

Tekster og metadata i Norchron

Ivar Berg og Kari Kinn

Norchron oppstartsworkshop
17. november 2025

Tekster i Norchron

Tekster i Norchron

Litt om

- ▶ Hvilke kjelder fins til norsk språkhistorie?
- ▶ Hva tar vi med i Norchron?
- ▶ Hva tar vi **ikke** med?

Kjelder til norsk språkhistorie

III. 1: Existing material sorted by text type and period.					
	Ancient Nordic (<i>urnordisk</i>)	Early Old Norwegian (<i>eldre gammalnorsk</i>)	Late Old Norwegian (<i>yngre gammalnorsk</i>)	Middle Norwegian (<i>mellomnorsk</i>)	Early Modern Norwegian (<i>eldre moderne norsk</i>)
Source type	200–700	700–1050	1050–1350	1350–1550	1550–1814
Runes	Print editions (outdated, no annotation)	Print editions (newer finds not included, no linguistic annotation)			
	Scandinavian Runic-Text Database & Runor (Uppsala/Stockholm), RuneS Database (Göttingen), Runic archives (Museum of Cultural History, UiO)				
Manuscripts			<i>Menota corpus</i> (digital editions, many partly or fully annotated)		
Diplomas			<i>Diplomatarium Norvegicum</i> (1847–; digitised print editions with no annotation)		
Dialect texts					Print editions (no encoding or annotation)

Runeinnskifter

- ▶ Ingen stor tekstmengde, men de eldste skriftlige kjeldene vi har
 - ▶ Eldre runer på urnordisk (til ca. 700)
 - ▶ Yngre runer på eldre/ynge gammalnorsk til 1400-tallet
- ▶ Mange, særlig eldre, er vanskelige å lese og tolke
 - ▶ Døme: Straumsbrynet (seint 500-tal): Tydelige runer, tre heilt ulike tolkingar
- ▶ Krever mye filologisk/runologisk arbeid
- ▶ Må lage ny, digital utgave

Gammalnorske handskrifter

- ▶ Mange er publisert i diplomatariske transkripsjoner i Menota (*Medieval Nordic Text Archive*)
 - ▶ Mange med morfologisk annotasjon
 - ▶ Noen også syntaktisk annotert (Menotec-prosjektet)
- ▶ De fleste “skjønnlitterære” tekster; mangler noen lovtekster
 - ▶ Noen **seine** lovhandskrifter hadde vært fint for geografisk/kronologisk bredde
 - ▶ Ellers er den “klassiske” gammalnorsken ca. 1270–1310 godt dekket

Diplom (i)

- ▶ Korte tekster med hovedsaklig rettslig (i vid forstand) innhold
 - ▶ Dokumentasjon av gardhandel og rettslige kunngjøringer
- ▶ Brukes i Skandinavia om “brev” mer generelt
 - ▶ Litt mer variert innhold i siste del av perioden
- ▶ Utgitt i *Diplomatarium Norvegicum* (DN) I–XXIII (1847–2011), totalt vel 20 000 tekster
 - ▶ Digitalisert, men utgavene (før bind XXI) følger ikke moderne filologisk standard
- ▶ DN inkluderer også tekster på andre språk (latin, lågtysk, svensk, dansk) og seine avskrifter

Diplom (ii)

- ▶ Diploma tom. 1310 publisert i Menota
 - ▶ Full morfologisk annotasjon
- ▶ Knapt 3000 diplom til rundt 1450 transkribert av Gamalnorsk ordboksverk
 - ▶ Mangler annotasjon
 - ▶ Uklare kriterier for utvalget; ser ut til å være mye sørøstlandsk (særlig sist i perioden)
- ▶ Må supplere for perioden 1450–1550
 - ▶ Velge ut diplom (etter hvilke kriterier?)
 - ▶ Kontrollere transkripsjoner fra DN mot originaler/foto

Dialekttekster fra dansketida

- ▶ Det eneste som blei skrevet på norsk i denne perioden
 - ▶ Avgrenser oss til det som direkte bygger på norsk talemål
- ▶ Lite sannsynlig at det fins mange andre (lange) tekster av samme type
 - ▶ Sjøl om det fins norske innslag i tekster på dansk
 - ▶ Performativ dialektbruk; dansk (etter beste evne) var normalen
- ▶ Utgitt av Venås (1990), Dalen & Hagland (1985), Bull (1942)
- ▶ Må lage digital utgave og annotere tekstene

Hva tar vi **ikke** med?

