

# ***Oslo-Bergen tagger***

## **1. BASIC INFORMATION**

### *1.1 Resource composition*

The Oslo-Bergen tagger is a robust morphological and syntactic tagger developed at the University of Oslo and at Uni Computing in Bergen over several years. The tagger consists of three main modules: a preprocessor with multitagger and compound analyser, a grammar module for morphological and syntactic disambiguation (Constraint Grammar) and a statistical module that removes the last of the remaining morphological ambiguity (only for Bokmål)

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

The downloadable resource contains the following elements:

- a) Readme file with instructions on how to use the tagger. The readme file also shows where to find and how to install the various parts of the tagger:
- b) Multitagger with lexicon for Norwegian Bokmål and Nynorsk. (At the moment, binary versions for Linux Intel 32-bit and 64-bit and og Mac OS X 64-bit are available)
- c) CG3-compiler (VISL-CG3) from the University of Southern Denmark in Odense
- d) OBT-stat (Statistical module that runs after the CG-rules)
- e) HunPos
- f) CG-rules for Bokmål and Nynorsk (morphological disambiguation). The Bokmål rules are available in two versions: One version for CG-grammar stand-alone and one version to be used together with OBT-stat.
- g) Shell Script for running the tagger

### *1.3 Character encoding*

Not relevant

## **2. ADMINISTRATIVE INFORMATION**

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

Janne Bondi Johannessen, j.b.johannessen@iln.uio.no

### *2.2 Copyright statement and information on IPR*

Downloadable on GPL terms

## **3. TECHNICAL INFORMATION**

### *3.1 Data structure of an entry*

Not relevant

### *3.2 Resource size (num. of rules)*

Not relevant

## **4. CONTENT INFORMATION**

### *4.1 Type of the resource (language (in(dependant)*

Language dependent

### *4.2 The natural language(s) for the resource is applicable (if language dependent)*

Norwegian bokmål and nynorsk (nbo, nno)

### *4.3 Domain(s)/register(s) of the resource*

Not relevant

### *4.4 Annotations in the resource (if an annotated resource)*

Not relevant

### *4.5 Intended application of the resource*

Syntactic and morphological analysis

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

Not relevant

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

<http://tekstlab.uio.no/obt-ny/english/read.html>