



NO-CLARIN: deltagelse fra UiB, Uni Digital og NHH

**Victoria Rosén, Knut Hofland, Gisle Andersen, Koenraad De
Smedt**

University of Bergen, Uni Digital, NHH

clarin@uib.no

**Nasjonalt møte
Oslo, Nasjonalbiblioteket
2010-6-18**



- Konsortiumdeltager i CLARIN
- Bedt av NFR om å koordinere samarbeidet i NO-CLARIN
- Ekspertiseområder bl.a.:
 - trebanker og datamaskinell grammatikk
 - andrespråkskorpus
 - terminologi
 - leksikalsk semantikk

Aktuelle ressursler/prosjekter



(sammen med Uni Digital m.m.)

- Holbergprosjektet (e-utgivelse)
- COLA (multimediakorpus, spansk)
- KIAP ('comparable' corpus, norsk, engelsk, fransk)
- Navnesamlinger
- Menota (Medieval Nordic Text Archive)
- ASK (andrespråkskorpus)

Aktuelle ressurser/prosjekter



INESS (Infrastructure for the Exploration of Syntax and Semantics)

- Ikke et forskningsprosjekt, men et infrastrukturprosjekt
- Skal lage en stor trebank for norsk
- Skal gjøre trebanker med avanserte søkemuligheter for norsk og andre språk lett tilgjengelige

Infrastrukturplaner



- Integrere INESS på europeisk nivå i CLARIN
- Tilby ASK til andre andrespråk gjennom europeisk samarbeid
- Konsolidere eksisterende samlinger gjennom kvalitetssikring og omkoding
- Koordinasjon av NO-CLARIN



Ekspertiseområder:

- avanserte korpusverktøy
- tekstkoding og digitale utgivelser
- IDP (interactive dynamic presentation)
- samarbeidsomgivelser

Aktuelle verktøy/prosjekter



- CORPUSCLE: neste generasjonsplattform for korpus, mulig avløser for Corpus Workbench, mer effektive algoritmer, kraftigere søkespråk, kan brukes på veldig store datamengder
- TCA2: parallelstilling med GUI, i Java
- Oslo-Bergen-tagger



Query [Concordance](#) [Word list](#) [Distribution](#) [Document](#) [Colligation](#)

Hit 1 - 30 of 416 . | Go to:

[next 30 hits](#)[Download](#)

Concordance type:

