

Indholdsfortegnelse

Finale Project: “Nationalistiske ord og emotionel tone i partipolitiske taler omkring Flygtningekrisen, 2015”	3
Forfattere	3
Keywords:	3
Abstract	3
1 Indledning	3
1.1 Problemformulering og motivation	3
1.2 Baggrund og historisk kontekst	4
1.3 Forventninger (til undersøgelsens resultater) og hypotese	5
2. Metode	5
2.1 Research Design	5
2.2 Software Framework and Materials	6
2.3 Data Acquisition and Processing	7
2.4 Udvælgelsen af nationalistiske ord	8
3. Empiriske Resultater	10
3.1 Analyse (1): Forekomst af nationalistiske ord på tværs af talerne	10
3.2 Analyse (2): De nationalistiske ords emotionelle anvendelse i talerne via Sentida	12
4. Diskussion	14
4.1 Resultatdiskussion	14
4.2 Metodediskussion	15
5. Konklusion	16
Author contributions	17
Ugentlige afleveringer	19
Deklaration:	19
Uge 08 - Regex og OpenRefine	19
Uge 10 - Danske konger (Tidyverse og Rmarkdown)	24
Uge 12 - Vores interaktive kort over danske museer	27
Uge 15 - Danske digitale arkiver	29
Litteraturliste	38
Data	40
Deklaration for anvendt Generativ Kunstig Intelligens (GAI)	41
Software metadata	41

Table 1 – Software metadata 41

Table 2 – Data metadata 43

Finale Project: “Nationalistiske ord og emotionel tone i partipolitiske taler omkring Flygtningekrisen, 2015”

Forfattere: Christine Clausen (202106872.), Anna Leth Christiansen(202206180.) og Helene Knudsen (202207738)

Keywords:

Nationalism, refugees, refugee crisis, immigration, Denmark, politics, Sentida, R-studios, tidyverse

Abstract

This assignment explores the use of nationalistic language in seven political speeches delivered by leaders of their respective parties during the months surrounding the Danish general election in 2015, a period in which the international refugee crisis began to impact Danish society.

Using the digital tool RStudio, the seven speeches are processed and analyzed with the aim of comparing them to our expectation that right-wing parties (DF, V, and K) employ nationalistic language more frequently and in a more positive context than left-wing parties (EL and R).

Through the use of RStudio and the Sentida package, two illustrations of our analysis are presented, revealing that the majority of our initial expectations were confirmed.

DF, V, and K use more nationalistic language compared to EL. The speech from R does not confirm any of our initial expectations. DF, V, and K stand out by employing more unique nationalistic terms (e.g., “grænsekontrol”) and by displaying the highest mean sentiment scores for sentences containing nationalistic words.

This study concludes that although the initial expectations were largely confirmed, the Sentida package alone is not a sufficient tool for thoroughly exploring nationalistic language in political speeches.

1 Indledning

1.1 Problemformulering og motivation

Partipolitiske taler er et middel til at skabe den offentlige diskurs og samtidig et produkt af heraf.

Politiske taler holdt i tiden omkring det danske folketingsvalg i 2015 er særligt interessante at undersøge, eftersom Danmark i 2015 var præget af flygtningekrisen. Vi vil vi undersøge, hvordan flygtningekrisen blev italesat overfor vælgerne - særligt er vi nysgerrige på hvordan de forskellige politiske partier adskilte sig i deres brug nationalistiske ord med henblik på at skabe en eventuel fælles national identitet.

Derfor vil vi i denne opgave undersøge følgende problemstilling:

Hvordan er forekomsten af nationalistiske ord fordelt på de syv taler fra folketingets partiformænd i perioden april-september 2015? Og er der sammenhæng mellem brug af nationalistiske ord og talernes emotionelle tone beregnet med Sentida-værktøjet i R-Studios?

1.2 Baggrund og historisk kontekst

I 2015 stod den Europæiske Union overfor en vanskelig udfordring: Italiens og Grækenlands systemer for flygtningemodtagelse kollapsede grundet for stort pres og det medførte et europæisk dilemma om, hvordan de mange mennesker på flugt skulle hjælpes og håndteres.

De mere end 1 million mennesker, der var flygtet fra krig og fattigdom i både Afrika og Mellemøsten, og især Syrien, havde ikke længere mulighed for at få asyl i Italien og Grækenland. Derfor bevægede de sig op gennem Europa gående ad motorvejene, og det skabte en panisk stemning i hele Europa (Chege, 2024). Mange lande begyndte som reaktion på flygtningestrømmene at etablere hegn og grænsekontroller, der skulle begrænse flygtningenes mulighed for at komme ind (Wiesner-Hanks et al. 2014. 995).

Flygtningekrisen kom også til at præge den danske værdipolitiske debat. Et ganske synligt eksempel på påvirkningen er, at Danmark i 2016 fik etableret en midlertidig grænsekontrol ved den dansk-tyske grænse - der i øvrigt stadig er aktiv 9 år efter oprettelsen (Chege, 2024). Derudover blev Dansk Folkeparti, der havde folketingets strammeste udlændingepolitik, landets næststørste parti med hele 20,9% af stemmerne ved folketingsvalget i 2015.

Vi husker selv flygtningekrisen i 2015 og folketingsvalget 2015 som første gang, vi oplevede nationalisme og beskyttelse af den danske nationale identitet som temaer i dansk politik.

Vi er interesserede i, hvordan sprog skaber virkelighed. Derfor vil vi i denne opgave undersøge syv partipolitiske taler fra politiske formænd i perioden omkring folketingsvalget 2015.

For at kunne analysere forekomsten af nationalistiske ord i talerne, har vi anvendt to teoretiske artikler som udgangspunkt for undersøgelsen. Første dele af Benedict Andersons *Imagined Communities* (1991) anvendes, da teorien giver perspektiv på arbejdet med, hvordan nationen italesættes som havende en kollektiv identitet.

Sosis og Alcorta arbejder i *Signaling, Solidarity, and the Sacred: The Evolution of Religious Behavior* (2003) med stabilitet i grupper og dermed overlevelse af disse. Artiklen har fokus på den religiøse gruppe, men det er oplagt at overføre denne teori til nationen som gruppe. De teoretiske artiklers begreber udfoldes i vores senere afsnit ”2.4 Udvalgelse af nationalistiske ord”.

1.3 Forventninger (til undersøgelsens resultater) og hypotese

I henhold til opgavens todelte problemformulering, har vi opstillet følgende forventninger til undersøgelsens resultater:

Forekomst af nationalistiske ord

Vi forventer, at partiformændenes politiske placering på hhv. den værdipolitiske højre- og venstrefløj vil have sammenhæng med, hvor ofte de bruger ord af nationalistisk karakter (Lange 2016). Vi forventer derfor, at de værdipolitisk højreorienterede partier, især Dansk Folkeparti og Venstre, i høj grad vil benytte ord af nationalistisk karakter med henblik på at tale til “danskerne” og opfordre til nationalt sammenhold i krisen. Omvendt forventer vi, at de værdipolitisk venstreorienterede partier, især Enhedslisten og Radikale Venstre, vil benytte færre ord af nationalistisk karakter.

Sammenhæng mellem brug af nationalistiske ord og talernes emotionelle tone

Vi forventer, at partipolitiske taler fra de værdipolitisk højreorienterede partiformænd (Dansk Folkeparti, Venstre og Konservative) vil vurderes til at have en overordnet positiv emotionel sentiment score for de sætninger, hvor nationalistiske ord indgår. Denne forventning skyldes at vi tror, at talernes formål er at vække patriotisk følelse i vælgerne, eftersom hvor Danmark er under pres grundet flygtningekrisen.

Vi forventer ikke at de værdipolitisk venstreorienterede partiformænd (Enhedslisten og Radikale Venstre), er ligeså interesserede i at vække en dansk patriotisme i vælgerne. Derfor tror vi, at deres taler vil have en væsentligt mere negativ sentiment score for de sætninger, hvor nationalistiske ord indgår.

2. Metode

2.1 Research Design

For at svare på problemformuleringen udvælger vi alle taler, der opfylder følgende krav:

- 1) skal være afholdt mellem april og september 2015.
- 2) skal være afholdt af en partiformand.
- 3) skal være tilgængelig online.

Med disse krav finder vi frem til 7 partiformænds taler. Det var ikke muligt at finde taler fra alle folketingets partiformænd i den afgrænsede periode eller at finde taler, der var holdt i samme forbindelse. Derfor har vi både taler fra årsmøder, 1. maj, folketingets åbning og folkemødet.

