

Nynorskorpuset – historikk og status

Oddrun Grønvik¹ og Christian-Emil Smith Ore²

Korpusseminaret 3.2.2021

Høgskulen i Volda og Universitetet i Bergen

¹UiB/NO-AH, e-post: oddrun.gronvik@iln.uio.no

²ILN/UiO, e-post: c.e.s.ore@iln.uio.no

Føremålet med Nynorskkorpuset

Nynorskkorpuset vart oppretta i 2005 for å styrkja og skunda på redigeringa av Norsk Ordbok.

Ordboksredigering er intensivt arbeid. Alt som lettar tilgjenge til og ordning av grunnlagstilfang, bidreg til kvalitet og ferdigstilling etter plan.

Nynorskkorpuset skal også vera

- del av dei nynorsk Språksamlingane, med all slags tekst, og høg kvalitet på både tekst og støtteinformasjon.
- eit tilbod til alle som vil gå inn i nynorsk tekst via eit elektronisk korpus. Nynorskkorpuset skal vera lettbrukt for alle.
- eit tilbod til forskarverda - uavhengig av fagdisiplin. Nynorskkorpuset kan brukast av språk- og litteraturforskarar, men også av andre.

Kva slags korpus treng Norsk Ordbok?

Norsk Ordbok: «ordbok for det norske folkemålet og det nynorske skriftmålet»

- All målføretekst etter reformasjonen og all nynorsk tekst etter 1850 er interessant som kjeldegrunnlag.
- Eit korpus skal vera søkbart, det inneber eit behov for at innhaldet har standardisert rettskriving og lemmatisering.
- MEN: Nynorskkorpuset skal dekkja heile den nynorske skrifthistoria, frå midt på 1800-talet. Det må innehalda tekst med rettskriving frå heile perioden.
- Nynorskkorpuset skal ha med dei sentrale tekstane i nynorsk skriftkultur, men også ha spreining i sjanger, emne, kjønn, geografi osv.
- Eit særskilt minoritetsomsyn: Finna saklitteratur om emne som er lite dekte i dei breie sjanrane «avis», «tidsskrift» og «skjønnlitteratur»

Sigmund Skard om Norsk Ordbok (1932)

Eit kulturmaal skal ikkje ha levt so mange hundrad aar, fyrr trongen vaknar til aa faa oversyn over dei uttrykksmiddel det raader over; det er baade eit vitskapleg og eit praktisk krav. Slike nasjonalordbøker finn vi alt i dei eldste kultursamfund

...

Ordsamlinga or litteraturen skal freista faa til vegar so stort eit tilfang av utskrifter or bøker at ein for kvart ord har oversyn over korleis det vert og har vorte nytta i nynorsk *skrift*.

...

Nynorsken har i siste mannsaldrane vorte utforma paa nær sagt alle umkverve av kulturlivet; skal denne voksteren syna att i ordboka, maa arbeidet leggjast ovleg vidt.

Krav til tekst til Nynorskkorpuset

Nynorskkorpuset skal ha all slags tekst, publisert, korrekturlesen og komplett.

Teksten skal vera autentisk (førsteutgåve eller utgåve føretrekt av forfattaren)

Teksten skal vera kjeldefest (med sikker identitet) og registrert i bibliografibasen til Norsk Ordbok, som også skal ha eit papireksempplar i kjeldeboksamlinga

Dette krev at

- avtaleverket er på plass
- korpustekst blir teken vare på av mottakar
- At ein har oversyn over dei som har ytt bidrag til Nynorskkorpuset, og som kan koma til å ta kontakt

Forfattarar
som har
levert tekst til
Nynorsk-
korpuset,
fordi dei stø
føremålet



Somme av forfattarane som har gjeve tekst til Nynorskkorpuset

Førre bilete frå øvst til venstre:

Aina Basso, Unn Røyneland, Johannes Heggland, Sigmund Løvåsen

Under frå venstre:

Åse Gruda og Sigmund Skard, Halldis Moren Vesaas, Kjartan Fløgstad.

Representantar for alle dei som har vore glade for Nynorskkorpuset, stilt opp for Norsk Ordbok og gjeve Norsk Ordbok ret til vederlagsfri bruk av tekst.

Dugnadsgjengen attom Norsk Ordbok og dei nynorske språksamlingane



TINGA TIDLEG: Knut Seim tinga det fyrste ordboksheftet frå Det Norske Samlaget i 1950. 66 år seinare sit han med det tolvte og siste bandet i ordboksserien over det norske folkemålet.

Holingen Knut Seim Tinga Norsk Ordbok frå fyrste heftet i 1950, etterlyste Olav Sletto i Nynorsk-korpuset, og var til stades ved lanseringa av band 12 onsdag 9. mars 2016. (Hallingdølen 29.3.2016 s. 30-31).

Nynorskkorpuset - innhald oppsummert

Korrekturlesen tekst – tekstutval etter redaksjonelle behov og tilgjenge på tekst

Alt er fritt tilgjengeleg via korpusapplikasjonen (innanfor sitatretten)

- 85 % frå etter 1975, 75 % frå etter 2000
- Mykje avistekst, mest nyare, men også noko som er eldre (før 1920)
- Brukande, men holut, dekning av klassisk nynorsk skjønnlitteratur 1860 – 1980.
- Bibelen frå 1921, Austlids lesebok (1889), fleire ikoniske verk med stor spreining
- Gamalt or Setesdal (GOS I-VI)
- Ein del for born og ungdom frå 1900 og framover (120 tekstar)
- Ein del sakprosa i bokform, lærebøker og oversynsverk
- Dårlegast dekning 1920-1970

Nynorskkorpuset som redigeringsstøtte

Norsk Ordbok skal lagra alt tilfang på artikkelnivå, helst sortert og annotert.

Dette er eit basiskrav i høve til vitskapleg etterrettelegheit.

Korleis kan Nynorskkorpuset bidra til å møte dette kravet, og samstundes skunda på Norsk Ordbok?