lexic

Show/edit annotations

F29

Ctrl+0 clear | Ctrl+1 F29:i | Ctrl+2 F29:L | Ctrl+3 F29:ij

Ctrl+4 F29:anna | Ctrl+5 F29:uty | Ctrl+6 F29:sein

document

D-00015 => I alle fall. </pers> <pers> Mm. </pers> <pers> Så vi må no ha noko her au. Men Molde i seg
 D-00015 kt, veit du. </pers> <pers> Så var det no om å gjere å finne på ting. </pers> <pers> Så kom vi
 D-00015 </pers> <pers> Det blei meir naturleg å </pers> <pers> bruke bokmål. </pers> <pers> Så det
 D-00015 : enden då, veit du, «hva, hva sier, sier du?» </pers> <pers> Hadde ikkje skjønt kva han sa, så
 D-00015 småungar. </pers> <pers> Men </pers> <pers> det var no litt dumt, synest eg, at den skulen
 D-00015 e så godt ho kunne. <para/> </pers> <pers> <para/> </pers> <pers> Ja, nei, da, vi hadde det
 D-00015 /pers> <pers> oppfører seg sjølve, var eg tenkt på å seie. Om dei har... </pers> <pers> Det er
 D-00015 dialekt, men så </pers> <pers> er ho så forfina på ein måte. </pers> <pers> Men det er no så
 D-00015 's> <pers> <para/> </pers> <pers> Nei, det var no nedlegging av, </pers> <pers> av, av eine
 D-00015 ile i Midsundet? </pers> <pers> Sånn at alle fekk plass. I staden for å begynne å flytte no eine
 D-00015 om er planen, er vel det at dei skal nedleggjast, </pers> <pers> skal nedleggjast begge desse
 D-00015 ossing og noko fræning og noko molding, veit du. Det er, det er sånn, <para/> </pers> <pers>
 D-00015 o molding, veit du. Det er, det er sånn, <para/> </pers> <pers> Blanda. </pers> <pers> sånn
 D-00015 > <pers> Nei, det vil ikkje eg seie, for </pers> <pers> Nei. </pers> <pers> vi blei no nokså fort
 D-00015 rs> <pers> Dei seier "i". </pers> <pers> Det er vel, </pers> <pers> dei var vel, det er vel meir
 D-00015 e skule alle frå Midøya. </pers> <pers> Vi hadde no eigen skule på Midøya. </pers> <pers> Ja,
 D-00015 <pers> Sånne ting au, trur eg berre. </pers> <pers> Ja, og så blei det det at søre midøyingane
 D-00015 i handla sikkert der og. </pers> <pers> Sånne ting au, trur eg berre. </pers> <pers> Ja, og så
 D-00015 ge ti. </pers> <pers> Og... </pers> <pers> Men skulen blei no nedlagd, og så </pers> <pers>
 D-00015 io nedlegging av, </pers> <pers> av, av eine barneskulen borte på Midøya då. Og alle elevane
 D-00015 <pers> Hadde eit lite problem då, veit du. Ho @name fekk ikkje bestandig vere med ho. For ho
 D-00015 Midøya og. </pers> <pers> Vi gjekk i, mykje i lag med sunnmøringane men </pers> <pers> det
 D-00015 Ja. </pers> <pers> Ja, eg synest eigentleg nynorsk er veldig fint då, men. </pers> <pers> Det
 D-00015 rståeleg nok ikkje alle som var heilt einige ti. </pers> <pers> Og... </pers> <pers> Men skulen
 D-00015 i Midsund? </pers> <pers> Nei, det vil ikkje eg seie, for </pers> <pers> Nei. </pers> <pers> vi
 D-00015 dei slutta med det. </pers> <pers> Tredje. Kanskje eg hadde begynt i tredje eller noko når det
 D-00015 e barneskulen, elevane der </pers> <pers> ut til Raknes, </pers> <pers> og slå i hop. For det
 D-00015 > skia var stort sett meir, <para/> </pers> <pers> Knekte. <para/> </pers> <pers> meir, meir
 D-00015 <pers> med </pers> <pers> Aukra og </pers> <pers> bli ein </pers> <pers> kommune, enn å
 D-00015 anskje betre å slå oss i saman </pers> <pers> med </pers> <pers> Aukra og </pers> <pers>

match

sjølv_{det} F29:L
 vel_{adv} F29:uty
 blei_{verb} F29:L
 blei_{verb} F29:L
 blei_{verb} F29:uty
 fole_{adj} F29:L
 klart_{adj|adv} F26:T.F29:L
 klart_{adj|adv} F26:T.F29:L
 barneskulen_{subst} F29:L
 barneskulen_{subst} F28:N.F29:L
 barneskulane_{subst} F28:sein.F29:L
 Blanda_{verb|verb|subst} F29:L
 blanda_{adj|subst} F29:L
 blanda_{verb|subst} F29:L
 blanda_{adj|adj|verb|subst} F29:L
 blei_{verb} F29:L
 blei_{subst} F29:L
 blei_{verb} F29:L
 blei_{verb|subst} F29:L
 blei_{verb|subst} F29:L
 blei_{verb} F29:L
 blei_{verb} F29:L
 blei_{verb} F29:L
 blei_{verb} F29:L
 blei_{verb} F29:sein
 blei_{verb} F29:sein
 blekk_{verb} F29:L
 bli_{verb} F29:L
 bli_{verb} F29:L

, veit du, sjølve byer
 på at vi skulle sy ein
 enklare, </pers> <p
 han irritert då. </pe
 nedlagd då. </pers>
 koseleg på Midøysk
 det at dei, dei får se
 deira si dialekt som
 borte på Midøya då.
 , elevane der </pers>
 . Alt skal </pers> <p
 . </pers> <pers> så
 at det er... </pers>
 med, </pers> <pers>
 [vore] </pers> <per
 det litt tull med diale
 det for langt å sykle
 det det at søre midø
 det vel starta opp s
 flytta ut til Raknes. </p
 for lita. <para/> </p
 hin vegen det bikka
 meir naturleg å </p
 no nedlagd, og så </p
 no nokså fort bland
 slutt på det. </pers>
 veldig trongt innpå
 enn tre. For dei var
 ein utkant under Mo
 ein </pers> <pers>