Til bearbejdning af materialet benytter vi indledningsvist Voyant Tools til at danne overblik over talernes indhold. Herfra arbejder vi hovedsageligt med R-Studios' værktøjer til at undersøge talernes nationalistiske sprogbrug. Det danske sentimentanalyseværktøj, Sentida, benyttes til at vurdere talernes emotionelle tone i relation til brug af nationalistiske ord. Derudover har vi forud for arbejdet i R-studios selv foretaget en nærlæsning af talerne for at kunne sammenligne vores egen vurdering af talerne med talernes bearbejdning i R-studios.

2.2 Software Framework and Materials

Dette projekt er kodet på Helenes MacBook Air M2 13" fra 2022, der benytter MacOS - sequoia 15.1.1 som styresystem. Størstedelen af projektet er udarbejdet på R-studios, version 2024.12.1+563 (2024.12.1+563). Vi har undervejs benyttet GAI-værktøjet Microsoft Copilot og Chat-gpt til sparring og gode forslag (se deklaration). Vi har dog redigeret i koderne og kodebeskrivelserne, så de passer til vores projekt.

Vi har desuden brugt følgende pakker:

ggwordcloud	word cloud værktøj til ggplot2
readtext	importere og arbejde med tekstfiler
sentida	se nedenfor
tidytext	gør det muligt at arbejde med tekst og frasortere ord uden betydning for analysen
tidyverse	overordnet pakke, der gør det muligt at hente mange af ovenstående pakker. Derfor taler mange af pakkerne godt sammen

Voyant Tools

Voyant Tools er et tekstgenkendelsesprogram, der gør det muligt at danne sig overblik over et stort tekstmateriale. Vi benytter Voyant Tools i vores indledende undersøgelser af talerne forud for udformningen af den endelige problemformulering. Derudover benytter vi programmets visualisering, den såkaldte "word-cloud", som inspirationskilde i udvælgelsen af nationalistiske ord.

RStudio og SENTIDA

Vi benytter analyseværktøjet *SENTIDA* i R-studios. Sentida gør det muligt at arbejde med sentimentanalyse af danske tekster. Programmets skabere anbefaler det til analyse af politiske tekster (Lauridsen et al., 2019. 17). *SENTIDA* bygger på *AFINN*, der er en dansk ordbog, hvori ord med en ladet værdi indgår og har fået tildelt en værdi på en skala fra meget negativ (-5) til meget positiv (5). *SENTIDA* tilføjer flere relevante ord til *AFINN* (og hertil værdier), som gør det muligt at anvende redskabet i flere domæner (ex. politiske tekster). Disse ord er individuelt udvalgt og værdisat af de danske skabere bag (Ibid., 40-41, 47). Teknologi bygger på *SnowballC*-pakken; her brydes sætninger op, og pakken har en teknik, der gør, at *SENTIDA*-leksikonet virker, trods ord med forskellige grammatiske former (Ibid., 45). *SENTIDA* forstår også kontekst i form af bl.a. negationer og tegnsætning, der påvirker ords ladning, men har problemer med at forstå ironi (ibid., 43, 50).

2.3 Data Acquisition and Processing

Opgavens data er som nævnt under punkt 2.1 fundet via offentlige kilder til nedskrevne taler. Vi har fravalgt videooptagelser af taler, da transskribering af disse vil være for omfattende til denne opgaves format. Vi mener, at de syv udvalgte taler er repræsentative for den politiske diversitet i Danmark i perioden.

Her følger en liste over valgte taler og deres ophavssted. Direkte links til talerne kan findes i litteraturlisten:

- DANSK FOLKEPARTI (DF): Kristian Thulesen Dahls tale ved DF's årsmøde 19. september 2015, Herning: www.dansketaler.dk
- DET KONSERVATIVE FOLKEPARTI (K): Søren Pape Poulsens tale ved Folketingets åbning, 6. oktober 2015, København: www.altinget.dk
- ENHEDSLISTEN (EL): Johanne Schmidt-Nielsens 1. maj-tale 1. maj 2015, ukendt sted: www.dansketaler.dk
- RADIKALE VENSTRE (R): Morten Østergaards tale på Folkemødet, 11. juni 2015, Allinge: www.dansketaler.dk
- SOCIALDEMOKRATIET (S): Mette Frederiksen tale ved Socialdemokraternes partikongres 19. september 2015, Aalborg: www.dansketaler.dk
- SOCIALISTISK FOLKEPARTI (SF): Pia Olsen Dyhrs tale ved Socialistisk Folkepartis landsmøde 19. april 2015, København: www.dansketaler.dk

- VENSTRE (V): Statsminister Lars Løkke Rasmussens tale v. Folketingets åbning, 6. oktober 2015, København: www.stm.dk

Forkortelsen i parentes indikerer navnet på pdf.filerne i GitHub datamappen.

2.4 Udvælgelsen af nationalistiske ord

Siden 2. verdenskrig har alle succesfulde revolutioner eller paradigmeskift defineret sig selv med nationalistiske termer (Anderson 1991. 2). Selvom dette er tilfældet, er der bred teoretisk enighed om, at det er svært at indkapsle, hvad der kan siges at være nationalisme og hvordan man kan definere det (Ibid. 4, 5).

Derfor har vi i arbejdet med udvælgelsen af ord, vi vil vurdere som nationalistiske, støttet os til de to teoretiske artikler, vores Voyant-wordcloud (se nedenfor), Microsoft Copilot og vores egen læsning af talerne.

For at komme frem til relevante ord i Voyant, har vi lavet en stop-word-liste. Vi har anvendt den overordnede liste fra undervisningen og tilføjet enkelte ord, der blev synlig i Voyant-wordcloud, som vi vurderede var fyldord uden analytisk værdi. Desuden har vi fjernet følgende ord fra den udleveret stop-word-liste: '*de, deres, dem, vores*'. Disse ord har en analytisk værdi, når vi arbejder med nationalisme og gruppedannelse (Anderson 1991. 17). Denne grænsedragning er med til at skabe en retorik, hvori der findes et '*os*' / '*vores*' og et '*dem*' / '*deres*', hvilket gør at der skabes et tilhørsforhold mellem de mange mennesker, som ikke kender hinanden, men er en del af samme fællesskab (Ibid. 17).



Figur 1: wordcloud lavet i voyant

Dette understøttes ligeledes af artiklen fra Sosis og Alcorta, da de nævner vigtigheden af, at man som individ i en gruppe må agere på en bestemt måde, som er i overensstemmelse med de værdier og normer, der gør sig gældende i fællesskabet. Ellers bliver man ekskluderet, og dermed defineret som en *free-rider* (Sosis og Alcorta 2003. 267). Når man i en gruppe opfattes som en *free-rider*, dækker det over, at et individ afviger fra den homogene gruppe, og dermed bliver anset for at være en trussel for fællesskabet, hvilket kan føre til eksklusion fra gruppen (Ibid., 265, 267).

De nævner ydermere, hvordan fælles værdier skaber stabilitet, hvilket også er et centralt element i ønsket om at lykkes med en nationalistisk retorik med et politisk formål (Ibid. 265).

I udvælgelsen af hvilke ord, vi har fundet relevante at undersøge, spiller alle disse fire komponenter en rolle. Anderson har særligt fokus på, hvordan en nation er et forestillet fællesskab, som tilhører en bestemt gruppe (Anderson 1991. 6); i vores case udgøres det forestillede fællesskab af det, man kunne kalde “den danske gruppe”. Vi har ud fra dette grundlag valgt ord som: ‘*grænser*’ og ‘*grænse*’. Desuden uddybes det, hvordan det forestillede fællesskab er afhængig af grænsedragninger - måden hvorpå fællesskabet forestilles afgør, hvem der er med i fællesskabet og hvem der står udenfor (Ibid.). Vi valgte derfor at inkludere ordene: ‘*dansk*’, ‘*danmark*’ og ‘*dansker*’.

Disse blev foreslået af Copilot og indgik i Voyant-word-cloud. Vi finder, at disse ord kan drage grænser mellem, hvad det forestillede fællesskab bygger på.

Derudover er ordene '*samfund*', '*regeringen*', '*velfærdssamfundet*' og '*bevare*' valgt ud fra den opdaterede word-cloud, som i højere grad repræsenterede talernes temaer, hvortil ordet '*tradition*' udelukkende er valgt ud fra Co-pilots forslag. Vi valgte også at tilføje ordet '*grænsekontrol*' efter selv at have læst de forskellige taler, da vi vurderede at dette ords forekomst og kontekst var relevant.

