

Superordbog og salmesang: ordbogslinkning i praksis

Mette-Marie Møller Svendsen, Nicolai Hartvig Sørensen & Thomas Troelsgård

Music and Language in Danish Hymns from the Reformation is a cross-disciplinary project based at the Society for Danish Language and Literature. Part of the project is the compilation of a “super-dictionary” consisting of a total of 12 digitised and interlinked dictionaries and glossaries relevant to the language used in the hymnbooks from the Reformation. This article provides a brief outline of the practical steps taken in the process of linking our lexical data, such as how the lemma equivalents are chosen, a few examples of how the linking is carried out, as well as a presentation of the tools used in the process.

1. Om projektet

Projektet *Musik og sprog i reformationstidens danske salmesang* er et tværvindskabeligt projekt hjemmehørende på Det Danske Sprog- og Litteraturselskab (DSL). Som en del af projektet udvikles en ”superordbog” for renæssancens sprog, som bliver tilgængelig som en ordbog i DSL’s ordbogsplatform der i forvejen inkluderer ni moderne og historiske ordbøger for dansk. Denne ordbog vil blive benyttet som ressource af den internetportal som udvikles i projektet, og som skal præsentere salmebøgerne (se nedenfor). Superordbogen kommer til at rumme i alt 12 digitaliserede ordbøger og gloselister som er relevante i forhold til sproget i reformationstidens salmebøger og generelt for perioden. Ordbogsværkerne omfatter dansk-dansk, dansk-latin, latin-dansk og latin-dansk-græsk. Projektet er støttet af Carlsbergfondet og Velux Fonden og forventes afsluttet i 2021.

Superordbogen udnytter at DSL er i gang med at linke alle Selskabets ordbøger sammen på lemmaniveau i en overordnet ”metaordbog” og alle superordbogens links udgør i praksis en delmængde af denne metaordbog.

På samme måde som den norske metaordbog giver et overblik over samtlige opslag på et ord fra en række forskellige materialer (Ore 2001), så giver superordbogen også mulighed for at skabe et indblik i sprogets udvikling ved hjælp af en mængde danske resurser. I øjeblikket er der stort internationalt fokus på de muligheder som linkning af forskellige leksikalske resurser tilbyder, både for leksikografer og brugere, som det fx kan ses hos Hudeček & Mihaljević (2019:74).

Internetportalen danner udgangspunkt for tre tværfaglige forskningsprojekter:

- Kontinuitet og nybrud i reformationstidens salmesang: I dette projekt fokuseres der på forholdet mellem middelalderligt og nyt melodistof, mellem latinske og danske tekster og mellem det gammeldanske ordforråd ift. nyskabelserne i de lutherske salmer.
- Varians i melodier og tekster: I dette projekt undersøges en eventuel korrelation mellem musikalske og litterære variationer, der kan tolkes som udtryk for forskellige opførelsespraksisser.
- Spredning af kirkens melodier og tekster: I dette projekt undersøges spredningen til den verdslige sang, specifikt genbrug af salmemelodier og salmernes indflydelse på visernes ordforråd og fraseologi.

Den avancerede internetportal kommer til at bidrage med nye værktøjer og materialer til forskningsprojekterne, og de tre tværfaglige projekter forventes at blive påbegyndt i 2020. Superordbogen bliver tilgængelig som en selvstændig ordbogsdel på portalen, hvor man kan slå ord op samlet i de relevante ordbøger. Desuden bliver den integreret i portalens tekstpræsentation, så man kan slå ord op direkte fra centrale tekster.

Tabel 1 viser de ordbøger der direkte eller indirekte inddrages ved opbygningen af superordbogen. I denne artikel gennemgås arbejdet med linkningen i praksis.

