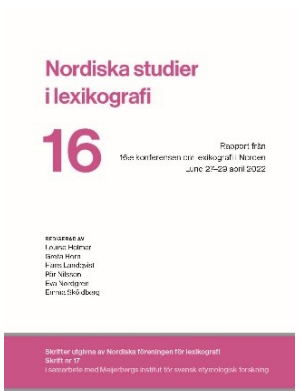


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Digitalisering af Jysk Ordbogs seddelsamling	
Forfatter:	Inger Schoonderbeek Hansen, Mette-Marie Møller Svendsen & Kristoffer Friis Bøegh	
Kilde:	Nordiska studier i lexicografi 16, 2023, s. 99–112	
URL:	https://tidsskrift.dk/nsil/issue/archive	

© Respektive författare, Nordiska föreningen för lexicografi och Meijerbergs institut för svensk etymologisk forskning, 2023

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavspersonen til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Digitalisering af Jysk Ordbogs seddelsamling

Inger Schoonderbeek Hansen, Mette-Marie Møller Svendsen &
Kristoffer Friis Bøegh

The historical, lexicographical documentation project, the Dictionary of the Jutlandic Dialects, *Jysk Ordbog*, is being edited on the basis of an extensive, unique, and *physical* collection. This collection comprises over three million paper slips containing excerpts of written and spoken linguistic data, reflecting the traditional culture and social conditions of the rural population of Jutland, Denmark, around 1850–1920. This article presents and discusses a pilot project in which almost 56,000 individual paper slips were scanned and digitized. The experiences of the project provide insights into how *Jysk Ordbog* and its different types of end-users will be able to benefit from digitization, and, in a more general documentary, lexicographical perspective, how digitization supports the project's overall goal of preserving cultural heritage.

NØGLEORD: dialektleksikografi, digitalisering, seddelsamling, Jysk Ordbog, kulturarv

1. Indledning

Den historiske dokumentationsordbog *Jysk Ordbog* (herefter JO) redigeres på baggrund af et omfattende, unikt og *fysisk* kildemateriale. Det består af ca. 3,1 millioner excerper af forskellige typer skrift- og talesproglige kilder der giver indblik i den jyske landbefolknings kultur- og samfundsforhold i perioden ca. 1850–1920. JO's kildemateriale er imidlertid ikke digitalt og matcher derfor ikke nutidens leksikografi (og sprogvidenskab generelt) hvor der i udgangspunkt tages afsæt i *digitale* tekster og korpora (jf. fx Svensén 2004, Atkins & Rundell 2008, van Keymeulen 2018). Overordnet set kan nutidens syn på tekst opsummeres således:

[T]exts in the 21st century are digitized. We digitize cultural heritage and literary classics, and new texts within mass communication, education, social media, journalism and literature are by default produced in digital formats (Bertelsen & Tannert 2021:i).

I nærværende artikel præsenteres arbejdet med et pilotprojekt hvor ca. 56.000 sedler fra JO's seddelsamling blev scannet og bearbejdet digitalt.¹ Overordnet set kommer artiklen ind på hvad digitalisering af seddelsamlingen betyder, og hvem der kan have gavn af den. De opnåede erfaringer med pilotprojektet giver et indblik i hvordan JO og dens brugere fremadrettet vil kunne drage nytte af digitaliseringen, og dermed hvordan denne understøtter projektets mål med bevaring af kulturarv og indgang til lingvistiske studier. Mens Hansen (2020) gav en statusrapport for JO med afsæt i spørgsmål om *hvem* ordbogen redigeres af og til, samt *hvorfor* det traditionelle videnskabelige ordbogsarbejde fortsat har og skaber værdi, supplerer nærværende artikel med afsæt i digitaliseringsprojektet med *hvordan* brugerne vil kunne tilgå det omfattende materiale.

I afsnit 2 skitseres JO's heterogene kildemateriale, og i afsnit 3 redegøres for pilotprojektets resultater. Det gælder såvel arbejdet med at ekstrahere metadata fra det scannede materiale, som hvilke konkrete udfordringer scanning af de redigerede og uredigerede sedler gav. I afsnit 4 drøftes digitaliseringens formål, dvs. hvilke nye forskningsperspektiver og andre anvendelsesmuligheder en digitaliseret seddelsamling vil kunne medføre. Desuden gives nogle eksempler på både simple og avancerede søgninger, der vil kunne fortages af diverse brugertyper, og der vil kunne give større og mere fleksible muligheder for at undersøge opslagsord af specifik sproghistorisk og/eller kulturhistorisk værdi. I afsnit 5 kommer vi med nogle afsluttende bemærkninger samt perspektiverer til fremtidige tiltag.

