AND OTHER INFORMATION

http://iness.uib.no/iness/main-page?session-id=231331326413533