Udgiver	Titel	Årstal	Omfang
Christiern Pedersen	<i>Vocabularium ad usum dacorum</i>	1510	13.000
Henrik Smith	<i>Hortulus synonymorum</i>	1520	1.000
Jon Tursen	<i>Vocabularius rerum</i>	1561	8.000
Henrik Smith	<i>Libellus Vocum Latinarum</i>	1563	6.200
Poul Nielsen Hingelberg	<i>Vocabulorum variorum expositio</i>	1576	3.500
Mads Pors	<i>De nomenclaturis Romanis</i>	1594	3.200
Poul Jensen Colding	<i>Dictionarium Herlovianum</i>	1626	30.000
Otto Kalkar	<i>Ordbog til det ældre danske sprog (2. udg. 1976)</i>	1881-1918	76.000
Jensen & Goldschmidt	<i>Latinsk-Dansk Ordbog (2. udg. 1920)</i>	1886	19.000
DSL	<i>Ordbog over det danske Sprog (ODS)</i>	1918-56	181.000
Kristian Sandfeld	Ordbog til <i>Hans Mogensens Oversættelse af Philippe de Commines Memoirer</i>	1919	5.500
Kristian Sandfeld	Ordbog til <i>En Ræffue Bog, Herman Weigeres Oversættelse af Reinke de Vos</i>	1923	6.100
Marius Kristensen	Ordbog til <i>Danske Viser fra Adelsvisebøger og Flyveblade 1530-1630</i>	1931	7.100
DSL	<i>Supplement til ODS (ODS-S)</i>	1992-2005	40.000
D. Duncker m.fl. (Københavns Univ.)	Ordlister udarbejdet i forb. med projektet <i>Dansk Folkevisekultur 1550-1700</i>	2001	4.500
DSL	<i>Den Danske Ordbog (DDO)</i>	2009-	100.000
DSL	<i>Gammeldansk Ordbog (GO)</i>	2015-	30.000
<i>Grøn baggrund = Ordbøger med dansk opslagsord</i>			
<i>Blå baggrund = Ordbøger med latinsk opslagsord</i>			
<i>Grå baggrund = Ordbøger udenfor projektet, der fungerer som "rygrad" for linkningen.</i>			

Tabel 1: Oversigt over de forskellige ordbøger der indgår i linkningen.

2. Ordbogslinkning: metoder og værktøjer

Artiklerne i de projektrelevante ordbøger linkes på lemmaniveau i superordbogen. Udgangspunktet for, eller rygraden i, linkningen er Den Danske Ordbog (DDO) og Ordbog over det danske Sprog (ODS), samt i mindre omfang supplementet til Ordbog over det danske Sprog (ODS-S) og Gammeldansk Ordbog (GO). For vokabularerne med latinske opslagsord opbygger vi forsøgsvis en mindre, latinsk superordbog, som en parallel til den danske, med Jensen & Goldschmidt fra 1920 (latin-dansk, 1. udgave udkom i 1886) som udgangspunkt.

Etymologi er det primære kriterium, og artikler med fælles etymologi knyttes sammen, også selvom der er stor variation i opslagsformerne; således samles fx *blugsel*, *blussel*, *blyghsel* og *blysel* alle i superartiklen *blusel*. Tilsvarende placeres *kamp* ('kampesten') og *kamp* ('krig, tvekamp') i hver sin superartikel.