2. Jysk Ordbogs kildemateriale

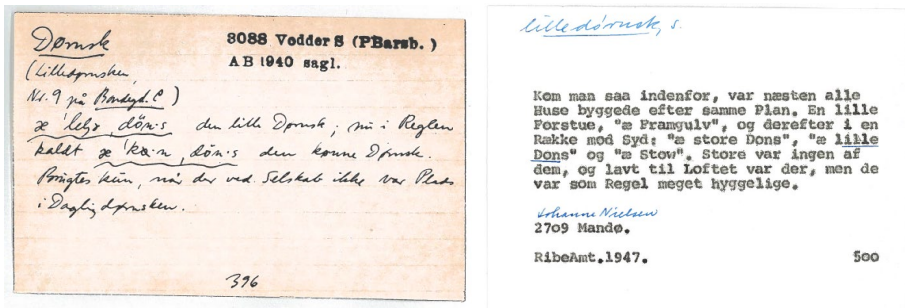
Det omfattende, empiriske indsamlingsarbejde til JO påbegyndes i 1932 af professor i dansk sprog ved Aarhus Universitet Peter Skautrup (1896–1982). JO er derfor et empirisk-videnskabeligt funderet projekt der afspejler dansk kulturarv sådan som den sprogligt er indlejret hos den jyske landbefolkning i perioden ca. 1850–1920, dvs. indtil afdialektaliseringen i Danmark tog fart som følge af bl.a. landbrugets mekanisering samt samfundets industrialisering og urbanisering (jf. fx Bøegh et al. 2023). De første prøvehæfter af ordbogen udkom 1970–1979, og siden 2000 er JO udkommet

¹ Stor tak til Kristoffer Nielbo og Peter Bjerregaard Vahlstrup, Aarhus Universitet samt studentermedhjælper Agnes Boel Nielsen for deres bidrag til pilotprojektet.

som online ordbog på www.jyskordbog.dk. Siden 2017 anvendes redaktionsværktøjet iLEX, og i skrivende stund (marts 2023) er 53 % af JO færdigredigeret og publiceret på <https://jysk.au.dk/jyskordbog>. Selvom JO publiceres online, findes selve kildematerialet fortsat i fysisk form (jf. Arboe 2003, Sørensen 2019, Hansen 2020, jysk.au.dk/jyskordbog).

JO's kildemateriale kan inddeles i tre hovedgrupper:

1) JO's seddelsamling eller centralkartotek (jf. JO, Ordbogens kilder) indeholder ca. 3,1 millioner papirsedler med excerpter fra bl.a. skønlitteratur på jysk dialekt, jysk typografi og lokalhistorie, dialektoptegnelser ved filologer mv. De enkelte papirsedler indeholder ord, udtalegengivelser, ordforbindelser og/eller korte citater af autentisk jysk dialekt. Sedlerne udgør et i høj grad heterogent materiale; nogle sedler er skrevet i hånden, andre på skrivemaskine, mens andre igen er indtastet på computer og derefter udprintet. De afspejler dermed mange årtiers dialektindsamling, forskellige faglige traditioner, varierende grader af systematik og akribi etc.; sedlernes indhold er ligeledes særdeles heterogent (se Figur 1).



FIGUR 1. To eksempler fra seddelsamlingen til opslagsordet *lille-dørns*: dialektologen Anders Bjerrums seddel fra Vodder (til venstre), og meddeler Johanne Nielsens beretning ”Livet paa Mandø” i Fra Ribe Amt, 1947 (til højre).

Sedlernes heterogenitet taget i betragtning må man regne med ”mange typer sproglige data på en ordbogsseddel”, som Henrik Hovmark (2011:301) fra *Ømålsordbogen*² formulerer det, hvilket fremhæver vigtigheden af at have et kritisk blik på sedlernes repræsentativitet. Sedlerne gengiver derfor ”et udsnit af virkeligheden som skal vurderes kritisk som kildegrund-

2 Jf. Gudiksen (2021) for en nylig oversigt over *Ømålsordbogen*s arbejde.

lag” (2011:305). De kan således aldrig anses som 100 % repræsentative, men de giver et kvalificeret bud på hvordan landbefolkningen talte.