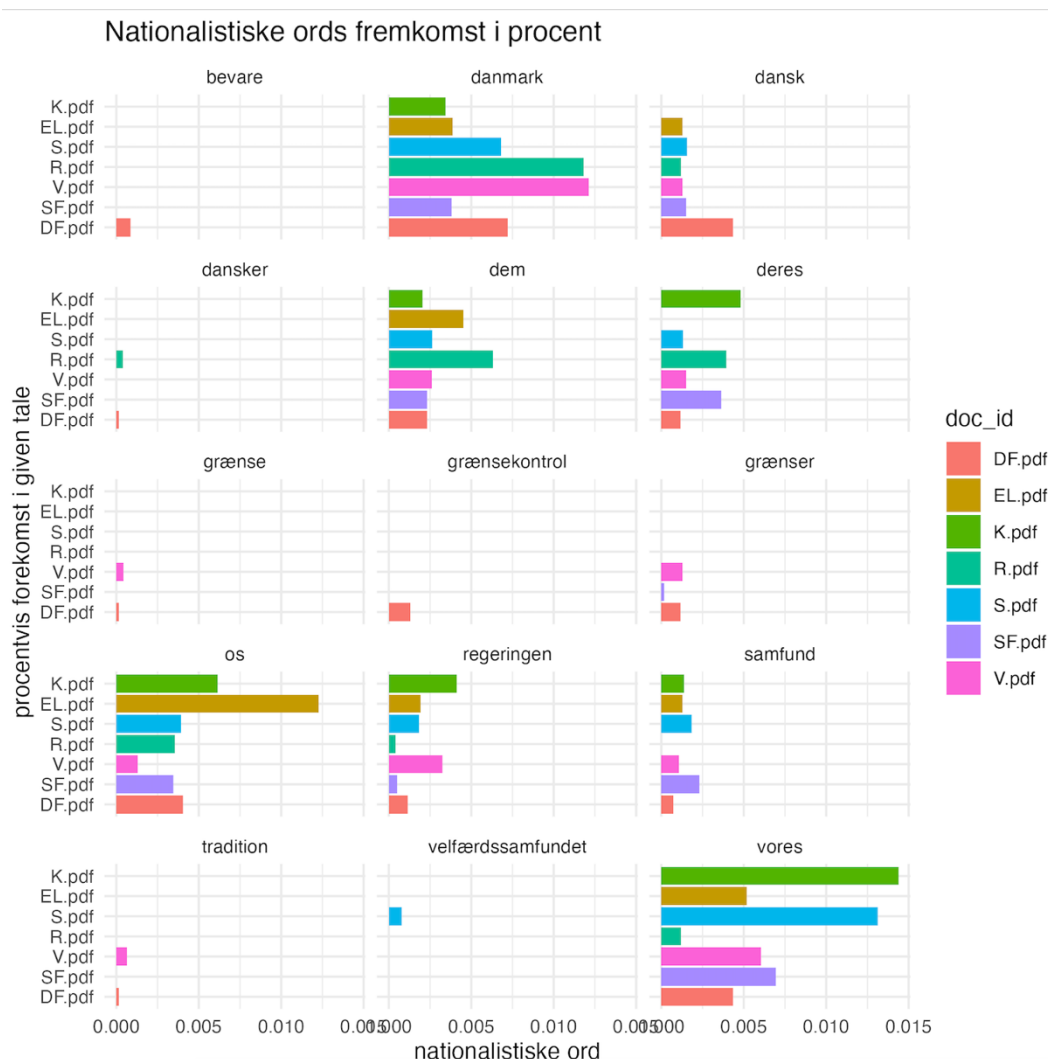
Det er vigtigt at nævne, at når vi omtaler ordene som 'nationalistiske', er det ikke et politisk begreb, men et analytisk begreb med baggrund i vores valgte teorier.

3. Empiriske Resultater

3.1 Analyse (1): Forekomst af nationalistiske ord på tværs af talerne

Fremgangsmåde

Vi henter først vores data i pdf-format til R-Studios. Herefter skal teksten splittes op i ord, for at vi kan arbejde med antallet af de forskellige nationalistiske ord, som vi har udvalgt. Disse ord opretter vi i en liste, for at R ved, hvilke ord, der skal tælles og efterfølgende sammenlignes. Da talerne ikke er lige lange, sørger vi for, at antallet af ord ses i procent og ikke i antal. Det er dette, der ender med at fremkomme af den endelige visualisering. Den viser forekomsten af forskellige ord i søjlediagrammer med samme enheder på x- og y-aksen, så de nemt kan sammenlignes. X-aksen udtrykker det nationalistiske ords fremkomst i procent, mens y-aksen viser data (partiet bag talen).



Figur 2: nationalistiske ords fremkomst i procent. Lavet i R

Resultatanalyse

Her ses en samlet visualisering af, hvor ofte de 15 udvalgte nationalistiske ord forekommer i hver af de politiske taler. Styrken ved visualiseringen er, at den viser et samlet overblik over brugen af de nationalistiske ord på tværs af partiformændenes taler, og at den samtidig kan bruges til at undersøge brugen af hvert enkelt ord.

Ordene *‘danmark’*, *‘dem’*, *‘os’* og *‘vores’* nævnes i alle syv politiske taler, uanset partiformændenes værdipolitisk ståsted. Dette kan tyde på, at disse ord ikke nødvendigvis relaterer sig til flygtningekrisen og partiernes værdipolitiske ståsted, men også bruges i andre kontekster. Samlet set er *‘danmark’*, *‘os’* og *‘vores’* de mest populære ord på tværs af partierne, men ordenes fremkomst mængde i talerne er bemærkelsesværdigt forskelligt. *‘Danmark’* bruges mest af Venstre og Radikale Venstre, og mindst af SF og Konservative.

Graferne for brugen af ordene 'os' og 'vores' er en interessant indgangsvinkel til at sammenligne talerne. Ordet 'os' beskriver en gruppe, mens ordet 'vores' beskriver noget, som gruppen besidder - altså kan ordene bruges til at sammenligne, hvilke partier der italesætter gruppen som enhed og hvilke partier, der italesætter det, gruppen besidder.

På dette område adskiller Enhedslistens tale sig markant fra Konservative og Socialdemokratiets taler. Enhedslistens tale bruger ordet 'os' langt oftere end alle de andre partier. Til sammenligning bruges 'vores' allerflest gange i Konservative og Socialdemokratiets taler, mens det benyttes markant mindre i Enhedslistens tale.

Denne sammenligning giver antydning af, at Enhedslistens partiformand har interesse i at danne gruppen 'os', der forbinder vælgerne til én enhed, mens Konservative og Socialdemokratiets partiformænd har større interesse i at tale til det, vælgerne besidder med henblik på at være beskyttere for det, der tilhører 'os'.

Et blik på visualiseringen kan også synliggøre, at de inkluderende gruppedannelses ord 'os' og 'vores' bruges markant oftere end de ekskluderende grænsedragningsord 'dem' og 'deres' på tværs af alle syv taler.