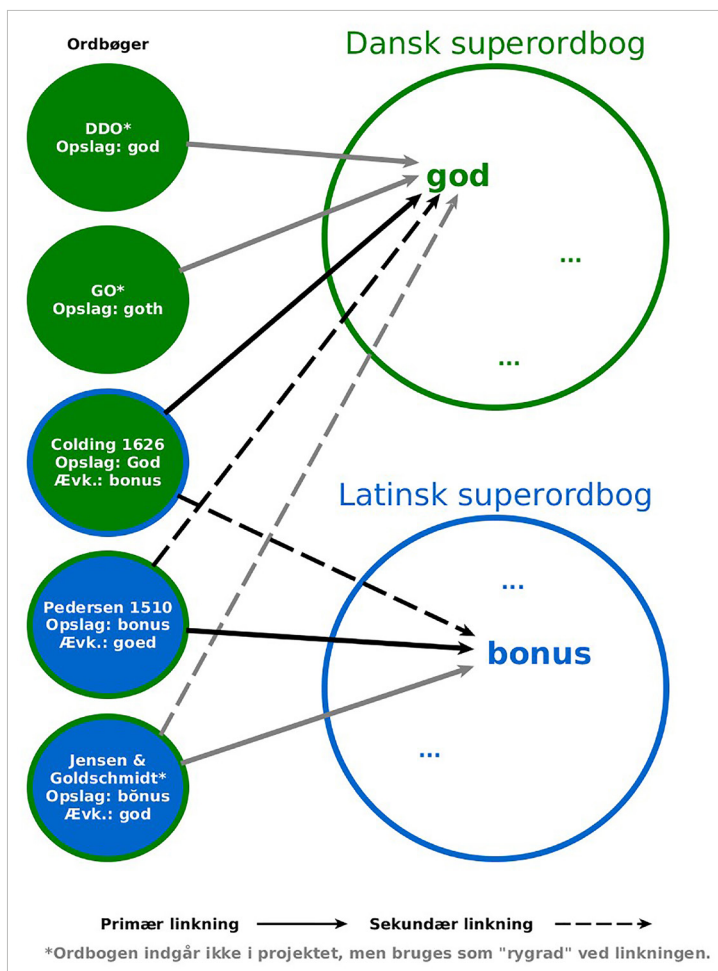
Hvis tiden tillader det, vil vi desuden forsøge at foretage en sekundær linkning, hvor artikler fra de bilingvale ordbøger via ækvivalenterne linkes til matchende artikler i målsproget.

Figur 1 viser et eksempel på linkning i form af lemmeet 'god' for hhv. dansk og latin. Her kan det ses hvordan DDO og GO fungerer som rygrad for linkningen, idet de udgør den knude som opslaget i *Dictionarium Herlovianum* (Col-ding 1626) linkes til.

Ved arbejdet med at etablere links til brug for superordbogen benyttes tre former for linkning af de leksikalske data: fuldautomatisk, halvautomatisk og manuel linkning. En redegørelse af de forskellige linkningsformer følger her.

2.1. Fuldautomatisk linkning

Artikler hvor der er match på form og ordklasse og hvor der ikke foreligger homografi, har vi valgt at linke fuldautomatisk med de potentielle fejl det kan medføre. Ord med datering i nyere tid fjernes automatisk i denne proces, så vi undgår at fx det gammeldanske substantiv *biker* 'bæger', linkes til DDO's *biker* 'motorcyklist'.



Figur 1: Figuren viser linkningsmuligheder for lemmaet 'god'.

2.2. Halvautomatisk linkning

For artikler hvor der ved inddragelse af flere kriterier end den rene opslagsform automatisk kan lokaliseres et eller flere mulige mål for linkningen, gennemføres en halvautomatisk linkning. Ved hjælp af et specialudviklet værktøj, navngivet *Konnektor*, kan redaktøren manuelt vælge og linke den eller de relevante målartikler. Kriterierne der bruges ved lokalisering af mulige målkandidater, omfatter:

- match på opslagsform og ordklasse (fx homografer)
- match på variantform der optræder i ordbogsartiklen

- match på genereret variantform, fx aa/å, x/ks, æ/e, ju/iu/y
- match på artikelindholdet generelt, baseret på strenglighed mellem tekstekstrakter fra kilde- og målartikel.

Artikler der skønnes ikke at have nogen matchende målartikel, gives et særligt mærke under linkningen i linkningsværktøjet, hvorefter de lægges ind som nye selvstændige lemmer, så vidt muligt normaliseret til moderne dansk for at lette fremtidig linkning. Tilsvarende sættes et særligt mærke ved problematiske artikler, hvilket typisk drejer sig om artikler hvor redaktøren ikke har et sikkert bud på linkningen. Endelig er der i værktøjet mulighed for manuelt at slå op blandt målartiklerne, og ad den vej linke manuelt til en målartikel der ikke er blandt de automatisk fundne kandidater.