Nynorskkorpuset gjev enkelt tilgjenge til råstoff og kjelder

Det er lett å ta inn nytt tilfang i sorteraren i redigeringsapplikasjonen

Det er lett å lagra bruksdøme i nye metaordboksartiklar

Metoden er «klipp og lim».

Nynorskkorpuset i Norsk Ordbok i – å bruksdøme og nokre tal

Status i dag:

Om lag 200 000 ordartiklar med status «ferdig»

203 000 bruksdøme med litterær kjelde

80 000 bruksdøme frå kjelder som kom med i Nynorskkorpuset 2005-2015

3300 ulike litterære kjelder for sitat

Av desse er 478 representerte i Nynorskkorpuset – litt under 15 %

Døma frå Nynorskkorpuset utgjer om lag 40 % av dei litterære bruksdøma.

Gjennomsnittleg tal belegg per kjelde for heilskapen – 62

Gjennomsnittleg tal belegg per kjelde frå Nynorskkorpuset – 168

Nynorskkorpuset vart hovudkjelda for litterære sitat.

Dei 10 mest brukte kjeldene for litterære sitat som finst i Nynorskkorpuset

Rang	Kjelde	Sjanger	Tal
1	Syn og Segn	Tidsskrift	8490
2	Dag og Tid	Avis	8349
3	Firda	Avis	3970
4	Fedraheimen	Avis	2661
5	Sogn Avis	Avis	2543
6	Garborg, Skrifter I-VI	Mest skjønnlitteratur	1992
7	Hallingdølen	Avis	1783
8	Den 17de Mai	Avis	1760
9	Vinje, Skrifter i Samling	Essayistikk og skjønnlitt.	1712
10	J. Skar: Gamalt or Setesdal	Folkeminne	1417
			34677

Ordbøkene er borte frå tabellen (dei høyrer heime i Ordbokshotellet). Både eldre og nyare tilfang er lyfta opp. **Raudmerkte** titlar er rykkja opp i bruk via Nynorskkorpuset. Sjangeren «folkeminne» er komen inn. Interessant i høve til normeringskrav til korpustekst. Det er skilnad på ordbøker og maskinell tekstanalyse.

10 mest siterte forfattarar med ein eller fleire tekstar i Nynorskkorpuset

Namn	Tal tekstar sitert	Tal tekstar sitert og i Nynorskkorpuset	Tal sitat
Olav Duun	30	22	4595
Arne Garborg	25	1 (Skrifter i Samling)	3081
Kjartan Fløgstad	12	9	2989
Aasmund Olavsson Vinje	7	1 (Skrifter i Samling)	2854
Kristofer Uppdal	20	16	2847
Tarjei Vesaas	36	33	2051
Rikard Berge	23	0	1890
Edvard Hoem	22	14	1747
Ivar Kleiven	9	2	1465
Jens Tvedt	18	2	1007
Kristoffer Janson	12	6	856

Merknader til «10 mest siterte forfattarar»

Blå farge – Rikard Berge er tett ekserpert og mykje sitert, men ikkje i Nynorskkorpuset.

Raud farge – Kristofer Uppdal var ei stigande sitatstjerne – dess meir tekst vi la inn i korpuset, dess meir vart han sitert.

Uppdal i Nynorskkorpuset er eit døme på samarbeid med ein forskar – Mads Claudi brukte Nynorskkorpuset i arbeidet med doktoravhandlinga si.

Korleis lagar ein korpus – føresetnader?

- Formål med korpuset og utvalskriterium
- Innhald
 - Medietypar (tekst, lyd, video)
 - Klarare opphavsrettar
- Datakvalitet
- Førebu tekstane
 - Vask, analyse, oppmerking manuell/maskinell
 - Val av kodeformat, (f.eks. TEI)
- Val av korpusprogramvare
 - Det finst mange tilgjengelege programpakker for korpus.

Programvare og implementering

- Modularisering
 - Bibliotek over originaltekstane
 - Automatisert innlegging av tekstar
 - Skilje grensesnitt frå søkemasineriet
 - Opent API for bruk og konstruksjon av alternative grensesnitt
 - Lenker til andre ressursar som tekstbeskrivingar (metadata)
- Ferdiglaga kontra eigenutvikla
 - Gratis, årleg lisens, open kode?
 - Eigenutvikla, kva hender når utviklaren(-ane) sluttar?

Nynorskkorpuset – føresetnadar

- Tekstane skal vere nærast mogleg dei komplette publiserte utgåvene
 - Koding i TEI
 - Publiseringsinformasjon og NO2014-bibliografikode i header
 - Sideskift, avsnitt og periodar (s-units) er markerte
- Nyare tekstar (etter 1938) er lemmatisert og POS-koda
- Tilgangsrettane er klarerte
 - Full tilgang for prosjektet NO2014
 - Litt avgrensa tilgang for allmenta (berre avsnitt)

Nynorskkorpuset – føresetnadar

- Korpuset skal full-integrerast i NO2014-prosjektet
 - Korpusgrensesnittet er ein saumlaus del av redigeringsverktøyet
 - Konkordansar og søk skal kunne lagrast som objekt i Metaordboka og i den semantiske sorteraren for NO2014
 - Det skal vere mogleg å lagre (delar av) konkordansar og kople dei til Norsk Ordbok gjennom Metaordboka for vitskapeleg dokumentasjon av redigeringa
- Korpuset skal ha eit web-basert grensesnitt for allmenta

Korpussystemet består av fire delar

1. Strukturert tekstbibliotek og fulltekstsøkesystem i ein relasjonsdatabase (her Oracle, kan vere MySql, PostgreSQL)
 - konstante empiriske data for etterprøving
2. Søkemaskineri (Corpus Workbench/CQP)
 - tradisjonelt, gratis, mykje brukt t.d. i Glossa
3. Web-basert grensesnitt mot 1 og 2 via API
 - 1 gir enkelt søk
 - 2 gir avansert søk med lemma og POS der det finst
4. Søk (basert på 3) og lagringsfunksjonalitet i redigeringsystemet (basert på 1) for Norsk Ordbok