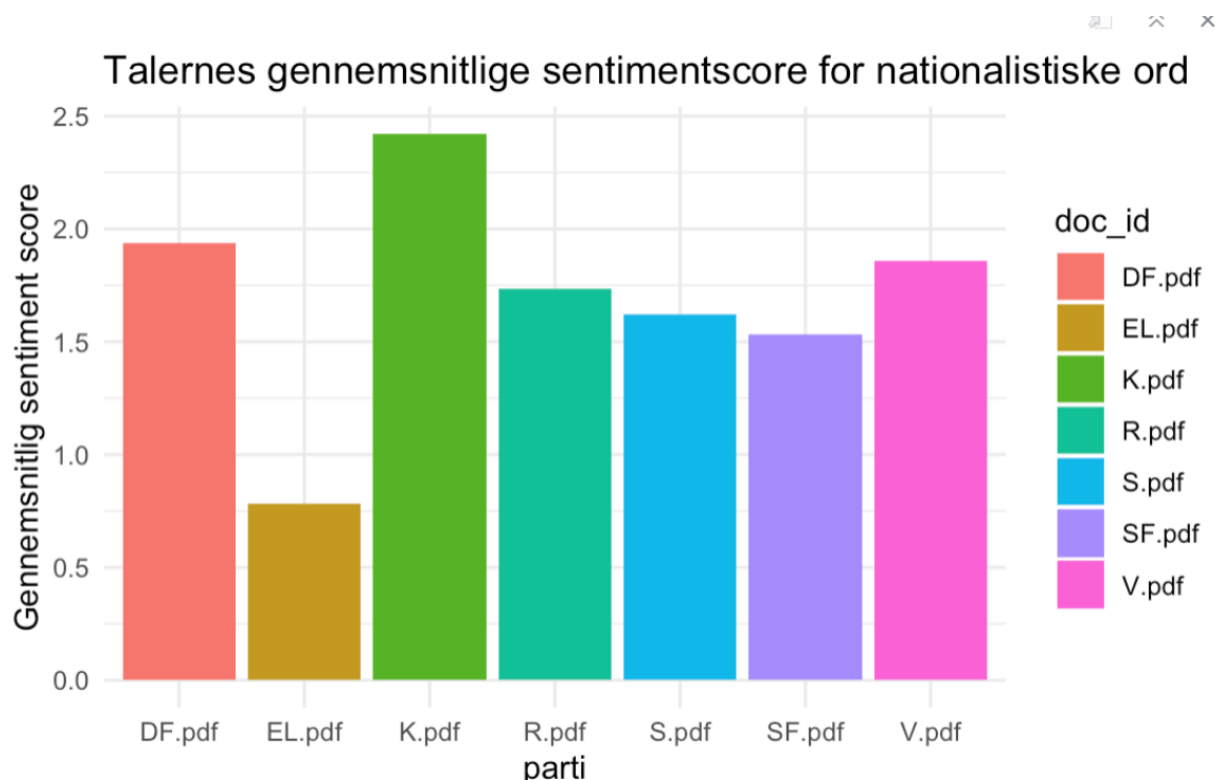
Visualiseringen giver også mulighed for overblik over, hvilke unikke ord, der kun bruges af et eller få andre partier. Dansk Folkepartis tale adskiller sig markant fra de andre partiformænds taler ved brug af unikke ord, da det er den eneste taler, der italesætter "grænsekontrol" - et tiltag, der senere blev vedtaget i Folketinget (Chege, 2024). Derudover benytter DF's partiformænd også de unikke nationalistiske ord 'grænse' og 'grænser'. Disse ord bruges også af Venstre (og meget lidt af SF).

3.2 Analyse (2): De nationalistiske ords emotionelle anvendelse i talerne via Sentida

Fremgangsmåde

Vi henter Sentida og indstiller pakken, så den kan forstå danske specialtegn. Teksten skal nu splittes op i sætninger, da Sentida analyser sentiment score ud fra ordets kontekst i sætningen. Herefter udregner Sentida sentiment score pr. sætning for hele talen. Vi grupperer disse efter parti. Derefter beder vi Sentida om at beregne den gennemsnitlige sentiment score pr. tale, men kun i de sætninger, der indeholder de nationalistiske ord. Det gør det muligt for os at lave en visualisering af, hvilken gennemsnitlig sentiment score, der er for de nationalistiske ord i hver enkelt tale, som vi her vil

arbejde analytisk med.



Figur 3: talernes gennemsnitlige sentimentscore for nationalistiske ord. Lavet med Sentida i R

Resultatanalyse

Her ses en visualisering af de politiske talers gennemsnitlige sentiment score.

I denne analyse er der udelukkende valgt sætninger, hvori der indgår ord fra listen med nationalistiske ord. Den gennemsnitlige sentiment score forklarer, hvordan Sentida gennemsnitligt vurderer den emotionelle ladning af de sætninger, hvori de nationalistiske ord indgår på en skala fra -5 til +5.

Visualiseringen ovenfor viser først og fremmest, at den gennemsnitlige sentiment score for alle syv politiske talers sætninger med indhold af nationalistiske ord er positive. Dette finder vi interessant, eftersom vi ikke havde forventet, at de værdipolitisk venstreorienterede partier (EL og SF) ville have en positiv gennemsnitlig sentiment score.

Konservatives tale vurderes til at have den mest positive stemning i de sammenhænge, hvor der indgår nationalistiske ord, efterfulgt af de næstmest positive taler, Dansk Folkeparti (DF) og Venstre (V). Alle tre talers nationalistiske sætninger har en gennemsnitlig sentiment score over +1,75, med Konservatives tale i top med en score på knap +2,5.

Enhedslistens tale (EL) adskiller sig dog markant fra de øvrige som talen med den klart laveste sentiment score på omkring +0,8.

Disse fund i analyse 2 underbygger vores fund i analyse 1. Her sammenlignede vi blandt andet Enhedslisten og Konservatives forskellige brug af ordene 'os' og 'vores'.

Konservatives tale, der bruger ordet 'vores' mest af alle partier, er også den tale, der har den højeste sentiment score, mens Enhedslistens tale, der bruger ordet 'os' mest, er talen med den laveste sentiment score.

4. Diskussion

Vi vil nu både diskutere, hvad vores resultater kan bruges til, og hvordan det er relevant i kontekst til vores valgte teori. Samtidig vil vi også diskutere vores metodiske arbejde og muligheden for denne metodes fremtidige brug i arbejdet med nationalisme i politisk retorik.

4.1 Resultatdiskussion

Det er særligt interessant at kigge på resultaterne fra analyse (1), hvor figur 2 viser fordelingen og fremkomsten af nationalistiske ord for alle syv taler. Udover ordet 'danmark' er de mest brugte ord 'vores' og 'os', hvilket tyder på, at der er et øget fokus på gruppedannelse. Størstedelen af partierne har også et større fokus på et fælles 'os' fremfor et ekskluderende 'dem'. Som Anderson påpeger i sin teori, er dette retoriske valg også en måde, hvorpå man kan definere et forestillet fællesskab, som også har til formål at skabe grænsedragninger. Gennem et retorisk 'os' og 'vores' skabes der implicit også en modsætning til det forestillede fællesskab og tilhørsforhold. Det øgede fokus på at skabe en gruppedannelse med fælles værdier, kan altså også være med til at tydeliggøre, hvem der afviger fra den definerede gruppe og dermed kan betegnes som en *free-rider* (Sosis og Alcorta 2003. 267).

Analyse 2 giver os også nogle interessante observationer. Sentidas sentiment score går fra -5 til 5. I vores sentimentanalyse af talerne bliver de vurderet til at have en sentimentværdi mellem +0,75 til ca. +2,5. Forskellen er altså relativ stor i, hvilken stemning Sentida vurderer, og det viser, at det vil være relevant at arbejde med yderligere sammenligning af talerne for at undersøge, hvordan stemningen knytter sig til de enkelte ord i figur 2.

V, DF, og K adskiller sig markant fra de andre partier og har en mere positiv stemning end de andre grupper. De tre partier har ikke nogen tydelige fællestræk i figur 2, men analyse 2 peger på, at de har en mere positiv fortælling om nationen. Da begreberne 'os', 'vores' og 'Danmark', er de ord,

der generelt fylder mest i talerne, kan det godt pege på, at det er nationen som et forestillet fællesskab, der bærer den positive fortælling, når DF, V og K anvender nationalistiske ord i talen. EL adskiller sig markant fra de andre partier: deres tales sentimentscore er under 1, og samtidig indgår der ikke mange forskellige nationalistiske ord i netop denne tale. Måske fylder nationen ikke meget i talen, eller måske er talen mere fokuseret på Danmarks rolle som hjælper i flygtningekrisen fremfor nation som indre fællesskab.

4.2 Metodediskussion

Der er flere ting, som er vigtige at tage højde for samt overveje i forståelsen af undersøgelsens resultater. Det er en væsentlig pointe, at talerne fra EL, R og SF er skrevet før folketingsvalget d. 18 juni 2015, men talerne fra DF, K, S og V er skrevet efter. Ligeledes skal der tages højde for at talerne er holdt forskellige steder og med forskellige formål, da det ikke har været muligt at finde taler fra alle partierne, som forholder sig til den samme præcise kontekst, på samme tid og på samme sted. Dette kan have betydning for ordlyden i talerne. Dertil er det vigtigt at påpege, at talerne ikke udelukkende forholder sig til flygtningekrisen, men også berører andre emner, hvorfor det er et valg vi har taget i gruppen om at lægge det specifikke fokus (flygtningekrisen) nedover talerne i forbindelse med vores opgave.

Vi har valgt taler ud fra kravet om, at de blev holdt ca. i samme periode og af samtidige partiformænd. Det har gjort vores dataudvalg småt, og indenfor denne ramme har længden på talen derfor ikke været en faktor, som vi har kunne tage hensyn til. Problematikken ved dette er behandlet ovenfor. Problemet kunne være løst ved at søge bredere efter taler (ex. holdt af andre politikere end partiformænd, holdt over en længere periode); men dette ville gøre det umuligt at undersøge taler netop omkring flygtningekrisen i 2015. En anden løsning kunne være at arbejde med uddrag af talerne udvalgt i en bestemt længde; men det vil gøre, at endnu mere kontekst mangler.