2) JO’s spørgelistesamling består af spørgelister udsendt 1949–2014, besvaret af mellem 300 og 1.000 meddelere født i perioden ca. 1880–1920. Denne kildetype bidrager primært til beskrivelsen af den geografiske udbredelse af ords betydning og/eller udtale, i JO illustreret ved hjælp af udbredelseskort. Spørgelistesamlingen er ligeledes exciperet og skrevet på sedler der indgår i seddelsamlingen. Endelig er spørgelistesamlingen delvist digitaliseret og kan tilgås via en lokal database på Peter Skautrup Centret (herefter PSC) i Århus.

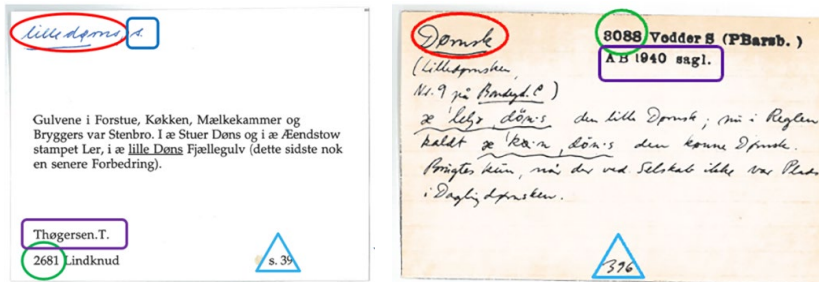
3) JO’s båndsamling består primært af optagelser fra hele landet indsamlet i perioden 1971–1976 og med dialekttalende født omkring 1900. En stor del af den blev indsamlet på basis af flere samarbejdsprojekter med Center for Dialektforskning, Københavns Universitet. Båndsamlingen er ligesom spørgelistesamlingen digitaliseret; en digital kopi af den opbevares af Det Kongelige Bibliotek, Aarhus (jf. Goldshtein & Puggaard 2019). Kun en mindre del af båndsamlingen er exciperet, dvs. transskriberet og udskrevet på sedler der indgår i JO’s seddelsamling.

3. Pilotprojektet

Pilotprojektet med digitaliseringen af JO’s seddelsamling er et samarbejde mellem PSC og Center for Humanities Computing Aarhus (herefter CHCAA). For PSC tjener digitalisering af den omfattende fysiske, og sårbare, seddelsamling et bevaringsformål, mens en digitaliseret seddelsamling i højere grad vil kunne stilles til rådighed for både fagfolk og alment interesserede end den fysiske samling kan i dag. Den vil yderligere kunne give bedre online søgeværktøjer for ordbogsredaktørerne, bl.a. vil man kunne lave krydssøgninger på metadata, fx ordklasse og kilder. CHCAA ville med pilotprojektet gerne teste kapacitet og brugbarhed af en specialscanner, samt tilpasse software de har udviklet til scanningsarbejdet. Dette ville give CHCAA gode erfaringer i forhold til fremtidige scanningsprojekter som vil kunne tilbydes til andre fysiske samlinger i bog- eller seddelform.

3.1. Metadata

JO's seddelmateriale er som nævnt heterogent. På en eksemplarisk seddel indgår et opslagsord, ordklasse, et topografisk nummer, navn på kilden, et excerpt og evt. et sidetal hvis excerptet er fra en trykt kilde. Alle sedler besidder dog langt fra altid alle nævnte metadata, og placeringen af de enkelte metadata på sedlen varierer meget (se Figur 2). Desuden er mange af sedlerne håndskrevne. Dette giver nogle udfordringer i forhold til en automatiseret afkodning af data fra sedlerne. Et af digitaliseringens mål er at *computer vision* anvendes til at aflæse udvalgte metadata fra seddelsamlingens enkelte sedler. I de fleste tilfælde vil computeren kunne tilgå JO's eksisterende lukkede inventarlistes, fx redaktionens lister over kilder, topografiske numre mv.



FIGUR 2. To repræsentative eksempler fra JO's seddelsamling til opslagsordet *lille-dørns*. Den røde cirkel viser metadata opslagsord, den blå firkant ordklasse, den grønne cirkel topografisk nummer, den lille firkant kilde, og den lyseblå trekant kildens sidetal.