Runer: Avgrenser oss til “lexically meaningful” (søknaden)
innskrifter

Gam.norsk: Prioriterer ikke nye tekster; eventuelt noen lovtekster

Mel.norsk: Kan ikke ta med alle diploma

- ▶ Endre Mørck (2018: 297) regner 6244 diplom fra dagens Norge i seinmiddelalderen
- ▶ Tar heller ikke med jordebøkene (mest stedsnavn)

Dansketid: Nøyer oss med publiserte utgaver

- ▶ Tar ikke med reine ordlister
- ▶ Leiter ikke etter nytt materiale

Gamle og nye tekster

- ▶ Inkluderer eksisterende tekster fra Menota
- ▶ Nye, digitale utgaver av runeinnskrifter, diplom (gammal- og mellomnorsk) og dialekttekster fra dansketida
- ▶ Legger til språklig annotasjon og ulike typer **metadata**

Metadata

Metadata (og språklig annotasjon)

- ▶ **Metadata:** informasjon **om** tekstene
 - ▶ f.eks. opphavssted, datering, sjanger/teksttype +++
 - ▶ Gjør det mulig å avgrense materialet etter gitte kriterier
 - ▶ Glossa: klikkbar metadatameny – man velger **verdier i ulike metadatakategorier** og avgrenser et delkorpus
- ▶ Språklig annotasjon: oppmerking (tagging) av selve tekstene
 - ▶ for hvert ord (token): ordklasse og lemma (“oppslagsord”)
 - ▶ for deler av materialet: morfologi (bøyingsformer) og syntaks (frase- og setningsoppbygging)
 - ▶ Gjør det mulig å søke etter ord – eller grammatiske fenomener uten ordformer
 - ▶ Men hovedfokus for oss nå er altså metadata

Metadatamenyer fra andre korpus: LBK

All 27081 texts (99952936 tokens) selected

Show

Publikasjon

Kategori

Utgiver

År

Sted

Oversatt?

Emne

FORFATTER ^

Navn

Kjønn

Fødselsår

Oversetter

DETALJER v

Hide filters

Reset form

Leksikografisk bokmålskorpus



Simple | Extended | CQP query

Search

Or...

- Nytt søkegrensesnitt!
- [Les om korpuset](#)
- [Les brukerveiledningen til korpuset](#)
- [Brukerlisens](#)
- [Rapporter om feil i korpuset](#)
- [Bruk den forrige versjonen av søkegrensesnittet](#)
- [Hvordan referere til korpuset](#)

Metadatamenyer fra andre korpus: Bokselskap-korpuset

All 414 texts (21876589 tokens) selected

Show

Open in Voyant

Tekst-ID

Tittel

Forfatter/redaktør



A.O. Vinje

Aasmund Olavsson Vinje

Aasta Hansteen

Absalon Pederssøn Beyer...

Alexander L. Kielland

Alvilde Prydz

Nynorsk eller bokmål/dansk

Kjønn

Sjanger

Førsteutgåve

Utgåve-årstal

Hide filters

Reset form

Bokselskap-korpuset

bokselskap

Simple | [Extended](#) | [CQP query](#)

Search

Or...

- Nytt søkegrensesnitt
- Tekstene i Bokselskap-korpuset er hentet fra bokselskap.no. Nettstedet er en ebokportal utviklet av Det norske språk- og litteraturselskap (NSL) i samarbeid med Nasjonalbibliotek. [Les mer her.](#)
- Korpuset vil bli oppdatert med nye tekster hvert år. Siste oppdatering: 10. februar 2025.
- [Rapporter feil i korpuset](#)
- [Hvordan referere til korpuset](#)

Metadatamenyer fra andre korpus: NDC (talespråk)

All 737 speakers (2754289 tokens) selected from 183 places in 5 countries

Show

Map

Open in Voyant

LOCATION ^

Place

Area

Region

Country

INFORMANT ^

Informant code

Birth year

Gender

Age

Age group

Recording Year

Genre

Hide filters

Reset form

Nordic Dialect Corpus v. 4.0



Nordic Dialect Corpus
Scandiasyn/NORMS

Simple | [Extended](#) | [CQP query](#)

Search

Or...

- New search interface!
- NB! This version - v. 4.0 - contains dialect recordings and transcriptions from 1998 - 2015 only. The older recordings from Målførearkivet are moved to [LIA Norwegian - Corpus of Old Dialect Recordings](#)
- [Go to v. 3.0](#) to search the NDC corpus with the old Målførearkiv recordings!
- [Read the User Manual for Nordic Dialect Corpus v. 4.0](#)
- [Homepage: Transcription guidelines, translation lists, etc](#)
- [Recording locations](#)
- [User license for the corpus](#)
- [Report errors in the corpus](#)
- Read or download the transcriptions:
 - [User license for the transcriptions](#)

Metadatamenyer i Norchron (i)

Hva slags metadata er åpenbart relevante i Norchron?