Når historiske kilder behandles med digitale hjælpemidler, vil man kunne beskyldes historikeren for, at dens rolle er udspillet. Men vores undersøgelse peger på det stik modsatte. Voyant-word-cloud og Copilot hjalp os med at udvælge nationalistiske ord, men det var ikke tilstrækkeligt. Vores læsning af talerne som historikere har gjort, at vi har tilføjet ord, der er vigtige pga. konteksten (ex. grænsekontrol). Den historiske samtid og nutid kender Sentida ikke, og en analyse uden denne vil blive unuanceret. Desuden er det nødvendigt at tilføje nationalisme-teori til den manuelle læsning for at blive opmærksom på, at ord der almindeligvis er stopord (*de, deres, os, vores*), her har en semantisk værdi, og derfor netop skal være objekt for analyse.

Sentida-ordbogen vurderer alle ord med samme værdi: *'grænsekontrol'* bliver ikke vurderet som et mere væsentligt ord end *'de'*, selvom sidstnævnte ord kan have mange betydninger, der ikke har med det undersøgende emne at gøre. Vi ved ikke, hvad der ligger til grund for denne værdiangivelse. Derfor er det også vigtigt, at vi i analysen har lagt vægt på, at *'grænsekontrol'* og *'bevare'* kun er ord, der findes i DF's tale: de ord skaber en stærkere nationalistisk diskurs og politisk agenda end ex. *'Danmark'* gør, som er et ord, der går igen i alle taler. Nedenfor vises to eksempler på, hvordan ordet *'Danmark'* har vidt forskellige betydninger i vores udvalgte taler:

"Et Danmark trygt for alle danskere" - Kristian Thulesen Dahl, Dansk Folkeparti
'Danmark' bliver her brugt nationalistisk, nærmest i en patriotisk kontekst: skaber en samlet gruppe af danskere.

"Lige om lidt skal der være folketingsvalg i Danmark. Alle os, der vil fællesskabet, vi skal sammen sørge for, at nøglerne til statsministeriet aldrig nogensinde igen havner ved Lars Løkke Rasmussen." - Johanne Schmidt-Nielsen, Enhedslisten

'Danmark' bliver her brugt i en praktisk og fordelingspolitisk kontekst: skaber en samlet gruppe af 'os der vil fællesskabet'.

Denne sondring kan kun lade sig gøre med metodisk og teoretisk historieforskning (Schrøder og Jensen 2022, 9).

At tilføje til dette er også, at formålet med Sentida bl.a. er, at der netop på forhånd er tildelt sentiment-værdier til en stor mængde danske ord. Ovenstående viser tydeligt, at konteksten er væsentlig for denne værditilskrivning, og derfor stiller vi os skeptisk overfor, hvorvidt Sentida i sin nuværende form er oplagt at bruge til politisk diskursanalyse, til trods for at det bliver anbefalet af skaberne (Lauridsen et al., 2019. 47). Det er i hvert fald tydeligt, at det skal indgå som et af flere redskaber og aldrig kan stå alene. Fremtidig forskning vil derfor med fordel kunne udvikle tilføjelser til Sentida med tematisk fokus som ex. nationalisme, hvis nogen ønsker at arbejde videre med et større tekstkorpus, der ikke tillader manuel gennemlæsning. Det vil også kræve, at man kan udvælge meget specifikke ord (ex. *'grænsekontrol'*), der har ikke har så bred semantisk betydning (ex. *'os'*)

5. Konklusion

Ud fra opgavens todelte problemformulering og vores forventninger til undersøgelsens resultater, kan vi sammenfatte, at mange af vores indledende hypoteser stemte overens med vores fund.

Analyse (1) viste til vores overraskelse, at ordene ‘*danmark*’, ‘*dem*’, ‘*vores*’ og ‘*os*’ blev brugt i alle partier, uafhængigt af partiets værdipolitiske ståsted. Vores indledende forventninger til brug af nationalistiske ord blev bekræftet ved, at de værdipolitisk højreorienterede partier (DF, V og K) tydeligt brugte flest nationalistiske ord. Især adskilte DF og V sig fra de øvrige partier ved at være de eneste, der italesatte ‘*grænser*’ og ‘*grænsekontrol*’. Konservatives tale udmærkede sig ved at være den tale, hvor ordet ‘*vores*’ blev brugt flest gange.

Talen fra det værdipolitisk venstreorienterede parti Enhedslisten havde som forventet en lav forekomst af nationalistiske ord, og benyttede primært ord som ‘*os*’ og ‘*danmark*’. Vi havde forventet, at Radikale Venstres brug af nationalistiske ord ville ligne Enhedslistens, men i realiteten mindede talens brug af nationalistiske ord mere om Socialdemokratiets tale.

Analyse (2) afslørede, at der som forventet var en sammenhæng mellem brug af nationalistiske ord og sentiment scoren i de sætninger, hvori ordene indgik. Konservatives, Venstres og Dansk Folkepartis taler blev af Sentida vurderet til at have de mest positive emotionelle toner med en gennemsnitlig sentiment score på over +1,75. Enhedslistens tale havde som forventet en mindre positiv emotionel stemning i de sætninger, hvor de nationalistiske ord indgik med en gennemsnitlig sentiment score på omkring +0,8. Vi forventede, at talen fra Radikale Venstre havde en lav sentimentscore grundet partiets venstreorienterede værdipolitik. Det blev dog modbevist med en sentimentscore på knap +1,75.

Vores undersøgelse viser altså, at der stadig er behov for dygtige historikere med viden om politik og nationalisme. Eksemplet med forskellig brug af ordet “Danmark” viste tydeligt, at Sentida ikke er et tilstrækkeligt værktøj til at vurdere, om partipolitiske taler benytter nationalistiske ord med en patriotisk og ekskluderende agenda eller med et inkluderende formål.

Dog var det for os et overraskende fund, at nationalistiske ord blev brugt i relativt høj grad på tværs af alle syv taler. Herfra må vi konkludere, at national identitet er et vigtigt tema uafhængigt af værdipolitisk baggrund, men at der er tydelige forskelle på, *hvordan* partiformændene benyttede nationalistiske ord i deres taler fra 2015.

Author contributions

Authors; Christine Clausen, Anna Leth and Helene Knudsen contributed to the study conception and design. Background research and literature search was done by Christine Clausen, Anna Leth

and Helene Knudsen. Data collection, processing, and analysis in R were performed by Christine Clausen, Anna Leth and Helene Knudsen. The first draft of the manuscript was written by all three of us and we all commented on previous versions of the manuscript. Christine Clausen, Anna Leth and Helene Knudsen read and approved the final manuscript.

Ugentlige afleveringer

Deklaration: Alle i gruppen (Christine Clausen, Anna Leth og Helene Knudsen) har samarbejdet om at lave løsningerne til alle de ugentlige afleveringer, selvom de er blevet afleveret individuelt i løbet af semesteret. Vi har efterfølgende gennemgået og sammenlignet vores ugentlige afleveringer for at sikre, at vi har forstået spørgsmålene og resultaterne på samme måde samt fundet frem til de bedste resultater. De følgende afleveringer er derfor skrevet, udarbejdet, udregnet og fortolket af os alle tre.

Uge 08 - Regex og OpenRefine

Spørgsmål 1: *What regular expressions do you use to extract all the dates in this blurb:*

<http://bit.ly/regexexercise2> and to put them into the following format YYYY-MM-DD?

Som det fremgår af spørgsmålet, fremkommer datoerne i datasættet i forskellige formater. Vha. Regular Expressions kan disse datoer formateres, således at de fremstår som YYYY-MM-DD.

Følgende kommando benyttes for at omskrive datoerne:



Ligning 1: screenshot fra <https://regex101.com/r/sBGaPT/1>

Udtrykket beskriver følgende:

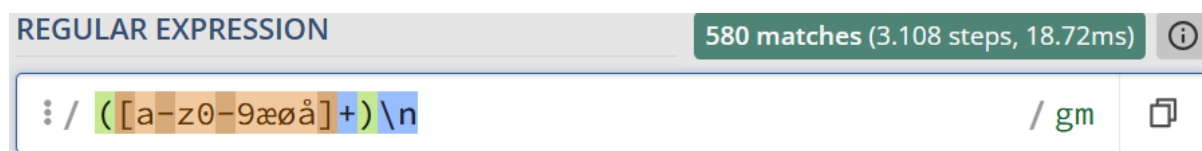
(\d+)	<p>\d: matcher et hvilket som helst tal.</p> <p>+: matcher det førnævnte udtryk (her \d), hvis det fremstår en eller flere gange.</p> <p>Dette fremkommer to gange for at få MM-DD (hhv. gruppe \$1 og \$2) til at fremkomme.</p> <p>() : gør at hvert udtryk defineres som en gruppe.</p>
.	<p>Matcher alle tegn.</p> <p>Dette gør, at hvad end datoerne i den originale tekst er adskilt af, vil det matche.</p>
\s?	<p>\s: matcher mellemrum eller ny linje; her mellemrum.</p>

	?: matcher førnævnte udtryk (her\s), hvis det fremkommer en eller nul gange, da der i nogle tilfælde i den originale tekst indgår mellemrum i datoangivelsen.
(\d{4})	\d: se række 1. {4}: matcher det foregående udtryk (her \d) præcis 4 gange for at få udtryk YYYY (gruppe \$3). (): se række 1.