Tekstbiblioteket med søkesystem

- Kvar tekst har ei lenke til bibliografien (NorMarc-format)
 - Har ofte lenke vidare til NB og Nasjonalbibligrfien
- Tekstane er delt opp i xml-element med ein kontekstvektor som syner taggehierarkiet
 - S-units (periode) er lågaste nivå
- Kvart eintydige ord (type) ligg i ein eigen tabell
- Kvar ordførekomst (token) ligg i ein eigen tabell
 - Lenke tilbake til s-unit (den lågaste eininga i kodinga)
 - Lenke til tilhøyrande eintydige ord (type)

Innførsel i NO2014-bibliografien (i NORMARC)

- *091 \$a**H.M.Ves.DFG**\$eHapax gjev ikkje grunnlag for artikkel\$fH.M.­Ves.­DFG
- *100 \$aVesaas, Halldis Moren (1907-1995)
- *245 \$aDu får gjera det, du\$bforteljing for barn\$cVesaas, Halldis Moren (1907-1995)
- *260 \$aOslo\$bAschehoug\$c1935
- *500 \$a126 s.
- *591 \$aspesifiser med sidetal\$bBand 12
- *655 \$alitteratur for born og ungdom
- *850 \$bNO2014\$zNynorskkorpuset 2014
- *856 \$uhttp://urn.nb.no/URN:NBN:no-nb_digibok_2007010400089
- *991 \$aBibsys

)

Koder i NORMARC-formatet:

091 \$aKode; utvalgskriterium; formatert ny kode 100 Navn 245 Tittel \$b annen tittelinformasjon, \$c ansvarsangivelse

260 \$a sted \$b Navn på forlegger, distributør, \$c år

655 Genre/form 500 Tekst 991 Katalog,

591 \$a Kommentar \$b med frå bnd nr.

850 \$aSamling \$z Lokal note

856 \$a URL

Tekstar og tekststruktur

Tabell over tekstar i korpuset med lenke til bibliografien

ID	DATUM	SHORTREF	CQP
...
30160	1935	H.M.Ves.DFG 1935	Y
...

Tabell over strukturen for kvar tekst i korpuset:

ID	XML_ID	DOCUMENT_ID	PAGE	TEXT
...
31436850	30160.2.1.21.3	30160	10	</S>
31436851	30160.2.1.21	30160	10	</P>
31436852	30160.2.1.22	30160	10	<P ID='30160.2.1.22'>
31436853	30160.2.1.22.1	30160	10	<S ID='30160.2.1.22.1'>
31436854	30160.2.1.22.1	30160	10	Dei tok køyrevegen op til kvenna .
31436855	30160.2.1.22.1	30160	10	</S>
31436856	30160.2.1.22.2	30160	10	<S ID='30160.2.1.22.2'>
...

XML_id identifiserer s-units eintydig, nyttast i koplinga til CQP og til større kontekst

Førekomst (token) og ord (type)

Tabell med tokens for setninga: «Dei tok køyrevegen op til kvenna .»

ID	W_ID	T_ID	POSITION	PAGE
...
106878628	84049	31436854	0	10
106878629	85588	31436854	1	10
106878630	658911	31436854	2	10
106878631	104415	31436854	3	10
106878632	84041	31436854	4	10
106878633	281219	31436854	5	10
106878634	84048	31436854	6	10
...

Standard metode for indeksering
til dømes brukt i Corpus Workbench/CQP

Tabell med types

ID	WORD	FREQ
...
658911	køyrevegen	47
...

Grensesnitt

- Resultatoppsett
 1. Konkordans med venstrekontekst, høgrekontekst og kjelde
 2. Resultatliste med frekvens, gir resultatoppsett 1 når ein vel ei resultatlinje
- Søkeoppsett
 - A. Enkelt (google-aktig)
 - Resultatoppsett 1 eller 2 avhengig av søk med eller utan jokerteikn
 - B. Avansert med innstillingar for resultatkonkordansen
 - Resultatoppsett 1
 - C. Avansert for frekvenssyning
 - Resultatoppsett 2

Opent grensesnitt

Norsk Ordbok

Ordbok over det norske folkemålet og det nynorske skriftmålet

VANLEG SØK

Kontekst:

ord sorterte etter

Linjetal

[Forklaring](#)

SØK MED FREKVENSSORTERING

[Forklaring](#)

Sjå eit oversyn over [teksttilfanget](#) i korpuset.

Sjå korleis tekstane i korpuset [fordeler seg per år](#).

Norsk Ordboks nynorskkorpus, avansert søk

Sjå rettleiinga nedst om du ønskjer å gjere avanserte søk i korpuset. Om du søker med frekvenssortering (søkjefeltet til høgre), får du fram talet på tilslag på den ordforma du søkte på. Søker du trunkert i dette feltet, kjem treffa opp i ei liste som er sortert på absolutt frekvens, med største frekvens øvst.

Nynorskkorpuset, vanleg søk

Trykkjer du på "søk" med det førehandsutfylte søkjeargumentet som allereie finst i søkjeboksen over, får du tilslag på alle førekomstane av ordet "hund" i dei tekstkjeldene som byrjar på "Ves.." i korpuset (til dømes Tarjei Vesaas og Vestnytt). Dømet

