

Bornholmsk ordbog

Metaordbogen som sproglig ressource

Alex Speed Kjeldsen
Institut for Nordiske Studier og Sprogvidenskab
Københavns Universitet

Traditionelle ømåls- og østdanske dialekter
Københavns Universitet, fredag d. 18. juni 2021



Dagens præsentation

- 1 Om det bornholmske ordbogsprojekt
- 2 bornholmskordbog.ku.dk som sproglig ressource
- 3 Metaordbogskonceptet i et bredere perspektiv

1. Om det bornholmske ordbogsprojekt

Bornholmsk Ordbog – revitalisering af et kuldsejlet projekt

- Fase 1 (infrastrukturbygning, sprogportal og bredere formidling):
 - Få infrastrukturen på plads
 - Aktiver bornholmere og skab opmærksomhed omkring projektet
 - Lancer bornholmskordbog.ku.dk
 - → Få politisk opbakning, større bevillinger
- Fase 2: Genoptag egentlig ordbogsredigering
- Fase 3: Udarbejdelse af ny bornholmsk grammatik



Frivillige som ressource

Arbejdsopgaver	Antal frivillige
Scanning	3
OCR-behandling af tekster og glossarer	4
Indtastning af glossarer	4
Korrekturlæsning af glossarer	4
Annotation af glossarer	5
Indtastning og annotation af tekster	5
Korrekturlæsning af tekster	12
Indlæsning af tekster	3
Transskription af talt bornholmsk	1
Arkivarbejde (scanning, registrering)	4
Registrering af radioudsendelsers indhold	4



Hvad har vi nået?

En kombination af frivillige, studentermedhjælpere, en VA (på deltid i ét år) og undertegnede har bl.a.:

- Indtastet og annoteret ca. 20.000 ordbogssedler (BO)
- Indtastet/indscannet (OCR-behandlet) og annoteret 17 glossarer/ordlister. Pt. er ca. 30.000 artikler blevet lagt på nettet, yderligere knap 20.000 artikler vil blive lagt på nettet i løbet af sommeren/efteråret.
- OCR-behandlet og korrekturlæst Thomsens og Wimmers bornholmske grammatik (en student arbejder nu på at annotere, omstrukturere og linke grammatiken til den digitale ordbog)
- Scannet, OCR-behandlet og korrekturlæst en stor mængde bornholmsksprogede tekster



Hvad har vi nået? (fortsat)

- Udviklet prototyper på en række sprogteknologiske ressourcer til bornholmsk (Leon Derzycynski, ITU).
- Indholdsregistreret ca. 80 lydoptagelser fra bornholmskebilleder.brk.dk (interviews + radioudsendelser med fokus på bornholmsk).
- Transskriberet en række optagelser af dialekttalende bornholmere, indspillede fra 1938 og frem, i PRAAT (alignment mellem lyd og transskription (dialektortografi og rigsmålgengivelse) på sætningsniveau).
- Indlæst en række tekster.
- Journaliseret/registreret materiale fra diverse samlinger (KU og eksterne).



Hvad skal der arbejdes på i den kommende tid?

- Tilføje de resterende glossarer.
- Rette lemmalisten og tilføje rigsmålnorm.
- Integrere interaktiv grammatik.
- Publicere flere tekster.
- Tilføje lydspor til flere tekster.
- Foretage sproglig annotation af (udvalgte) tekster (lemmatisering og part-of-speech på basis af NLP-værktøjer) og linke teksternes ordformer til ordbogen (som kendt fra ebøger).
- Indføre flere tekstgengivelsesniveauer (standardiseret dialektortografi, rigsmålsgengivelse).
- Inkludere talesprogskorpusset.



Tekster, der er publicerede eller under forberedelse

Teksttyper	Antal ord	%-del udg.
Dialektkomedier	627.102	0 %
Romaner og længere fortællinger	232.000	100 %
Kortere fortællinger (sagn, eventyr, anekdoter)	307.024	23 %
Diverse tekster i dagblade og tidsskrifter	329.684	0 %
Digte	23.886	53 %
Andre tekster	2.150	0 %
total	1.521.846	6 %



2. bornholmskordbog.ku.dk som sproglig ressource

bornholmskordbog.ku.dk

- 1 Kort overordnet præsentation af bornholmskordbog.ku.dk
- 2 Nogle eksempler på brug af avanceret søgning
 - Søgning i ordliste (orddannelse): 1) [neutrale sb. på -ene](#); 2) [adj. på -urijn](#)
 - Fritekstsøgning i glossarer:
 - Rigsmålsord, fx [pindsvin](#)
 - Overordnede forhold, fx [nedsættende sprogbrug](#) og specifikke steder (fx [Gudhjem](#))
 - Søgning i tekster og citater
 - brugen af [farruer](#) vs. [færdier](#)
 - syntaktiske konstruktioner, fx [te + infinitiv](#) vs. [te å + infinitiv](#)



3. Metaordbogs-konceptet i et bredere perspektiv



En samlet nordisk metaordbog som grundlag for studier af sproglig variation i tid og rum

Man ville fx kunne studere:

- forskelle i ordforråd og orddannelse
- variation i form (lyd, bøjning, syntaks) i tid og rum
- variation i betydning i tid og rum
- specifikke ordsfærer og begrebsområder på tværs af sprog/dialekter i et synkront eller diakront perspektiv
- lån og sprogkontakt i forskellige regioner