Efterfølgende vælger vi at substituere med udtryk \$3-\$1-\$2 for at datoerne kommer i formatet YYYY-MM-DD. Den endelige rettede tekst kan ses via linket indsat ved: *ligning 1*.

Spørgsmål 2: Write a regular expression to convert the stopwordslist (list of most frequent Danish words) from Voyant in <http://bit.ly/regexexercise3> into a neat stopwords list for R (which comprises "words" separated by commas, such as <http://bit.ly/regexexercise4>).

Vi har brugt regex for at få den samlede voyant-stopliste (liste adskilt af linjeskift) skrevet om til en R-stopliste (tekst adskilt af "ord", og mellemrum). Vi har lavet nedenstående udtryk for at omskrive:



Ligning 2: screenshot fra <https://regex101.com/r/x5knWJ/1>

([a-z0-9æøå]+)	[]: definerer den liste af tegn, som man ønsker at finde. a-z: matcher ethvert lille bogstav fra a-z. Alle ord med disse bogstaver er inkluderet. 0-9: matcher ethvert tal.
----------------	---

	<p>æøå: tilføjes manuelt som andre tegn, da det ikke-danske program ikke kender dem som bogstaver i en alfabetisk rækkefølge.</p> <p>+: matcher det førnævnte udtryk (her tal og bogstaver indenfor klammen), hvis det fremstår en eller flere gange. Derfor inkluderer udtrykket alle ord.</p> <p>() : gør, at hvert udtryk defineres som en gruppe.</p>
\n	Matcher en ny linje, da ordene i voyant-listen er adskilt af linjeskift.

Se den rettede liste via linket ved: *ligning 2*.

Vi tilføjer til sidst "\$1" i substitutionslinjen, fordi gruppen, som vi har beskrevet ovenfor, alle er ord, der skal omgives af citationstegn, for at den lever op til kravene for en R-stopliste.

Det er nødvendigt at bemærke, at vores regex ikke gør det muligt at inkludere de tilfælde, hvor der indgår andre specialtegn end dem, som vi har inkluderet i ligningen ex. ejefaldet i *cd's*.

Spørgsmål 3: Then take the stopwordlist from R <http://bit.ly/regexexercise4> and convert it into a Voyant list (words on separate line without interpunction)

Vi bruger stort set samme regex-indtastning som i spørgsmål 2. Der er dog nogle ændringer for at konvertere teksten fra en R-stopliste (tekst adskilt af "ord", og mellemrum) til en voyant-stopliste (liste adskilt af linjeskift). Vi vil derfor nedenfor kun beskrive de ændringer, som vi har lavet:

.	Vi har tilføjet tegnet en gang før () og to gange efter (). Det gør, at vi kan fjerne de symboler, som i R-listen står omkring alle ordene.
\n	Har vi fjernet, da R-listen ikke indeholder noget linjeskift, som vi ønsker at slette.

Vi tilføjer til sidst \$1\n i substitutionslinjen, fordi gruppen som vi har beskrevet ovenfor, alle er ord, der skal stå på en ny linje, hvis listen skal leve op til kravene for en voyant-stopliste. Se den bearbejdet ordliste her: <https://regex101.com/r/5W1mMP/1>

Spørgsmål 4: Does OpenRefine alter the raw data during sorting and filtering?

OpenRefine ændrer ikke rådata gennem sortering og filtrering uden brugerens involvering. Dog kommer OpenRefine med forslag til sortering og filtrering, der skal godkendes af brugeren.

OpenRefine er et godt værktøj til at sammenslå, sortere og filtrere rådata, hvilket vi tilsammen kan kalde at bearbejde data. Eksempelvis kan vi aktivt vælge at slå kategorier sammen (ex. landsoldat og nationalsoldat), hvis vi vurderer, at det vil understøtte den historiske kontekst eller det spørgsmål, som vi gerne vil svare på. Alt manipulering af rådata kræver tolkning og analyse, og man kan derfor trække dataen i mange retninger og bruge den til mange formål.

Filtreringsprocessens formål er altså at forberede datasættet til de næste skridt. Samtidig bliver regnearket langt mere læseligt for iagttageren, når de hurtigt og effektivt skal finde data.

Spørgsmål 5.1: Fix the [interviews dataset](#) in OpenRefine enough to answer this question: "Which two months are reported as the most water-deprived/dryest by the interviewed farmer households?"

Datasættet skal bearbejdes, så alle forskellige stavemåder, tegnangivelser osv. ikke forstyrrer aflæsningen af måneder og så hver enkelt måned (og NULL og (blank)) kun forekommer som én kategori.

Ud fra vores redigerede data kan vi konkludere, at de to tørreste måneder i 2016 var september og oktober. De fremkommer hhv. 70 og 74 gange i interviewene som de måneder, hvor der ikke er nok vand. Konklusionen er dog behæftet med stor usikkerhed, da kategorien *months_no_what* i 126 tilfælde er angivet som blank og i 45 tilfælde er angivet som NULL.

Spørgsmål 5.2: Real-Data Challenge: What are the 10 most frequent occupations "erhverv" among unmarried men or women of 20-30 years in [1801 Aarhus](#) census dataset? (hint: first select either

men or women to shrink the dataset to a manageable size, then filter by age, and then use merging to cut the erhvervvariation ruthlessly.)

De 10 mest forekommende erhverv for ugifte mænd i alderen 20-30 år:



erhverv	
161 choices Sort by: name count	
Soldat	458
Tjenestekarl	42
Væverarbejde	27
Rytter	26
Skrædder	25
tjener familien	24
karl på gård	22
Matros	20
Snedkerarbejde	20
Skriverarbejde	17

Det er nødvendigt at bemærke, at konklusionen er behæftet med en del usikkerhed. Der er 1880 tilfælde, hvor mandens erhverv ikke er angivet (blank), og samtidig er kategorien *soldat* udtryk for mange forskellige typer af soldat. Mange af mændene er soldat samtidig med, at de varetager et andet erhverv, og det kan være en forklaring på, at der er så mange med netop soldater-erhvervet.

De 10 mest forekommende erhverv for ugifte kvinder i alderen 20-30 år:

Denne konklusion er også behæftet med en del usikkerhed. Størstedelen af kvindernes erhverv falder i kategorien NULL, som vi har lavet som én samlet kategori for alle de angivelser, der ikke er et erhverv (ex. krøbling). Desuden er der 2142 tilfælde, hvor kvindens erhverv ikke er registrere



erhverv	
25 choices Sort by: name count	
NULL	58
Tjenestepige	38
Væver	24
Tjener forældrene	15
Spinder	13
Håndarbejde (sy)	12
Husjomfrue	9
Husholderske	8
Lever af sine midler	8
Skræder	7
Køkkenpige	6

(blank).

Uge 10 - Danske konger (Tidyverse og Rmarkdown)

Det første vi gør i arbejdet med danske konger er at uploade vores oversigt over danske konger i CSV-format til R ved brug af tidyverse-værktøjet.

Dernæst svarer vi på følgende spørgsmål omkring vores kings-objekt samt data:

Which of these functions is a tidyverse function?

```
# FILL IN THE CODE BELOW and review the outputs

kings1 <- read.csv("data/danske_konger.csv")

kings2 <- read_csv("data/danske_konger.csv")

kings3 <- read.csv2("data/danske_konger.csv")

kings4 <- read_csv2("data/danske_konger.csv")
```

Alle de funktioner med underscore 'read_csv' er tidyverse versioner, hvilket omfatter følgende to funktioner: (read_csv) og (read_csv2). Dette eliminerer funktionerne Kings1 og Kings3, da de lyder: read.csv eller read.csv2. Den version, som vi skal bruge af de to tilgængelige, er derfor Kings4, da vi i læsningen af denne funktion, får de bedste resultater i form af en tabel, som passer til vores data, da den er til brug på kommaseparerede filer. Det er derfor denne funktion, vi i et nyt objekt, som vi blot kalder kings, kan arbejde videre med.