Ves.HF 1971, 84	deg - - Ætt- tavla di <i>hund</i> ! Er det sant at myrkret
Ves.HF 1971, 83	Myrkret løyner . Ætt- tavla di <i>hund</i> ! Ingen ser det . Er
VestNytt 2005	opna . - Det er ein <i>hund</i> ! fortel ho . Etter ei
Ves.Br 1966, 57	vart slikka på av ein blind <i>hund</i> , Tor- vil . I draumen
Ves.Br 1966, 44	endå er det ein altfor veldig <i>hund</i> , den sandfarga hunden . Midt
Ves.Br 1966, 49	til å skaka ? Ein sandfarga <i>hund</i> , det er alt . No
Ves.HF 1971, 83	er det ein hund ? Ein <i>hund</i> , ikkje meir , ein hund
Ves.KRH 1935, 79	kjend skraping . Det var ingen <i>hund</i> , men Helge som ikkje rakk
Ves.LH 1936, 126	hans var augo i ein banka <i>hund</i> , og biletet vart etter det
Ves.Br 1966, 45	ikkje mitt . Går det ein <i>hund</i> , så går det ein hund
Ves.GG 1925, 192	snute kom og kjælte . Ein <i>hund</i> , tenkte han , korleis hev
Ves.Br 1966, 45	hund , så går det ein <i>hund</i> - min er han ikkje .
VestNytt 2005	seier mora , som sjølv har <i>hund</i> . - Både son min og
Ves.Br 1966, 137	Vi vil bli kvitt det , <i>hund</i> . 22 ANDRE MØTET Aud og
Ves.BN 1968, 107	. Eg seier der er ingen <i>hund</i> . Der er heller ingen handel
Ves.S 1950, 203	eit bein og han sjølv ein <i>hund</i> . Det byrjar å skine i
Ves.Br 1966, 48	ser ein ? Ein sandfarga høgbeint <i>hund</i> . Det er alt . Det
Ves.Br 1966, 45	blir lyfte . Der er ingen <i>hund</i> . Det var berre våre armods
Ves.BN 1968, 107	fjortande natt . Der er ingen <i>hund</i> . Eg seier der er ingen
Ves.BN 1968, 108	kveld . Har aldri funnist nokon <i>hund</i> . Eg seier der er ingen
Ves.GG 1925, 190	ein skugge , er det ein <i>hund</i> . Ein hund her i fjellet
Ves.Br 1966, 49	. Det er ein flott veldig <i>hund</i> . Gjennomskaka ser vi han .
Ves.KH 1929, 66	paa husbonden so takksamt som ein <i>hund</i> . I vedskjolet var tvo slags
Ves.BN 1968, 108	. Der er ikkje snerten av <i>hund</i> . Ikkje snerten av hund .
Ves.Br 1966, 44	er klart det må kallast ein <i>hund</i> . Ingen ser han - lest
Ves.HH 1938, 18	at det gøyr som ein utestengd <i>hund</i> . Klas heisar nåler . Det

Resultatoppsett – frekvensliste

Søkjeord: [word="hund.*" & ref="Ves.*"]

hundre	90
hund	62
hundra	57
hunden	46
hundane	17
hundar	14
hundrevis	13
hundravis	5
hundreår	5
hundekjeks	4
hundredelar	4
hundeeigarar	3
hundeeigaren	3
hundrad	3
hundra-år	2
hundehus	2
hundreåret	2
hundreårgamle	2
hundens	2
hundeskrekk	2
hundeglam	2
hundretals	2
hundrade	2
hundra-året	1

Resultatoppsett – frekvensliste med konk.

Søkjeord: [word="hund.*" & ref="Ves.*"]

hundre	90
hund	62
hundra	57
hunden	46
hundane	17
hundar	14
hundrevis	13
hundravis	5
hundreår	5
hundekjeks	4
hundredelar	4
hundeeigarar	3

Søkjeord: A = [word="hund.*" & ref="Ves.*"]; B= "hunden"; C= intersection A B; cat C

Ves.BN 1968, 100	få stega mot siglaren . Då <i>hunden</i> kjem borttil , slepper har
Ves.BN 1968, 95	som eit skadd medvet . Men <i>hunden</i> har han kontakt med . Et
Ves.GG 1925, 195	just bui hans dette . Og <i>hunden</i> hadde funne husbonden :
Ves.Br 1966, 137	Det har ingen sagt . Om <i>hunden</i> rører deg , har det ingen
Ves.KH 1929, 38	skotvare , sa han som aatte <i>hunden</i> . - Ja det hev du
Ves.BN 1968, 95	mildare . Det viser seg at <i>hunden</i> følgjer med bak buskane
Ves.GG 1925, 194	, urøleg , alt til at <i>hunden</i> kom og stakk snuten bur
Ves.Br 1966, 74	sat vidare . Slikka på av <i>hunden</i> - kva er det for noko
Ves.Br 1966, 107	? Ein er slikka på av <i>hunden</i> . Det er ikkje sant !
Ves.Br 1966, 134	? Ein er slikka på av <i>hunden</i> . Anten ein drøymer eller
Ves.BN 1968, 94	liene svarar det . Då blir <i>hunden</i> vill , anten det er redsle
Ves.BN 1968, 93	rad med signal frå den bortgøymde <i>hunden</i> . Hus ? Ikkje hus å

NO2014 nynorsk korpus - Google Chrome

Not secure | no2014.uib.no/cgi-bin/korpuset/cqp-context.cgi?textid=13369219&s...

Ves.LEL 1953

Er du **hunden** ? Hundelekket gnagar ovnsfoten under draum , og utrekt på teppet ligg dyret uskuldig , uskuldig og blundar - men jakttraumen går med kvink og brårykk . Markene er sæle og blodige .