- ▶ **Tekst-ID** for hver tekst
- ▶ **Datering** av tekstene – men ikke ukomplisert:
 - ▶ Usikkerhet rundt datering
 - ▶ Vi daterer et gitt manuskript/diplom osv. – men språkformen kan reflektere eldre forelegg
- ▶ **Sted**
 - ▶ Presis informasjon ned til gårdsnavn tilgjengelig for mange diplomer (jf. Gammeltoft i morgen)
 - ▶ Ikke like presist i andre tilfeller
- ▶ **Sjanger/teksttype**
 - ▶ Men hvordan defineres de best?
 - ▶ Og hvor detaljert? F.eks. ulike typer diplomer?

Metadatamenyer i Norchron (ii)

Hva slags metadata kan ellers (muligens) være interessante?

- ▶ Er teksten **oversatt** til norsk fra et annet språk?
- ▶ **“Originalt”** dokument (f.eks. de fleste diplomer) eller skrevet av etter et forelegg?
 - ▶ Jf. datering – det er relevant, ikke minst i språkforskning, å vite om en tekst er en avskrift
- ▶ **Skriftsystem** – runer eller latinske bokstaver?
- ▶ **Materiale** – f.eks. stein, pergament, papir?

“Forfatter” er ikke en brukbar kategori – men diplomer kan ofte knyttes til én eller flere **utstedere**.

Metadatamenyer i Norchron (iii)

Hvordan skal informasjonen (evt.) formuleres i menyene?

- ▶ **Datering** – hvor detaljert?
 - ▶ Ulik informasjon tilgjengelig for ulike tekster
 - ▶ **Tidsintervaller** er trolig nyttig – hvor store intervaller?
- ▶ **“Original”** tekst?
 - ▶ En binær kategori (ja/nei)?
 - ▶ Skal vi (i tillegg eller istedenfor) prøve å angi datering for/spesifisere et evt. forelegg?
 - ▶ Interessant teksttype: **Vidisser** – avskrift av diplom med referanse til forelegg. Relasjon mlm. tekster i menyen?
- ▶ **Utsteder** (for diplomer)
 - ▶ Svært mange ulike personer
 - ▶ Kan/bør dette være søkbart? Evt. tittel/rang? (Mørck 1999, Blaxter 2017)

Avveining av ulike hensyn

- ▶ Norchron inkluderer mange ulike teksttyper – med ulike typer av (tilgjengelige) metadata
- ▶ Diplomer skiller seg ut med mye og detaljert informasjon
- ▶ Glossa gir i prinsippet mulighet til avgrense søk etter alle slags metadata – i alle tenkelige kombinasjoner
- ▶ **Men** veldig mange søkemuligheter kan gjøre grensesnittet uoversiktlig
- ▶ Metadatakategorier som kun er aktuelle for én teksttype kan virke forvirrende – på den annen side er det svært nyttig om man jobber med denne teksttypen!

“Deliverables”

Resultat av prosjektet (ifølge søknaden)

- ▶ Ca. 3000–5000 ord fra runeinnskriftene (nytt)
- ▶ Ca. 200 000 ord på gammalnorsk med full morfologisk og syntaktisk annotasjon fra eldre prosjekt, ytterligere 300 000 med morfologisk annotasjon
- ▶ Ca. 1 mill. nye ord
 - ▶ Alle tekster med minimum lemmatisering (inkl. ordklasse) og (automatiske) glosser
 - ▶ Ca. 200 000 ord fra mellomnorsk og tidleg moderne norsk med full morfologisk og syntaktisk annotasjon
- ▶ Søkbar i et brukervennlig grensesnitt
 - ▶ Språklig (grammatisk og leksikalsk)
 - ▶ Ulike typer metadata

Diskusjon

Spørsmål til diskusjon

Tekster

- ▶ Er det tekster vi absolutt bør ta med?
- ▶ Hvordan bør vi prioritere hvilke nye diplom vi tar med?

Metadata

- ▶ Hva slags metadata er nyttige for dere?
 - ▶ Av kategoriene som er nevnt ...
 - ▶ ... eller helt andre!
- ▶ Hvor detaljerte verdier i søkefeltene?
- ▶ Hva slags metadata bør prioriteres høyest, om vi må velge?