2.3. Om Konnektor

Den halvautomatiske del af linkningen udføres ved hjælp af det specialudviklede værktøj Konnektor. Programmet blev oprindeligt udviklet til lemmatisering af ord i en række TEI-opmærkede middelaldertekster; til hvert ord blev vist en række mulige lemmakandidater, fundet ved at matche formen mod diverse ordlister vedr. perioden under prioriteret inddragelse af bøjningsparadigmer, lydregler, strenglighed m.m. Siden har værktøjet fået flere funktionaliteter, og det er blevet udvidet til at kunne arbejde på forskellige typer af input, så længe et simpelt XML-format overholdes. Således kan alle opgaver der består i at vælge 0, 1 eller flere mål til et givet udgangspunkt udføres i Konnektor, og programmet bruges i dag i flere DSL-projekter, både af sproglig, historisk og litterær art.

Tre eksempler på typiske situationer der opstår ved halvautomatisk linkning ved brug af Konnektor følger her.

- **Ét match fundet**

Kildeartiklen i Kalkars Ordbog er *uge*, sb.: ”samme som islandsk vika”. Konnektor finder match på form og ordklasse, og kun én mulig målartikel, nemlig *uge*, sb. Linkningen gennemføres ved at acceptere mulighed 1:

- 1) *uge*, sb. /DDO: 1. periode som strækker sig over syv på hinanden følgende [dage].

- **Flere matches fundet**

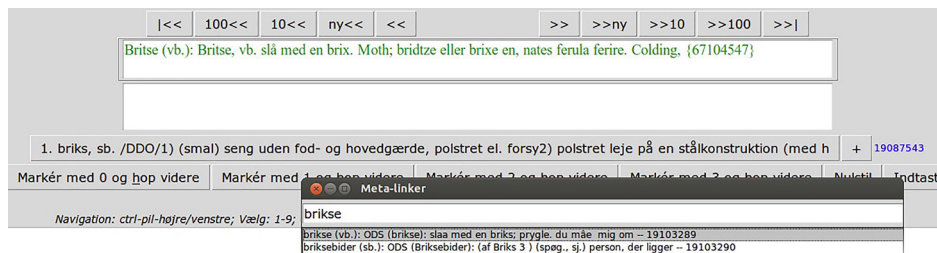
Kildeartiklen i Kalkars Ordbog er *skræppe*, sb. med betydningen 'taske'. Der er match på både form og ordklasse ved de første tre bud fra Konnektor, her findes dog flere mulige målar artikler:

- 1) *skræppe*, sb. /ODS: person, der skræpper op, taler meget og er støjende
- 2) *skræppe*, sb. /DDO: op til 1 m høj (flerårig) ukrudtsplante
- 3) *skræppe*, sb. /ODS: vadsæk, ransel olgn. ell (nu især) madkasse
- 4) *skræbe*, vb. om mennesker: græde (stærkt og vedholdende)
- 5) *skræppe*, vb. /DDO: give høje, skingre og gennemtrængende skrig

Mulighed 3 vælges, da kildeartiklens "taske" svarer til 'vadsæk, ransel' i målar-tiklen.

- **Ingen matches fundet**

Kildeartiklen i Kalkars Ordbog er *britse*, vb.: 'slå med en brix', se figur 2. Der foreslås ingen korrekt målarartikel i Konnektor, som kun foreslår *briks*, sb: '(smal) seng uden fod- og hovedgærde'. Det mistænkes at varianten *brikse* kan findes; *brikse* findes via et manuelt opslag i Konnektor, og der linkes til den korrekte artikel.



Figur 2: Halvautomatisk linkning af lemmaet *britse*, vb. i linkningsværktøjet Konnektor.