What is the result of running class() on the kings object created with a tidyverse function

Resultatet af at køre kommandoen class(kings) giver følgende output:

```
kings <- kings4

class(kings)

## [1] "spec_tbl_df" "tbl_df"      "tbl"        "data.frame"
```


“Spec_tbl_df” viser, at dataen er en tabel (som data.frame og tbl), men den er nu lavet til en mere avanceret version, spec_tbl_df, der inkluderer mere detaljeret information om selve datasættet og formatet.

Det mest væsentlige ved denne kommando er, at vi nu kan se, at vores data er i et tibel (tbl) format, som tidyverse forstår som numerisk. Dette er præcis, hvad vi ønsker, da det er essentielt, at tidyverse forstår dataen som numerisk, hvis vi ønsker at kunne bruge dataen til at udregne forskellige ting, som vi skal senere hen.

How many columns does the object have when created with these different functions?

Vi bruger kommandoen “ncol(kings)” til - gennem tidyverse - at tælle antallet af kolonner i vores objekt. Dette viser os, at der er 6 kolonner i vores funktion:

```
ncol(kings)
```

```
## [1] 6
```

Show the dataset so that we can see how R interprets each column

Ved at skrive kommandoen glimpse(kings) kan vi få et overblik over, hvordan R forstår hver kolonne. Dette viser at R forstår hver kolonne som numerisk (dbl) undtagen kolonnen “danske regenter”, som fortolkes som characters (chr), hvilket stemmer overens med vores forventning og krav til datasættet:

```
glimpse(kings)
```

```
## Rows: 58
## Columns: 6
## $ `Danske regenter`      <chr> "Frederik_X", "Margrete_ll", "Frederik...
## $ Fødselår               <dbl> 1968, 1940, 1899, 1870, 1843, 1818, 18...
## $ Dødsår                 <dbl> NA, NA, 1972, 1949, 1912, 1906, 1863, ...
## $ `Regeringperiodens_ begyndelse` <dbl> 2024, 1973, 1947, 1912, 1906, 1863, 18...
## $ Regeringperiodens_ende   <dbl> NA, 2024, 1972, 1947, 1912, 1906, 1863...
## $ `Varighed (antal år)`    <dbl> 1, 52, 25, 35, 6, 43, 14, 9, 31, 42, 1...
```

Efter at vi har redegjort for vores kings-objekt og data, vil vi nu besvare følgende arbejdsspørgsmål:

Calculate the duration of reign for all the kings in your table

For at udregne regeringsperiodens varighed for alle monarker benytter vi mutate-funktionen ved at trække kolonnen “Regeringsperiodens_begyndelse” fra kolonnen “Regeringsperiodens_ende”. Dette gør vi samtidig med, at vi via funktionen danner en ny kolonne, hvor resultatet for alle monarker står. Desuden sørger vi for, at fjerne NA's i de kolonner vi bruger til udregningen:

```
#your code
kings_duration_no_na <- kings %>%
  mutate(duration = Regeringsperiodens_ende - `Regeringsperiodens_begyndelse`)

class(kings_duration_no_na$duration)
```

Calculate the average duration of reign for all rulers

```
# YOUR CODE
average <- mean(kings_duration_no_na$duration, na.rm = TRUE)
```

Den gennemsnitlige regeringstid for alle regenter er 19,39286 år inkl. Frederik d. 10, hvis regeringsperiode endnu ikke er slut.

How many and which kings enjoyed a longer-than-average duration of reign?

Ud fra resultatet kan jeg se, at der er 26 konger, som har regeret i flere år end gennemsnittet på 19,39286 år. Se nogle af dem nedenfor (for at se alle 26 monarker, tilgå Github-link):

```
## # A tibble: 26 × 7
##   `Danske regenter`   Fødselår Dødsår `Regeringperiodens_ begyndelse`
##   <chr>              <dbl>   <dbl>          <dbl>
## 1 Christian_IV_myndig 1577    1648          1596
## 2 Margrete_II        1940     NA           1973
## 3 Christian_IX        1818    1906          1863
## 4 Erik VIII_af_Pommern 1382    1459          1396
## 5 Christian_VII       1749    1808          1766
## 6 Valdemar II_Sejr    1170    1241          1202
## 7 Christian_X         1870    1949          1912
## 8 Valdemar IV_Atterdag 1320    1375          1340
## 9 Christian_I         1426    1481          1448
## 10 Erik VI_Menved     1272    1319          1286
## # i 16 more rows
## #   <dbl> <dbl> <dbl>
```

How many days did the three longest-ruling monarchs rule?

De tre længst regerende regenter er følgende: Christian IV_myndig, Margrethe II, Christian IX. Se antallet af dage, hvor de hver især har regeret nedenfor. Vi har angivet antallet af dage på et år som: 365,25.

Christian IV_myndig og Margrethe II har begge regeret i 18980 dage, mens Christian IX har regeret i 15695 dage.

Uge 12 - Vores interaktive kort over danske museer

Assignment: Significance: Explain in 250 words what your visualisation (sentiment chart, map, or colonial legacy timeseries or regression chart) represents and what are its implications vis-a-vis your [historical] research question? Also, consider briefly what your your visual does not show



Dette kort er til dig, der:

- Er ven med Anna, Christine eller Helene, og ønsker at få hendes gode anbefalinger til seværdige (og ikke-seværdige) museer i Danmark.
- Skal på ferie et specifikt sted i Danmark, og ønsker at besøge en håndfuld gode museer nær den specifikke lokalitet.

Ovenstående billede viser et interaktivt kort, fremstillet ud fra vores personlige oplevelser med besøg på diverse museer i Danmark. Kortets interaktive funktioner gør brugeren i stand til hurtigt at danne sig overblik over de enkelte museers geografiske placering. Ved at trykke på én af lokationerne, får brugeren vist en oversigt, der indeholder museets titel, vores korte, subjektive beskrivelse af museets indhold, museets region (relevant ift. Det danske sundhedsvæsen) og vores subjektive vurdering af museets oplevelsesværdi på en skala fra 1-5.

Fordelen ved et interaktivt kort som dette er, at museernes synlige geografiske placering kan øge tilgængeligheden for brugeren og måske motivere brugeren til at besøge ét eller flere museer i nærheden af brugerens lokalitet. Dette korts største begrænsning er, at beskrivelserne af museer er meget subjektive, eftersom de afspejler vores personlige oplevelser på museerne - og måske denne oplevelse ville være en anden, hvis vi besøgte museet i dag. Derudover har vi langt fra besøgt alle

museer i Danmark, og af denne årsag kan der mangle vigtige museer set fra et historisk og kulturelt perspektiv.

Uge 15 - Danske digitale arkiver

Deklaration: Vi har valgt at strukturere opgaven ud fra de tre dele, som instruktionerne lægger op til. Del 2 har vi primært skrevet som sammenhængende tekst uden at inddrage spørgsmålene eksplicit for en bedre læseoplevelse. Liste over alle spørgsmål kan ses i GitHub. Vi har valgt at fremhæve et eksempel fra hvert arkiv, da vi mener, at det er dækkende for at svare på spørgsmålene og er repræsentativt for arkivernes materiale og vores historiske arbejde.

DEL 1: redegørelse for Schriver og Jensen (2022)

Schriver og Jensens artikel undersøger det stigende digitaliseringsarbejde i arkivverdenen med udgangspunkt i Det Kongelige Biblioteks projekt *Danmark set fra Luften – Før Google* (DSFL), hvor historiske luftfotos er blevet digitaliseret (Schriver og Jensen 2022, 5-10). Artiklen kombinerer teoretiske og historiske refleksioner med en konkret case og peger på, at digitaliserede arkiver ofte behandles ukritisk af historikere. Ifølge forfatterne opfattes arkivet fejlagtigt som en objektiv kilde, selvom det i virkeligheden er et historisk og analytisk produkt – og dermed udgør et subjekt i sig selv i den videnskabelige proces (ibid., 6-7).

Digitalisering skal ifølge DSFL-projektet både demokratisere adgangen til kulturarv og lette historikeres arbejde, men Schriver og Jensen fremhæver, at man som historiker må være opmærksom på, hvad der *ikke* digitaliseres, samt hvordan materialet klassificeres og hvilken metadata, der følger med (ibid., 8-9, 21). I DSFL er udvælgelsen af materiale præget af folkelig interesse, forskningsværdi og kommercielle hensyn (ibid., 12-13). Dette er analytiske valg, som påvirker det samlede arkiv og dets tilgængelighed. Derudover er metadata om digitaliseringsprocessen mangelfuld og ustandardiseret (ibid., 17, 19, 22-23).