54
54
209
50
3, 91
3, 91
, 38
, 99
, 83
, 190
52
98

gråten bak steinen . Men den *hunden* var som livs levande enn
, Aud . Det er den *hunden* . Å koma bort til Aud
Eg kom berre i hug den *hunden* for totre veker sidan . -
Som etter eit linjalstrek kjem den *hunden* , sakte , så nær etter
sæle og blodige . Er du *hunden* ? Det står ingen over deg
EIN DRØYMANDE HUND ? Er du *hunden* ? Hundelekket gnagar ov
klemde enda til snuten av den eine *hunden* ned i mosen og sa fillebi
ropar det . Endå veldigare enn *hunden* . - Hoi ! Hei !
rykte i honom . So er *hunden* yver honom og slikkar .
rykte i honom , so er *hunden* yver honom og slikkar .
etter ein linjal går den flotte *hunden* vegen sin tilbake . Måne
ikkje til orde heller , for *hunden* hissar seg sjølv opp meir

Rettleiar for utfylling av søkeargument

AVANSERT SØK

[word="gjen.*" & msd="SAFF.B"]

Vanlig konkordans (KWIC) Kontekst: ord sorterte etter Linjetal
 frekvenssøk

gjen ✕

ordform: ord grunnord

byrjing slutt del heile alle slags

Substantiv Adjektiv Partisipp Determinativ Verb Pronomen Andre

- substantiv
- appellativ proprium
- hankjønn hokjønn inkjekjønn
- eintal fleirtal
- genitiv alt anna
- ubunden bunden

Nullstill

Korpuset, Metaordboka og Norsk Ordbok

- Metaordboka er ein felles indeks til heterogent kjeldemateriale
 - 771 886 artiklar
 - berre nynorsk 433.462,
 - berre bokmål 224.950,
 - begge 113.451
 - Artikkeltitlar:
 - oppslagsord, ordklasse, språk, normering
 - Artikkeltittel
 - Lenker til kjeldedatabasar i alt 12 ulike typar inkludert:
 - Norsk Ordbok (i-å) og ekserpt frå Nynorskkorpuset
 - Kvar kjeldedatabase er ein sjølvstendig database med eit eige grensesnitt

Korpuset, Metaordboka og Norsk Ordbok

- S-unit er den minste eininga
 - Den eintydige identifikatoren til s-units knytt saman tekstene i Corpus Workbench med tekstsysteemet i Oracle
 - Kontekstvektoren gjer det mogleg å syne større og større kontekstar for ein gitt ordførekomst
- Korpuslinjer (ekserpt) kan lagrast
 - Tekst-id, korpus, s-unit-id, søkeargument, tilslagsord
- Korpuslinjer kan koplast til
 - Metaordboka
 - den semantiske sorteraren i redigeringsverktøyet for Norsk Ordbok

Eksempel på lagra konkordanslinjer

ID	TEXTID	XML_ID	CORP US	SEARCHSTR	REF	TEXT
217583	8691911	27187.2.1.1.2.2.10.1	CQP	"fri.*" "ut"	Ezra/Neh 1921 1,, 10	fria ut
217582	9144145	27255.2.1.1.6.3.7.2	CQP	[word="veggfli.*"]	FløgstadKS 1991 25-26	veggflisene
217581	13776268	28772.2.1.7.3.7.1	CQP	[word="veggfli.*"]	DrangeRL 2007, 118	veggflisene
217580	13715024	28766.2.1.2.32.25.2.5	CQP	[word="(V v)(e æ)lfo.*"]	A.NordalJ 1989, 271	velforsynt
217579	9734496	27372.2.1.1.6.2.103.6	CQP	[word="(V v)(e æ)lfo.*"]	E.HoemO 2000, 54	velforsynt
217578	8471118	27027.2.1.1.3.142	CQP	[word="(B b)(a e)r.*e"] [word="vent"]	SS 1998, 146	bare vent
217577	28616837	30027.2.1.278.5.3	CQP	[word="(B b)(a e)r.*e"] [word="vent"]	NPK 2008	berre vent

Tabellen kan med få endringar og nyttas med CorpuscleLex som korpussøkesystem. Ein treng ein kolonne med s-unit identifikatoren nytta i CorpuscleLex.

Korpussøk integrert i redigeringsverktøyet

Universitetenes samlingsdatabaser, Språk - [Søk i korpus]

Metaordboka Normaliser Utforsker NO2014 Nynorskkorpus Søk i korpus

Navigatør NO2014 nynorskkorpus Smart venstresortering

NO 2014 nynorskkorpus

Kontekst: 6 ord Sortert etter høgre Linjetal stikkeprøve Søk: [word="hund.*" & msd="SA.*"] Søk Dokumentasjon

Dra kolonner hit for å gruppere resultatet

Litteratur-id	Venstrekontekst	Tilslag	Høgrekontekst
DT 2001	tradisjonar som iallfall gjekk tilbake til	hundreårsskiftet	. Halvdan Koht , i den
P_Lars_LienNF 2009, 302	med fotografiet på i åra etter	hundreårsskiftet	. I 2004 presenterte for eksl
NoAllk X 491-492	føre 1900 , andre stader etter	hundreårsskiftet	. Kring 1930 var talet på
HoelMM 2009, 301	som kom til i fylkesorganiseringsbølga kring	hundreårsskiftet	. Laget heldt det gåande nc
Gund.SN 1983, 76	på Bergensbanen stod ferdig like etter	hundreårsskiftet	. Like eins vart det bygd
DT 2002	eit nytt høgdepunkt frå og med	hundreårsskiftet	. Likevel er det ikkje sjølv
DT 2001	til ei polarmakt kring det førre	hundreårsskiftet	. Otto Sverdrup , Nansen o
HoelMM 2009, 155	fløya i Venstre i åra kring	hundreårsskiftet	. Som lokalpolitikar , storting
L_Opstad 1990, 24	stelte om vi går fram mot	hundreårsskiftet	1500 / 1600 . I 1593
DT 2000	av verda det siste året før	hundreårsskiftet	2000 årstalet har tre nullar v
A.Bj.SV [2004], 61	I manns minne . Daglegliv ved	hundreårsskiftet	II " . Det Norske Samlaget
Gund.SN 1983, 72	lagt på tverrgåande sviller . Rundt	hundreårsskiftet	begynte ein òg å impregner
P.KvammeTA 1999 39,2	absolutte barndom . I åra rundt	hundreårsskiftet	budde han i Frankrike , som
SoA 2007	relansera den store personlegdomen frå førre	hundreårsskiftet	både i bok og på film
HoelMM 2009, 367	arbeidarvenlege fløya i Venstre . Etter	hundreårsskiftet	er dei radikale målmiljøa si ti
Gund.SN 1983, 17	posttilsette , og tida fram til	hundreårsskiftet	er derfor prega av små fram
FurreVH 1992, 20	med om lag 50 % frå	hundreårsskiftet	fram til 1920 . Den nye
A_Fidjest_Sivle 2007, 283	seg stor og viktig , innan	hundreårsskiftet	hadde ho allereie over 5000
SS 1975:9 550,1	frå øya . Bortimot 300_000 ved	hundreårsskiftet	har skrumpa inn til mindre e
HoelMM 2009, 157	17de Mai . finn ein før	hundreårsskiftet	heilt klart ei arunneaaande

Nivå: 0

Det er Et liv i isen av Jan Ove Ekeberg ,
soga om polarkokken Adolf Henrik
Lindstrøm , mannen som var med på alle
dei store polferdene som gjorde landet
vårt til ei polarmakt kring det førre
hundreårsskiftet .