3.2. Scanningsresultater

Til pilotprojektet benyttede vi en seddelscanner med en kapacitet på 250 sedler i minuttet. Dertil brugte vi software specialfremstillet af CHCAA til at sortere det scannede materiale og indtaste udvalgte metadata. En studentermedhjælper stod for håndteringen af seddelmaterialet og scanningen, fx skulle hun indtaste udvalgte metadata. Undervejs i forløbet førte hun logbog over hvilke skuffer og typer materiale hun scannede, hvor lang tid det tog, og hvilke problemer hun stødte på i løbet af processen. Til pilotprojektet stræbte vi efter at vælge et så forskelligartet materiale

som muligt da vi ville undersøge hvilke faktorer der kunne påvirke scanningsprocessen. Fx valgte vi nogle skuffer med få sedler per opslagsord, dvs. mange, men korte opslagsord som studentermedhjælperen skulle behandle, og sammenlignede disse erfaringer med skuffer med blot et enkelt eller få opslagsord, dvs. med flere tusinde sedler per opslagsord. Dertil lavede vi også stikprøver i forhold til redigeret og uredigeret seddelmateriale. Det heterogene seddelmateriale gav umiddelbart en del udfordringer. Her følger nogle eksempler:

- 1) Seddelmateriale med trykt information på begge sider. Scanneren er indrettet til at kunne registrere tekst på begge sider af en seddel, men da papirkvaliteten på flere af sedlerne er helt tynd, ville det resultere i en del fejlscanninger og i værste fald i den dobbelte arbejdsmængde hvis samtlige sedler skulle scannes to gange. Løsningen blev at scanneren blev indstillet til kun at scanne én side hvorfor studentermedhjælperen skulle monitorere scanningsprocessen og holde øje med sedler med tryk på begge sider.
- 2) Seddelmateriale sammensat af to (eller flere) stykker papir. Scanning af sedler bestående af flere sammenklipsede eller -hæftede stykker papir kan resultere i papirstop, eller i værste fald iturevne sedler. En løsning på problemet var at studentermedhjælperen overførte data fra sedlen til ét stykke papir, fx ved at scanne sedlerne på en almindelig scanner og efterfølgende klippe det til.
- 3) Uredigeret seddelmateriale. Til trods for at seddelsamlingen er struktureret både alfabetisk og topografisk, kan uredigeret materiale mangle finsortering og derfor stå i forskellige skuffer; det gælder fx materiale til ordformerne *lodden* og *lådden* der begge hører til opslagsordet *loden*. Visse homonymer er heller ikke opdelt endnu hvorfor studentermedhjælperen skulle foretage flere beslutninger i scanningsprocessen.

JO's 3,1 millioner sedler er fordelt på 1.656 skuffer, dvs. ca. 1.900 sedler per skuffe, der varierer i størrelse. Studentermedhjælperen brugte 74 timer på projektet hvori hun scannede 25,5 skuffer med 55.897 sedler, fordelt på 2.557 opslagsord. Det gav os altså ca. 755 indscannede sedler i timen eller 1/3 skuffe i timen. Det betyder at vi i løbet af pilotprojektet fik scan-

net ca. 2 % af seddelsamlingen. En scanning af hele seddelsamlingen vil derfor kræve ca. 3.700 studentertimer. Selvom timeantallet er højt, og vi sandsynligvis vil møde nye problemstillinger ved en opskalering af projektet, så har vi løst mange udfordringer: Redaktionen har i fællesskab diskuteret og fastlagt tilgange til at håndtere seddelmateriale der giver ekstra udfordringer, og som derfor ikke kommer til at kræve en masse ressourcer fremover. CHCAA har justeret softwaren efterhånden som studentermedhjælpen er stødt på noget der yderligere kunne optimere indtastningsarbejdet. Arbejdsgangen er blevet optimeret undervejs i projektet, og gode vejledninger til processen er blevet udfærdiget, så arbejdet ved opskalering hurtigt kan igangsættes. Ved en opskalering vil der også blive arbejdet med automatisk aflæsning og fejltjek af metadata (jf. afsnit 3.1).

4. Hvilket formål har en digitaliseret seddelsamling?

De enkelte opslag i JO fungerer, som Asgerd Gudiksen påpeger i sin artikel om *Ømålsordbogens* samlinger som sproghistorisk kilde, som en ”indgang til [den bagvedliggende] samling” (Gudiksen 2021:25). Hver artikel baserer sig derfor på forskning der formidles både til den alment interesserede bruger og til fagfællen, lokalhistorikeren etc. (jf. Svendsen 2023). I nærværende afsnit fremhæves digitaliseringens potentiale i et forsknings- og brugerorienteret perspektiv (jf. desuden Hovmark 2023).

