

BETAVERSJON AV SPRÅKBANKENS GULLKORPUS FOR BOKMÅL OG NYNORSK

Korpusene utvikles ved Nasjonalbiblioteket/Språkbanken i samarbeid med Tekstlaboratoriet ved Universitetet i Oslo. Denne dumpen av korpusene ble tatt 6. august 2012.

Innhold

20120806_nob_gullkorpus.conll	Gullkorpus for bokmål
20120806_nno_gullkorpus.conll	Gullkorpus for nynorsk

Om filformatet

Korpusene er i conll-format. For mer om dette formatet, se ufal.mff.cuni.cz/conll2009-st/task-description.html.

Kort om innholdet

Korpusene består av tekst som er manuelt annotert for morfologiske trekk, syntaktiske funksjoner og hierarkisk struktur. Den morfologiske annotasjonen følger i all hovedsak Oslo-Bergen-taggeren (tekstlab.uio.no/obt-ny/). Dependensgrammatikk er brukt som formalisme for den syntaktiske annotasjonen. Mange av analysevalgene følger *Norsk referansegrammatikk* (Jan Terje Faarlund, Svein Lie og Kjell Ivar Vannebo, Universitetsforlaget, 1997). En detaljert annotasjonsveiledning vil etter hvert bli gjort tilgjengelig.

Tekstutvalg

Korpusene består inntil videre av avistekst fra sent 90-tall fram til i dag, og stortingsreferater fra samme periode.

Bokmål

Aftenposten
Dagbladet
Klassekampen
Stortingsreferater

Nynorsk

Dag og Tid
Vest-Telemark blad
Klassekampen

Vi vil etter hvert legge ut en detaljert tekstoversikt på Språkbankens nettsider.

For spørsmål, kommentarer og feilrapporter, vennligst kontakt Språkbanken på sprakbanken@nb.no.