DT 2001
Dag og Tid
Oslo, 1962-

Sist brukte
Verktøy

ceo@eddhauge.uib.no

Korpussøk integrert i redigeringsverktøyet

Universitetenes samlingsdatabaser, Språk - [Søk i korpus]

Metaordboka Normaliser Utforsker NO2014 Nynorskkorpus Søk i korpus

Navigator NO2014 nynorskkorpus Smart venstresortering

NO 2014 nynorskkorpus

Kontekst: 6 ord Sortert etter høgre Linjetal stikkeprøve Søk: [word="hund.*" & msd="SA.*"] Søk Dokumentasjon

Dra kolonner hit for å gruppere resultatet

Litteratur-id	Venstrekontekst	Tilslag	Høgrekontekst
DT 2001	tradisjonar som iallfall gjekk tilbake til	hundreårsskiftet	. Halvdan Koht , i den
P_Lars_LienNF 2009, 302	med fotografiet på i åra etter	hundreårsskiftet	. I 2004 presenterte for eks
NoAllk X 491-492	føre 1900 , andre stader etter	hundreårsskiftet	. Kring 1930 var talet på
HoelMM 2009, 301	som kom til i fylkesorganiseringsbølga kring	hundreårsskiftet	. Laget heldt det gåande nc
Gund.SN 1983, 76	på Bergensbanen stod ferdig like etter	hundreårsskiftet	. Like eins vart det bygd
DT 2002	eit nytt høgdepunkt frå og med	hundreårsskiftet	. Likevel er det ikkje sjølv
DT 2001	til ei polamakt kring det førre	hundreårsskiftet	. Otto Sverdrup , Nansen o
HoelMM 2009, 155	fløya i Venstre i åra kring	hundreårsskiftet	. Som lokalpolitikar , storting
L_Opstad 1990, 24	stelte om vi går fram mot	hundreårsskiftet	1500 / 1600 . I 1593
DT 2000	av verda det siste året før	hundreårsskiftet	2000 årstallet har tre nullar v
A.Bj.SV [2004], 61	I manns minne . Daglegliv ved	hundreårsskiftet	II " . Det Norske Samlaget
Gund.SN 1983, 72	lagt på tvergåande sviller . Rundt	hundreårsskiftet	begynte ein òg å impregner
P.KvammeTA 1999 39,2	absolutte barndom . I åra rundt	hundreårsskiftet	budde han i Frankrike , som
SoA 2007	relansera den store personlegdomen frå førre	hundreårsskiftet	både i bok og på film
HoelMM 2009, 367	arbeidarvenlege fløya i Venstre . Etter	hundreårsskiftet	er dei radikale målmiljøa si ti
Gund.SN 1983, 17	posttilsette , og tida fram til	hundreårsskiftet	er derfor prega av små fram
FurreVH 1992, 20	med om lag 50 % frå	hundreårsskiftet	fram til 1920 . Den nye
A_Fidjest_Sivle 2007, 283	seg stor og viktig , innan	hundreårsskiftet	hadde ho allereie over 5000
SS 1975:9 550,1	frå øya . Bortimot 300_000 ved	hundreårsskiftet	har skrumpa inn til mindre e
HoelMM 2009, 157	17de Mai . finn ein før	hundreårsskiftet	heilt klart ei arunnelegande

Nivå: 1

Det er merkeleg korleis hendingar og opplevingar flettar seg saman i ein spesiell situasjon . Her ligg ein hugstur på sjukehus og ventar på å få operert ei tå som det har sett seg vondt i . Då sendar fjernsynet Ni liv , saga om Jan Baalsrud som skar av seg tærne for å hindre at det vart kaldbrann . Så kjem ein god ven innom , og til trøyst har han med seg ei bok . Det er Et liv i isen av Jan Ova Elshaug , saga om

DT 2001

Dag og Tid

Oslo, 1962-

Sist brukte

Verktøy

ceo@eddhauge.uib.no

Korpussøk integrert i redigeringsverktøyet

Universitetenes samlingsdatabaser, Språk - [Søk i korpus]

Metaordboka Normaliser Utforsker NO2014 Nynorskkorpus Søk i korpus

Navigator NO2014 nynorskkorpus Smart venstresortering

NO 2014 nynorskkorpus

Kontekst: 6 ord Sortert etter høgre Linjetal stikkeprøve Søk: [word="hund.*" & msd="SA.*"] Søk Dokumentasjon