4.1. Fremtidig brugerflade

En digital seddelsamling vil på den ene side gavne JO's redaktionsarbejde, især hvis den kan udformes som et dynamisk værktøj hvor uredigerede sedler kan sorteres og opdateres (jf. Hovmark & Gudiksen 2018). På den anden side vil den også kunne bruges af kolleger, ikke mindst som grundlag for egne sproghistoriske og kulturhistoriske studier. Endelig vil den almindeligt interesserede bruger kunne tilgodeses med oplysninger om dansk talesprogs historie, struktur og kulturafspejling.

JO's brugere bør kunne få adgang til en brugerflade hvor de kan søge på forskellige typer data både som simple eller mere avancerede søgninger. Søgninger skal kunne trække på oplysninger i de to tilgængelige databaser: redigeringsplatformen iLEX og den fremtidige digitaliserede seddelsamling. Løsningen skal give mulighed for at vise seddelmaterialet, lige-

som brugerne på sigt via ordbogens kortmateriale skal kunne se et ords udbredelse, sådan som de er vant til i JO. Dermed vil det materiale redaktionen igennem årtier har indsamlet og bearbejdet leksikografisk, kunne nå ud til såvel en bredere offentlighed som det mere snævre lingvistiske fagmiljø. Nedenfor reflekteres over disse to brugertyper med afsæt i nogle eksempler.

4.2. Simple søgninger

Hvis en alment interesseret bruger ønsker at vide mere om betydningen og udbredelsen af et specifikt ord, kan vedkommende slå ordet op i JO. Ifølge JO betyder fx ordet *gammeløl* 'lagret, stærkt øl', og der oplyses yderligere: "som regel brygget i marts (el. tidligere) og anvendt i høst- og jule-tiden, stedvis også ved gilder; ofte tilsat sukker (el. kandis) og brændevin (el. rom)". Betydningsangivelsen følges op af fem citater, dvs. belæg som redaktøren har udvalgt fra seddelsamlingen (jf. også Gudiksen 2021:25). Hvis brugeren derudover kunne få adgang til en digitaliseret version af de relevante sedler, inkl. de øvrige eksempler på *gammeløl* som redaktionen af bl.a. formidlingsmæssige grunde har fravalgt fra selve ordbogsartiklen, ville vedkommende have mulighed for selv at gennemgå opslagsordets belæg. Brugeren kunne ønske at vide mere om hvordan man bryggede øl i det traditionelle jyske landbosamfund. Da artiklen *øl* endnu ikke er redigeret i JO (som er nået til *l-*), ville en søgning i JO's digitaliserede seddelsamling på opslagsordet *øl*, evt. inkl. ordklasseangivelse, give et relativt stort antal excerpter, fx:

gammeltøl, s

man da 'ha: 'tɔy, 'gaməl 'ɔl, de blɔw
 da slán on 'pɔt 'rom, 'ix, 'eyn de
 blɔw spu:nt 'eél / 'sá wa 'de jo 'hi: 'l'at
 'sá 'stow 'de 'da: te 'hɔst / 'de sku 'nɔ: s
 te? á blyw 'strap.

1382 Bedsted s.Hassingh. MS 217.6.1
 Medd.f.1881. Opt.1956.Udskr:Torsten Balle.

Gammeltøl, s

Den Dag, man fik ophøstet, "strøg
 man for Kaalen", og Madmor maatte da
 ud med Gammeltøl og Æbleskiver; om
 Aftenen holdt man et Gilde, "Opskør"
 med en bedre Nadver til Folkene,
 senere holdtes Høstgilder.

181

AarbVejle. 1943.