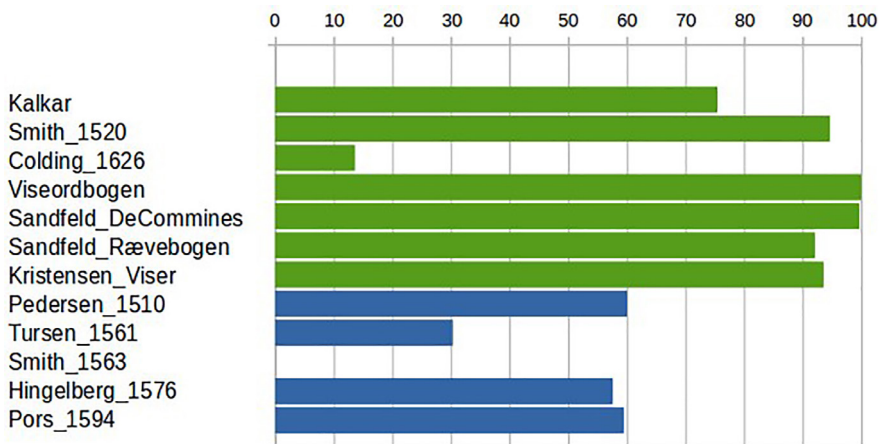
2.4. Manuel linkning

Artikler der ikke kan matches ved automatisk eller halvautomatisk linkning, samt artikler for hvilke der ikke er fundet nogle forslag til match, udskydes til manuel linkning. Denne linkning foretages som en senere "anden runde" af en kollega med speciale i 1500-tallets danske sprog. I praksis foregår linkningen som en fornyet kørsel i Konnektor, med den forskel at de fleste af disse problematiske

artikler først analyseres, fx ved kontrolopslag i andre værker, og derefter enten linkes manuelt til en eksisterende målartikel, eller markeres som nyt lemma. Et eksempel kunne være *tog* fra Kalkars Ordbog, i den artikel der begynder med: ”Tog, no.; i (på) t. (svf på tok) o: i blinde, løst hen”. Her bliver redaktøren præsenteret for ti forslag til match, hvoraf ingen giver mening, og frem for at bruge tid på analyse og kontrolopslag, sendes artiklen videre til en senere manuel linkningsrunde.

3. Status på projektet

Linkningen er i gang for de fleste ordbøger, og figur 3 viser status for linkningen i oktober 2019. I en del af de ældre ordbøger er lemmatisering og artikelstruktur inkonsekvent, og en større del af artiklerne har derfor ikke kunnet linkes under de hel- og halvautomatiske linkningsrunder.



Figur 3: Oversigt over status i procent for linkningen på projektet i oktober 2019.

4. Perspektiver

I og med at flere større ordbogsressurser – herunder DDO og ODS – indgår i den overordnede danske metaordbog, er grunden lagt for en bedre udnyttelse af ordbogsdata i DSL’s onlineresurser. Hvor opslag på tværs af ressourcer hidtil primært har været helt strengbaseret, kan nu også de linkede artikler (med afvigende opslagsformer) inddrages. Linkningen af ordbogsressurserne giver desuden mulighed for at levere nye typer af dataudtræk til brug i sprogteknologiske projekter,

hvor data fra forskellige resurser kombineres. Dette er allerede aktuelt, idet DSL deltager i det fælleseuropæiske projekt ELEXIS, bl.a. med leverance af data til forsøg med automatisk matching af betydningsbeskrivelser på tværs af sprog og perioder.

For tiden kan der kun søges enkeltvis i DSL's ordbogsressurser eller via ordbogsportaler som fx sproget.dk. En oplagt mulighed for videre udnyttelse af linkningen i fremtiden vil være opbygningen af en ny, samlet og integreret ordbogsressource for alle DSL's ordbøger. Ligesom den nye superordbog kan hjælpe brugere med interesse i reformationstidens sprog, vil lignende ordbøger med fokus på andre perioder kunne være til gavn for alle der arbejder med det ældre danske sprog.