Dra kolonner hit for å gruppere resultatet

Litteratur-id	Venstrekontekst	Tilslag	Høgrekontekst
DT 2001	tradisjonar som iallfall gjekk tilbake til	hundreårsskiftet	. Halvdan Koht , i den
P_Lars_LienNF 2009, 302	med fotografiet på i åra etter	hundreårsskiftet	. I 2004 presenterte for eks
NoAllk X 491-492	føre 1900 , andre stader etter	hundreårsskiftet	. Kring 1930 var talet på
HoelMM 2009, 301	som kom til i fylkesorganiseringsbølga kring	hundreårsskiftet	. Laget heldt det gåande nc
Gund.SN 1983, 76	på Bergensbanen stod ferdig like etter	hundreårsskiftet	. Like eins vart det bygd
DT 2002	eit nytt høgdepunkt frå og med	hundreårsskiftet	. Likevel er det ikkje sjølv
DT 2001	til ei polamakt kring det førre	hundreårsskiftet	. Otto Sverdrup , Nansen o
HoelMM 2009, 155	fløya i Venstre i åra kring	hundreårsskiftet	. Som lokalpolitikar , storting
L_Opstad 1990, 24	stelte om vi går fram mot	hundreårsskiftet	1500 / 1600 . I 1593
DT 2000	av verda det siste året før	hundreårsskiftet	2000 årstale har tre nullar v
A.Bj.SV [2004], 61	I manns minne . Daglegliv ved	hundreårsskiftet	II " . Det Norske Samlaget
Gund.SN 1983, 72	lagt på tvergåande sviller . Rundt	hundreårsskiftet	begynte ein òg å impregner
P.KvammeTA 1999 39,2	absolutte barndom . I åra rundt	hundreårsskiftet	budde han i Frankrike , som
SoA 2007	relansera den store personlegdomen frå førre	hundreårsskiftet	både i bok og på film
HoelMM 2009, 367	arbeidarvenlege fløya i Venstre . Etter	hundreårsskiftet	er dei radikale målmiljøa si ti
Gund.SN 1983, 17	posttilsette , og tida fram til	hundreårsskiftet	er derfor prega av små fram
FurreVH 1992, 20	med om lag 50 % frå	hundreårsskiftet	fram til 1920 . Den nye
A_Fidjest_Sivle 2007, 283	seg stor og viktig , innan	hundreårsskiftet	hadde ho allereie over 5000
SS 1975:9 550,1	frå øya . Bortimot 300_000 ved	hundreårsskiftet	har skrumpa inn til mindre e
HoelMM 2009, 157	17de Mai . finn ein før	hundreårsskiftet	heilt klart ei arunneaaande

Nivå: 2

Sterk kost til trøyst

Det er merkeleg korleis hendingar og opplevingar flettar seg saman i ein spesiell situasjon . Her ligg ein hugstur på sjukehus og ventar på å få operert ei tå som det har sett seg vondt i . Då sendar fjernsynet Ni liv , saga om Jan Baalsrud som skar av seg tærne for å hindre at det vart kaldbrann . Så kjem ein god ven innom , og til trøyst

DT 2001

Dag og Tid

Oslo, 1962-

Sist brukte

Verktøy

ceo@eddhauge.uib.no

Bruksdøme kan flyttast frå korpus til artikkel

NO 2014 nynorskkorpus

Kontekst: 6 ord Sortert etter høgre Linjetal stikkeprøve Søk:

Sortering glo NK Operasjonar

Ikona	Tekst	Ny kolonne
	glod (Ves.Brannen 1961, 175)	
	glod (Ves.HM 1945, 83)	
	glod (Ves.S 1950, 21)	
	glør (BySBR 1971, 55)	
	glør (Setr.DSM 1946, 214)	
	glørne (FløgstadD7K 1986, 183)	

Dra kolonner hit for å gruppere resultatet

Men han hadde også lært seg å styra eldhugen sin . gøyma både den og sinnet djupt i hjarta liksom ein eld under oska , klar til å flamma opp i forbitring når ein ny urett bles til glørne .

NO 2014 nynorskkorpus

Kontekst: 6 ord Sortert etter høgre Linjetal stikkeprøve Søk:

word="glør *" & med="SA *"

Dra kolonner hit for å gruppere resultatet

Litteratur-id	Venstrekontekst	Tilslag	Høgrekontekst
Farbr_P 1997, 199	Men under ligg det framleis store	glør	! Ein gloande gr
Sundl.TB 1960, 16	til å blåsa litt til synde-	glørne	i kvar manns hj
Setr.DSM 1946, 214	snøen , og fjelltoppane i sør	glør	som glødd stål .
BySBR 1971, 55	Åsgard ! Det er som tanken	glør	i han , da han g
FløgstadD7K 1986, 183	når ein ny urett bles til	glørne	. Og ingen , såg
FløgstadA 1996 98-99	Romsås og ser videoflimmeret blåsa til	glørne	av dei latente d
Sundl.TB 1960, 109	ved det , pustar heller til	glørne	og flirer stygt , e
Firda 2000	mm . Det trengst kol til	glør	og ved til eld og
SoA 2008	vind har gjeve ny næring til	glørne	slik at elden kur
SoA 2008	vinden har gjeve ny næring til	glørne	og at kastevind
MonaBGT 2004, 117	yrkesstatus . Dette var pust til	glørne	i den begynnan
E.Oft.FV 1954, 20	Seth seg framover og sparka til	glørne	. Då slo Bedagi
Hovdh.FGA 1993, 85	. Somme gamle budeier tok tre	glør	frå peisen som c
SoA 2008	søndag er Molde . Politiet trur	glør	frå bål var brann
DT 2001	. **** Våre viljor , ulmande	glør	og bankande hj
DT 2001	Irak er bensin til alle ulmande	glør	i regionen . For
SoA 2008	etter nokre timar skrapar dei ut	glørne	, gjer omnen rei
Sunnh 2007	slik han skulle . Det var	glør	rett under peise
Hordaland 2010	når kveld , og det var	glør	att i peisen . Sn

Ein kan flytta korpusbelegg med kopier og lim (Ctrl-C og Ctrl-V) frå korpuset til artikkel-sorteraren, der tilfanget som ligg attom kvar ordboksartikkel ligg samla. Slik kjem Norsk Ordbok i møte kravet om vitskapleg etterrettelegheit

- og frå Nynorskkorpuset til Metaordboka

The screenshot shows the 'Søk i korpus' application interface. The main window is titled 'Søk i korpus' and displays search results for the term 'reimaskøyte' in the 'NO 2014 nynorskkorpus'. The interface includes a search bar, a list of search results, and a detailed view of the selected entry.