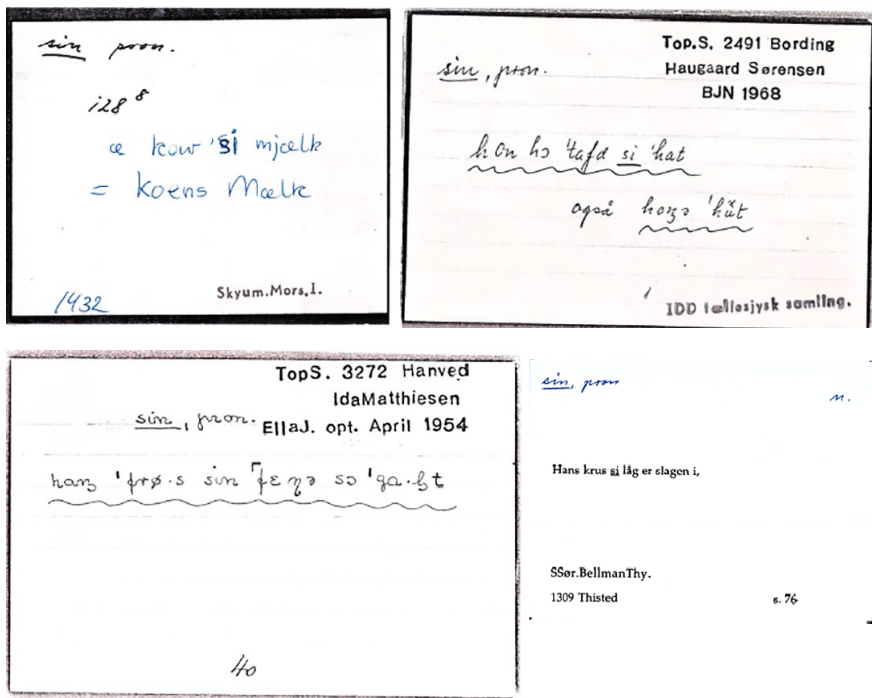
1. gammeltøl, s.
 Gammeløl - sørlig gotthorj-
 ning, der lagretes uist et
 helt Aar. Gammeløl næv en
 Feind i". Opmindelig vor det
 uist en rigtig Feind til at
 være Seckeren om med i
 Skænk for en Feind. Kan jeg
 har altid faaet den Opfat-
 tending 2224.

gammeltøl, s.
 Snaps ukendt. — an 'ol-hörel, et glas öl med en dram i.
 'gaməl'öl, gammeltöl, öl, som er brygget året i forvejen (og som
 særlig gemtes til hosttiden); 'gaməl'öl næ an 'pej' 'töl = ölholv.
 Skautrup, H. I.
 2481 174

FIGUR 3. Et udvalg af excerper under et forestillet søgeresultat af opslagsordet *gammel-øl*.

Sedlerne ville være topografisk ordnede, men som det fremgår ovenfor, kan excerperterne af forskellige årsager være svære at læse (jf. afsnit 2). Derfor bør sedlerne suppleres med oplysninger, linkmuligheder mv., så brugeren vil kunne få et indblik i fx kildernes alder, proveniens og pålidelighed.

Man kan ligeledes forestille sig en situation hvor en fagfælle, fx en lingvistkollega, gerne vil vide mere om et jysk grammatisk fænomen, fx brugen af de refleksive pronominer *hans*, *hendes* og *sin*. Lingvistens søgning i JO resulterer i oplysninger under hhv. opslagsordene *han* og *hun* hvor oplysninger om det refleksive *hans* hhv. *hendes* er bragt under artiklens allersidste betydningsangivelse. Her viser eksempler at *hans* og *hendes* i jysk kan referere til sætningens subjekt når det er hhv. *han* og *hun*. Dette mønster afviger fra rigsdansk hvor det refleksive pronomen *sin* bruges, hvilket nævnes i ordbogsartiklens afsnit om etymologi. Hvis lingvisten gerne vil vide mere om brugen af det endnu uredigerede *sin*, fx om det overhovedet bruges i jysk, så vil hun kunne søge på opslagsordet *sin* i den digitale seddelsamling og finde et relativt stort antal excerper fra samlingen, fx:



FIGUR 4. Forestillet søgeresultat for pronomenet *sin* – et udvalg af excerpter.³

Også dette søgeresultat bør suppleres med oplysninger der gør det muligt for lingvisten at forholde sig lige så kildekritisk til materialet som JO's redaktører.

4.3. Avancerede søgninger

Brugerne, uanset om de er ikke-lingvister eller lingvister, skal også kunne foretage avancerede søgninger i den digitaliserede seddelsamling. Til dette formål er det afgørende at kunne afgrænse søgninger ved hjælp af filtre. Umiddelbart vurderer vi følgende filtre som mest relevante: opslagsord, ordklasse, topografisk nummer og kilde. Senere kan der evt. tilføjes yder-