Aktuelle ressursler/prosjekter



- ICAME (engelsk)
- COLT (engelsk), COLA (spansk), multimedia
- WAB (Wittgensteinarkivet)
- NSPC (norsk-spansk)
- Norsk aviskorpus (norsk, >800 M ord)

Noen ressursler i eldre systemer trenger omkoding til nye systemer

Planer ifm. infrastruktur



- Tilby Corpuscle som nytt korpusplattform nasjonalt og internasjonalt, utvide til støtte for parallellkorpus, multimedia
- Tilby TCA2, videreutvikle med bl.a. egne moduler for parallellstilling på ordnivå og termekstraksjon
- Gjøre IDP tilgjengelig nasjonalt og internasjonalt på alt TEI-XML materiale (gjennom IDE-SONE)
- Aviskorpuset: konvertering, produksjon av n-gram, drift og brukerstøtte
- Tilby ICAME i Corpuscle som standard ressurs



Ekspertiseområder:

- Terminologi og fagspråk
- Oversettelse

Aktuelle ressurser/prosjekter



- NHH termbase
- KB-N parallellkorpus og termbase
- Termportalen, NOT-basen, RTT-materialet
- Translatørkorpus NHH (ikke ferdig)
- Norsk aviskorpus

Planer ifm. infrastruktur



Termsøk — Tempportalen.no • Mozilla Firefox

http://tempportalen.uib.no/soek

Tempportalen

Forside | Ressurser | Termsøk | Om prosjektet | Aktuelt | Kontakt oss

Du er her: [Forside](#) → [Termsøk](#)

Termsøk

I prototypen av Termsøk er det mulig å søke både i enkeltressurser og på tvers av flere ressurser. Søkene kan spesifiseres etter domene. Ressursene i dagens prototype består av fire store og flerspråklige terminologidatabaser. Termsøk er passordbeskyttet så lenge Tempportalen er i valideringsfasen, men vil deretter være fritt tilgjengelig for alle eksterne brukere.

Søk

Søkestreng: (oblig.)

Høyretrunkert:

Termbase:

Språk:

Felt:

(Totalt 513 treff)

| Base | Språk | Felt | Tekst |
|---------|-------|------|----------------------------------|
| KBN | nob | hvd | oljekorrigert overskudd |
| KBN | nob | hvd | oljekorrigert budsjettunderskudd |
| KBN | nob | hvd | oljereelatert industri |
| KBN | nob | hvd | oljeinntekter |
| KBN | nob | hvd | oljevirksohmheten |
| KBN | nob | hvd | oljefond |
| KBN | nob | hvd | oljemarked |
| KBN | nob | hvd | oljenæringen |
| KBN | nob | hvd | oljeskatt |
| KBN | nob | hvd | oljekorrigert budsjettunderskudd |
| NOTBASE | nob | hvd | oljeeksporterende land |
| NOTBASE | nob | hvd | oljeforbruk |
| NOTBASE | nob | hvd | oljeforsyning |
| NOTBASE | nob | hvd | oljekrise |
| NOTBASE | nob | hvd | olje- og energiminister |
| NOTBASE | nob | hvd | oljepolitikk |
| NOTBASE | nob | hvd | oljereserve |
| NOTBASE | nob | hvd | oljeselskap |
| NOTBASE | nob | hvd | oljesparing |
| NOTBASE | nob | hvd | oljevolumfaktor |
| NOTBASE | nob | hvd | oljebasert slam |
| NOTBASE | nob | hvd | oljefat |

Fullført

English (United Kingdom)

Nokia Nseries PC Su... | Terminor og temp... | Leksikograf_Tampere | Tampere_2009 | Termsøk — Tempo...

Planer ifm. infrastruktur



- Integrering med EuroTermBank (Tilde; CLARA)
- Utvide termportalen til nasjonal og på sikt europeisk portal



Takk for oppmerksomheten!

NO-CLARIN has received funding from the Research Council of Norway
CLARIN has received funding from
the European Community's Seventh Framework Programme
under grant agreement n° 212230