Search Parameters:

- NO2014 nynorskkorpus
- Smart venstresortering
- Kontekst: 6 ord
- Sortert etter: høgre
- Linjetal: stikkeprøve
- Søk:

Search Results Table:

Litteratur-id	Venstrekontekst	Tilslag	Hø
AustL.B 1889, 228	hunden og eit par le' r-	reimar	til a
A.NordaJ 1989, 168	til andre ting , så som	reimar	til h
Tv.BP, 203	paa - alt det rygg og	reimar	tølc
Up.AE 1920, 186	fram ein mon til . Og	reimar	tøy
Ap 1921 22,, 25	hadde bunde honom til slag med	reimarne	, så
Maul.N 1919 6-7	sjaa paa jarn- hjelmarne og bøffel-	reimarne	og
SoA 2008	og trikotar medan me kom med	reimaskøytene	og
SoA 2009	fresvikjene for det meste stille med	reimaskøyter	. Ei
SoA 2009	blitt smått med skøytegang . Sokalla	reimaskøyter	var
FløgstadFF 1980, 208	smergelskiva . Det var også ein	reimdriven	fres
Du.IS 1923, 77	ofte slik og slarka utan å	reime	skc
A.NordaJ 1989, 133	skar til av høveleg skinn (reime-	skir
Up.Ku_II 1947, 178	eg . I strøppen på liv-	reimen	. Tu
HalseT 1927, 94	nedpå brygga så det small i	reimendane	. "I
KohtGE, 25	fekk han løyst , og ingen	reimende	so i
Flod.EF 1968, 134	så drog han i den vesle	reimenden	har
Flod.EF 1968, 56	dei , - dei å opp	reimene	. Er
Flod.EF 1968, 216	til alt mogleg som jarntråden og	reimene	er c
NorheimA 2008, 96	ligg alt ferdig . Dei innvendige	reimene	er l.
J.Ørj.P 1988, 87	eins tor kniv , og på	reimene	ove
H.Sandel 1980, 13	. Muhammeds navle . Eg strama	reimene	run
H.Sandel 1980, 150	følqde med . Eq kontrollerte at	reimene	sat

Metadata for 'reimaskøyte':

- Oppslag-id: 11691020
- Ordform: reimaskøyte
- Målform: Nynorsk
- Normalnivå: Hovudform
- Frå: 01.01.1938
- Til: 31.12.9999
- Grammatikk: subst fem ubf sg appell

Oppslag List:

- reim*skøise f ok 2
- reim*skifter m ok 1
- reim*skinn, reimskinn n, n ok 15
- reim*skive, reim*skive, rem f, f r ok 23
- reim*skive*krans m ok 1
- reim*sko, reim*sko m, n ok 11
- reim*skokle f ok 1
- reimskrue m ok 1
- reim*skurd m ok 2
- reim*skøyte f ok 3
- reimaskøyter Ukje ok
- reimaskøyter Ukje ok
- reimaskøytene Ukje ok

Thumbnail: A small image showing a yellow and black pattern, likely a logo or a decorative element related to the search results.

Nynorskkorpuset – dei viktigaste punktene

- Forståing for at nynorske tekstsamlingar er viktig infrastruktur
- Gode kontaktar til teksteigarar og tekstprodusentar
- Ei velpleidd tekstsamling av høg kvalitet med klarerte tilgangsrettar
- Eit modulært korpussystem knytt til redigeringsverktøyet for Norsk Ordbok og Metaordboka
- Tilgang for allmenta til materialet

Kva no – kva treng vi?

- Meir tekstmateriale
- Andre og betre verktøy for tekstanalyse
 - Morfo-syntaktisk analysator for moderne og eldre tekst
 - Andre og/eller fleire korpussystem
- Fleire typar kjeldemateriale
 - Trebankar
 - Talekorpus
 - Målføremateriale (integrere LIA-materialet)
- Nytt **grensesnitt** til samlingsdatabasane
 - For dei dataressursane som allereie er i systemet og for nye medietypar
 - Modularisert og med opne lenka data teken på alvor

Nokre kjelder

- Grønvik, Oddrun (2006): Verknader av digitalisering på materialvurdering, redaksjonell metode og opplæring. I: Lorentzen, Henrik, og Lars Trap-Jensen (red): Nordiske Studier i Leksikografi. Rapport fra Konference 24.-28. maj 2005. Skrift nr 9. København. s. 129-142.
- Komplette samling. Hallingdølen 29.3.2016 s. 30-31
- Ore, Christian-Emil Smith. Nettordbøker og Norsk Ordbok – hvordan etablere en vitenskapelig nettordbok. I: Nordiska studier i lexicografi 11. Rapport från Konferensen om lexicografi i Norden, Lund 24–27 maj 2011. Nordisk forening for lexicografi 2012 ISBN 978-91-85333-42-4. s. 486-499
- Ore, Christian-Emil Smith. Edisjonsfilologi, fjernlesning og digitale utgaver. I: Textkritik som analysemetode. Bidrag till en konferens anordnad av Nordiskt Nätverk för Editionsfilologer 2-4 oktober 2015. Svenska vitterhetssamfundet 2017 ISBN 978-91-7230-187-0. s. 26-44
- Ore, Christian-Emil Smith; Grønvik, Oddrun. Comparing Orthographies in Space and Time through Lexicographic Resources. I: Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts. Znanstvena založba Filozofske fakultete 2018 ISBN 978-961-06-0097-8. s. 159-172
- Ridings, Daniel & Oddrun Grønvik (2012): A corpus based method for a diachronic study of the central vocabulary of New Norwegian. I: Eaker, Birgit et al (red): Nordiska studier i lexicografi 11. Rapport från Konferensen om lexicografi i Norden Lund 24-27 maj 2011. s. 524-534.
- Skard, Sigmund (1932): Norsk Ordbok: Historie, plan, arbeidsskipnad. Oslo. Det Norske Samlaget.