³ Bent Jul Nielsen lavede en større undersøgelse af pronomenet *sin* i jysk (Nielsen 1986). Resultaterne viste at *sin* kan optræde som reflektivt pronomen i jysk, men at det da som regel henviser til subjekter der ikke er *han* eller *hun*, jf. fx *e kow' stâ'r i si bo's* (optegnelse af Torsten Balle i Thorsted, Thy, ca. 1960) eller *den (dvs. vædderen) havde jo gjort sin tjeneste* – hvorfor den skulle slagtes (Anders Bjerrum, optegnelse fra Vodder, Vestsønderjylland, ca. 1930'erne).

ligere filtre som fx emne og semantisk kategori. Mens simple søgninger på metadata som opslagsord plus ordklasse allerede foreligger i pilotprojektet (jf. afsnit 3), kunne den almindelige bruger anvende et filter som topografisk nummer, dvs. en søgning på geografisk variation inden for det jyske område, for at skærpe sit spørgsmål til fx ”Brugte man betegnelsen *gammeløl* i Vendsyssel?” Brugeren finder oplysninger om hvad de specifikke topografiske numre refererer til i en lukket liste; listen indgår i en af JO’s redaktionsmanualer og skal blot gøres tilgængelig. Lingvisten kunne bruge filteret ’topografisk nummer’ til fx at undersøge om der findes forskel på brugen af *hans* og *sin* i hhv. vestjysk og østjysk, måske ud fra en hypotese om at østjysk i flere tilfælde strukturelt stemmer overens med rigsdansk, jf. fx artikelbrug: østjysk *hus-et* og vestjysk *æ hus*. Dette filter bør på sigt kobles sammen med det omfattende kortmateriale der er tilgængeligt i JO. Filteret kilde kan bl.a. muliggøre søgning på historisk variation i materialet. Hvis brugeren har fundet et ord i en specifik kilde, fx hos en dialektforfatter som Jeppe Aakjær, skal vedkommende have mulighed for udelukkende at kunne søge i den del af JO’s seddelsamling der består af excerpter fra Jeppe Aakjærs tekster. Filtret skal hjælpe lingvisten i sin søgning til fx at kunne datere et specifikt sprogligt fænomen. En liste over JO’s kilder fremgår allerede på JO’s hjemmeside, og brugeren kan finde oplysninger om kildens alder, hvad den beskriver etc.

Fremtidige søgemuligheder kunne fx også inkludere søgning efter emne, som pt. hverken eksisterer som tags for seddelsamlingen eller i JO som sådan, men som anvendes i visse andre ordbøger. Ydermere kunne man forestille sig søgninger efter bestemte semantiske kategorier; dette er pt. heller ikke muligt, men kunne opnås ved at gøre det muligt udelukkende at søge i opslagsordets betydningsafsnit. Dette ville umiddelbart være muligt i det færdigredigerede men ikke i det uredigerede seddelmateriale.

5. Afslutning

Digital humaniora og digitaliseret kulturarv er kommet for at blive, og de bidrager til nye, forskningsmæssige indsigter. Digitalisering af arkiv- og seddelsamlinger har gennem de seneste 25-30 år fundet sted på forskningsinstitutioner rundt om i Europa. I en nordisk kontekst var et af de nok tidligste projekter det norske Dokumentationsprojekt (Ore & Kristiansen 1998), mens et af de senere projekter er digitalisering af *Ømålsordbogens*

seddelsamling, der dog endnu mangler en digital infrastruktur i form af en brugerflade (jf. Hovmark & Gudiksen 2018). Formidlingsmæssigt vinder vores projekt ved at unikt og empirisk-videnskabeligt materiale om det gamle landbosamfund i Jylland kan nå ud til et langt større publikum med såvel faglig som almen interesse. Det både form- og indholdsmæssigt heterogene materiale fordrer dog vejledninger til de enkelte kilders optegnelser, korte emnebeskrivelser, så brugerne bedre kan forstå og bruge materialet. Pilotprojektet har givet en smagsprøve på hvad vi kan forvente ved en opskalering af pilotprojektet med digitalisering af JO's seddelsamling. En fremtidig målsætning er ydermere at koble de eksisterende og sammenlignelige ordbogsressourcer i såvel Danmark som i Norden (jf. også Svendsen 2023).