Litteratur

Ordbøger m.m.

Colding, Poul J. (1626): *Dictionarium Herlovianum*. København: Salomon Sartor.

DDO = *Den Danske Ordbog*. Onlineversion: <www.ordnet.dk/ddo> (oktober 2019).

Duncker, Dorthe (1999): *Visernes vej: Sammenhæng mellem visebøger? I: Flemming Lundgreen-Nielsen & Hanne Ruus (red.): Svøbt i mår: dansk folkevisekultur 1550-1700*. København: C.A. Reitzel.

ELEXIS: European lexicographic infrastructure <www.elex.is>.

GO = *Gammeldansk Ordbog*. Onlineversion: <www.gammeldanskordbog.dk> (oktober 2019).

Hingelberg, Poul Nielsen (1576): *Vocabulorum variorum expositio*. København: Mads Vingaard.

Jensen, J. Th. & M. J. Goldschmidt (1920): *Latinsk-Dansk Ordbog, 2. udgave*. Onlineversion: <www.latinskordbog.dk> (oktober 2019).

Kalkar, Otto (1976): *Ordbog til det ældre danske sprog (1300-1700), 2. udgave, bind I-VI*. København: Universitets-Jubilæets danske Samfund, Akademisk Forlag. Onlineversion: <www.kalkarsordbog.dk> (oktober 2019).

Kristensen, Marius (1931/1978-79): *Ordbog til Danske Viser fra Adelsvisebøger og Flyveblade 1530-1630*. København: DSL, Gyldendal/C.A. Reitzel.

ODS = *Ordbog over det danske Sprog*. Onlineversion: <www.ordnet.dk/ods> (oktober 2019).

ODS-S = *Supplement til ODS*. Onlineversion: <www.ordnet.dk/ods> (oktober 2019).

Pedersen, Christiern (1510): *Vocabularium ad usum dacorum*. Paris: anon.

- Pors, Mads (1594): *De nomenclaturis Romanis*. Frankfurt a. M.: Joh. Feyerabend.
- Sandfeld, Kristian (1919): *Ordbog til Hans Mogensens Oversættelse af Philippe de Commynes Memoirer*. København: DSL, Gyldendal.
- Sandfeld, Kristian (1923): *Ordbog til En Ræffue Bog, Herman Weigeres Oversættelse af Reinke de Vos*. København: DSL, Gyldendal.
- Smith, Henrik (1520): *Hortulus synonymorum*. København: Melchior Blumme.
- Smith, Henrik (1563): *Libellus Vocum Latinarum*. København: Hans Zimmermann.
- Tursen, Jon (1561): *Vocabularius rerum*. København: Lorenz Benedicht.

Anden litteratur

- Grønvik, Oddrun & Christian-Emil Smith Ore (2018): Bokmål og nynorsk samindeksert – Metaordboka som verktøy for jamføring og utforskning av ordtilfang. I: *Nordiske studier i leksikografi* 14. Reykjavik: Nordisk forening for leksikografi, s. 87-95.
- Hudeček, Lana & Milica Mihaljević (2019): Croatian Web Dictionary – Mrežnik – Linking with Other Language Resources. I: Kosem, I., T. Zingano Kuhn, M. Correia, J. P. Ferreria, M. Jansen, I. Pereira, J. Kallas, M. Jakubiček, S. Krek & C. Tiberius (red.): *Electronic lexicography in the 21st century*. Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o., s. 72-98.
- Ore, Christian-Emil (2001): Metaordboken – et rammeværk for Norsk Ordbok. I: *Nordiske studier i leksikografi* 5. Gøteborg: Nordisk forening for leksikografi, s. 250-270.

Mette-Marie Møller Svendsen
assisterende redaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
mettemariesvendsen@gmail.com

Nicolai Hartvig Sørensen
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
nhs@dsl.dk

Thomas Troelsgård
seniorredaktør
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
tt@dsl.dk