

Overview of corpus searches

Copyright © Elisabet Engdahl & Anu Laanemets 2016

Licensed under a Creative Commons CC-BY License

This note describes the corpus searches and the annotation codes for the datasets in **Danska.xls**, **Norska.xls** and **Svenska.xls**. The datasets are referred to in

Engdahl, E. & A. Laanemets: Opersonlig passiv i danska, norska och svenska – en korpusstudie. *Norsk Lingvistisk Tidsskrift*.

Engdahl, E.: Expletive passives in Scandinavian – with and without objects.

Danska.xls

KorpusDK: <http://ordnet.dk/korpusdk>

Example search string:

```
[word="der" ] [lemma="blive" ] []{0,1} [word="ingen" | word="meget" |  
word="mange" | word="nogen" | word="ingenting" | word="intet" |  
word="noget" | word="nogle" | pos="ART" | pos="PERS" | pos="NUM" ]  
[]{0,1} [morph="PCP2_PAS" ] [pos!="N" ]
```

Annotation codes:

Column A: o = impersonal passive
 q = irrelevant
Column B: 0 = negated
 1 = not negated

Norska.xls

Norwegian *Bokmål*: *Leksikografisk bokmålskorpus*:

<http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk>

Example search string:

```
[ ((word="det" %c)) ] [ ((lemma="bli" %c)) ] []{0,1}  
[ ((word="intet" %c)) ] []{0,1} [ ((type="perf-part-  
v")) ] [ !((ordkl="s")) ]
```

The results are shown with one tab for each of the following: {ingen, mange, mye, noe, noen, meget, DET, PRO}

Norwegian *Nynorsk*: *Oslo Corpus of Tagged Norwegian Texts*

http://www.tekstlab.uio.no/norsk/korpus/nynorsk/oktntn_reg.html

Search strings:

```
"det" [(tagg=".bli\".*" | tagg=".* .bli\".*")] [] {0,1}
([word="ingenmyelmangelnoen?"][(tagg=".* (pronldet).*")] [] {0,1} [tagg=".*perf-part.*"]
[tagg != ".* subst.*"]
```

```
[(tagg=".bli\".*" | tagg=".* .bli\".*")] "det" [] {0,1}
([word="ingenmyelmangelnoen?"][(tagg=".* (pronldet).*")] [] {0,1} [tagg=".*perf-part.*"]
[tagg != ".* subst.*"]
```

The results are shown under the tab *nynorska*.

Svenska.xls

Korp: <http://spraakbanken.gu.se/korp/>

Example search string for *bli-passive*:

```
[word = "det" %c] [lex contains "bli..vb.1"] []{0,1} [(lex contains
"ingen..pn.1" | lex contains "mycken..pn.1" | lex contains
"mången..pn.1" | lex contains "någon..pn.1")] []{0,1} [pos = "PC"]
[pos != "NN"]
```

The results are shown with one tab for *Skönlitteratur* (novels), *GP* (the newspaper *Göteborgs-Posten*) and *Bloggmix*; only a selection of the hits are included and only the first 100, or 200, are annotated.

Annotation codes for *bli-passive*:

Column A: o = impersonal passive
q = irrelevant
x = double

Column B: 0 = negated
1 = not negated

Column C: 1 = verb is GÖRA ('do')
0 = other verbs

Example search string for *s-passive*:

```
[word = "det" %c][lemma not contains ".*s" & msd = "VB.*SFO"]
```

The results are shown with one tab for *Skönlitteratur* (novels), *GP* (the newspaper *Göteborgs-Posten*) and *Bloggmix*; only a selection of the hits are included and only the first 100, or 200, are annotated.

Annotation codes for *s-passive*:

Column A: ob = impersonal passive with oblique
oi = intransitive
ot = transitive
ox = extraposed clause
r = referential *det* or definite article
x = double