Litteratur

- Arboe, Torben 2003. Jysk Ordbog på internettet. *Nordiske Studier i Leksikografi* 6, 31–41.
- Atkins, B.T. Sue & Michael Rundell 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bertelsen, Ulf Dalvad & Morten Tannert 2021. Introduction: The Role of Computational Methods within the Humanities and Social Sciences. *Scandinavian Studies in Language* 12(1), i–iv.
- Bøegh, Kristoffer Friis, Peter Bakker, Inger Schoonderbeek Hansen & Carsten Levisen 2023. The Beginning of Quantitative Sociolinguistics in the Nineteenth Century: The Dane Anker Jensen (1878–1937) and his pioneering study “The Linguistic Situation in the Parish of Aaby, Aarhus County” (1898). *Historiographia Linguistica* 49(2/3), 336–354.
- Goldshstein, Yonatan & Rasmus Puggaard 2019. Overblik over danske dialektoptagelser. *Ord & Sag* 39, 18–28.
- Gudiksen, Asgerd 2021. Ømålsordbogens samlinger som sproghistorisk kilde. *Danske Talesprog* 21, 11–28.
- Hansen, Inger Schoonderbeek 2011. Livslange indsatser – om arbejdet hen imod Jysk Ordbog. I: Arboe, Torben & Inger Schoonderbeek Hansen (red.), *Jysk, ømål, rigsdansk mv. Studier i dansk sprog med sideblik til nordisk og tysk. Festskrift til Viggo Sørensen og Ove Rasmussen*. Århus: Peter Skautrup Centret for Jysk Dialektforskning, Nordisk Institut, Aarhus Universitet, 299–316.

- Hansen, Inger Schoonderbeek 2020. Jysk Ordbog – af hvem, til hvem og hvorfor? I: Sandström, Caroline, Ulla-Maija Forsberg, Charlotta af Hällström-Reijonen, Maria Lehtonen, & Klaas Ruppe (red.), *Nordiska studier i lexikografi* 15 (Skrifter utgivna av Nordiska föreningen för lexikografi. Skrift nr 16). Helsingfors, 135–143.
- Hovmark, Henrik 2011. Data og repræsentativitet i ordbogsarbejdet. I: Eaker, Birgit, Lennart Larsson & Anki Mattisson (red.), *Nordiska studier i lexikografi* 11. *Rapport från Konferens om lexikografi i Norden*. Lund, 296–308.
- Hovmark, Henrik 2023. Seddelsamlinger – historisk arkivmateriale eller levende resurse? I: Holmer, Louise et al. (red.), *Nordiska studier i lexikografi* 16. Lund & Göteborg: Nordiska föreningen för lexikografi, 165–176.
- Hovmark, Henrik & Asgerd Gudiksen 2018. Digitization of the Collection at Ømålsordbogen – the Dictionary of Danish Insular Dialects: Challenges and Opportunities. *CEUR Workshop Proceedings 2084*, 341–348.
- JO = *Jysk Ordbog*. 2000–. <www.jyskordbog.dk>. Tilgået februar 2023.
- Nielsen, Bent Jul 1986. Om pronominet *sin* i jysk. *Danske folkemaal* 28, 41–101.
- Ordbog over det danske sprog*. 1918–2005. <www.ordnet.dk/ods>. Tilgået februar 2023.
- Ore, Christian Emil & Nina Kristiansen 1998. *Dokumentasjonsprosjektet. Sluttrapport 1992–1997*. Oslo: <<https://www.dokpro.uio.no/sluttrapp.pdf>>. Tilgået februar 2023.
- Svendsen, Mette-Marie Møller 2023. Brugernes blik på Jysk Ordbog: En brugerundersøgelse i leksikografiens tegn. I: Holmer, Louise et al. (red.), *Nordiska studier i lexikografi* 16. Lund & Göteborg: Nordiska föreningen för lexikografi, 389–401.
- Svensén, Bo 2004. *Handbok i lexikografi: Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- Sørensen, Viggo 2019. *Redaktionsregler for Jysk Ordbog i iLEX-regi*. <www.jysk.au.dk/publikationer/retningslinjer-for-jysk-ordbogs-redaktion>. Tilgået februar 2023.
- van Keymeulen, Jacques 2018. The Dialect Dictionary. I: Boberg, Charles, John Nerbonne & Dominic Watt (red.), *The Handbook of Dialectology*. Hoboken, NJ: John Wiley & Sons, 39–56.

Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omkringliggende øer 1992ff. Gudiksen, Asger, Henrik Hovmark & Karen Margrethe Pedersen (red.), Center for Dialektforskning, Institut for Nordiske Studier og Sprogvidenskab, Københavns Universitet. København: C.A. Reitzels Forlag (bd. 1-8 og tillægsbind), Odense: Syddansk Universitetsforlag (bd. 9-12).