Artiklen konkluderer, at selvom DSFL indeholder værdifuldt materiale, mangler der gennemsigtighed i de metodiske overvejelser bag digitaliseringen. Historikere må derfor anerkende, at de arbejder med et allerede fortolket kildemateriale og lade arkivkritik indgå som en aktiv del af den analytiske proces.

DEL 2: digitaliseringscases - Frihedsmuseet om besættelsestiden i Næstved og NæstvedArkiverne om besættelsestiden.

Frihedsmuseet

Modstandsdatabase og digitaliseringsstrategi

Frihedsmuseet blev i 1957 givet som en gave fra den danske modstandsbevægelse til staten og blev underlagt Nationalmuseet. Det er derfor statsfinansieret. Ifølge museets hjemmeside er formålet at formidle valget om at gå ind i modstand, med udgangspunkt i den danske modstandsbevægelse under 2. verdenskrig. Det er relevant at overveje økonomiens betydning for digitale arkiver, da publikumsinteresse i høj grad styrer, hvad der bliver digitaliseret (Jensen, s. 255).

I denne sammenhæng er den digitaliserede "Modstandsdatabase" særlig relevant. Ifølge fanen "Gennemgået litteratur" bygger databasen på over 15.000 bøger og tidsskrifter, gennemgået af frivillige. Jensen (2021) peger på, at frivillige ofte står for transskribering i Danmark, og dermed også præger, hvad der bliver tilgængeligt online (s. 255). Det rejser spørgsmål om, hvilket materiale de frivillige finder relevant at digitalisere. Pomerantz (2015) beskriver med begrebet *Resource Discovery*, hvordan vurderingen af, hvad der er relevant information, ofte er subjektiv (s. 16). Det er centralt, fordi historikere tilstræber deskriptiv fremstilling snarere end subjektiv tolkning.

Det har været vanskeligt at finde en egentlig digitaliseringsstrategi for museet. Det nærmeste er Tilgængelighedserklæringen på museets hjemmeside, hvor det fremgår, at websiden delvist lever op til webtilgængelighedsstandardEN 301 549.

Frihedsmuseets fotoarkiv - En undersøgelse af tilgængeligt webmateriale om Næstved

Om fotoarkivet

Frihedsmuseets fotoarkiv er Danmarks største specialsamling vedrørende besættelsestiden, og indeholder mere end 60.000 billeder, der er scannet ind i høj opløsning (TIF, 600 dpi) og tilføjet til hjemmesidens digitale arkiv.

I vores søgen efter information om arkivskaberen, nåede vi undersiden "*Fotos og rettigheder i relation til Frihedsmuseets arkiv*". Denne side bidrager med information om

digitaliseringsprocessen. De fleste billeder på hjemmesiden stammer fra private familiefotoalbums, som slægtninge afleverer til arkivet. Af denne årsag er det i de fleste tilfælde vanskeligt at finde frem til fotografen.

I tilfælde af publicering af billeder fra arkivet, skal Frihedsmuseet (og fotografen, i fald denne kendes) angives som kilde til billedet.

Eksempel på arkivets indhold om Næstved

For at finde frem til konkrete fotos, der relaterer sig til besættelsestiden i Næstved, benyttede vi søgeordet "Næstved" på søgemaskinen i Frihedsmuseets fotoarkiv. Hertil valgte vi filteret "præcis placering" og afkrydsede alle medier som interessante. Vi fik 59 resultater på denne søgning. Søgeresultaterne er alle beskrevet med samme metadata, eftersom at alle billeder er tagget med emneordet "Næstved".

Af disse 59 resultater har vi her udvalgt et eksempel på resultatet af søgningen på "Næstved".



Gruppe 11D 35A	Modtagergrupper - Sjælland
Mus. nr.	
Emne	Modtagerfolk på parade på Næstved torv. <i>foran Hotel VINHUSET i</i>
Dato	Maj 1945
Fotograf	
Neg. nr.	
Særlige bemærkninger	

Kilde: Frihedsmuseets fotoarkiv.

(Link: <https://samlinger.natmus.dk/fhm/asset/241013>)

Billedets titel er “*Frihedskæmperparade foran Hotel Vinhuset i Næstved i maj 1945*”. Billedet er publiceret af brugeren Brian Larsen d. 16/2/21 og tagget med flere emneord. Brian Larsen har desuden tilføjet koordinater for, hvor han formoder at billedet er taget. Disse koordinater virker rimelige, eftersom billedet er taget foran en kendt lokalitet - Hotel Vinhuset.

På billedets originale arkiveringspapir, der også er digitaliseret, er der en tabel med to kolonner. Venstre kolonne indeholder kategorier for billedets metadata (Gruppe, Mus. nr., Emne, Dato, Fotograf, Neg. nr., Særlige bemærkninger) og højre kolonne indeholder enkelte informationer, der står ud for angivne kategorier.

Billedet er på arkiveringspapiret dateret til “Maj 1945”. Dog er billedet i den digitale arkivering dateret “10. maj 1945 - 31. maj 1945”. Hvor de digitaliserede datoer stammer fra er ikke angivet.

På websiden for hvert enkelt billede i fotoarkivet findes en tabel over “Detaljer” om billedet. Her er det muligt at få oplysninger om billedets licens (i dette tilfælde “All Rights Reserved”), kilde (Nationalmuseet Danmark), Fil-ID (FHM-241017), filnavn (35A5700011-01.tif), original filstørrelse (8 MB) og original opløsning (1875x1530 pixels). Det er ikke muligt at benytte nogle af disse oplysninger som web-genvej, og det er derfor ikke muligt at se indholdet af en eventuel samling eller serie, som billedet er en del af. En mulig vej til relateret information ville derfor om nødvendigt være at kontakte Frihedsmuseets efter flere informationer, eller rette henvendelse direkte til brugeren Brian Larsen for at høre mere om billedets ophav.

NæstvedArkiverne

Kildedatabase og digitaliseringsstrategi

Næstvedarkiverne har en guldgrube af informationer i deres Arkivstrategi 2020-2025. Her er både information om arkivets formål, dets metode til kildeindsamling, digitale strategier og tilgængelighed. Det er vigtigt at påpege, at strategien er normativ, så vi ved ikke, om det i praksis er sådan, at arkivet fungerer. Det forventer vi dog, da arkivet er forpligtet overfor både Arkivloven og arkivets vedtægter (Indledning). På den måde optræder der en del strukturel metadata, som giver information om, hvordan arkiv-strategien er organiseret (Pomerantz 2015, 17).

Arkivets formål er at tilgængeliggøre lokalhistorie, der både skal kunne bruges af borgere, medier, myndigheder og forskere på lokalt såvel som nationalt niveau (2.1). Samlingens materialer har det formål at give den større offentlighed mulighed for at interagere med den kulturarv, institutionerne forvalter (Schrøver og Jensen 2022, 5).

For at opnå mest mulig tilgængelighed bliver materiale løbende digitaliseret af arkivets medarbejdere i samarbejde med frivillige (3.1). Både kilder, der er født digitale og analoge, bliver arkiveret i det digitale arkiv (2). Det er dog sværere at indsamle analoge kilder fra private ex. foreninger, og der foretages altså en vurdering af, hvorvidt kilderne har høj nok værdi for arkivet til at lægge en så stor indsats i indsamlingen (1). Vurderingerne sker systematisk, og det fremgår af arkivet, at dette bl.a. sker ved tematisk opdeling.

Det er hertil vigtigt at huske, at digitale arkiver ikke nødvendigvis indeholder samme materiale som deres analoge sidestykke. Det er vigtigt, at aktørerne tydeligt får kommunikeret til brugerne, at ikke alle arkivets materialer er digitale, og at der er overensstemmelse mellem fysiske og digitale samlinger (Schrøver og Jensen 2022, 17).

Ved digitalisering har arkivet autenticitet i fokus, hvilket skal være med til at sikre arkivalierne for fremtiden (1.1). Det ses især ved, at de har scannet de analoge tekster ind og ikke blot renskrevet dem. Formålet med denne digitalisering er tilgængeliggørelse (2.1), og det vil vi behandle videre i kontekst til metadata nedenfor. De scannede tekster kan læses vha. Online FlippingBook. En del af strategien for denne metadata er ikke blot at præsentere kilderne, men også bruge dem til at fortælle lokalhistorie (2.2). Strategien understreger trods ovenstående, at digitale kilder ikke er anvendelige for alle, og at de analoge kilder derfor stadig skal være tilgængelige for borgeren (2).

Det står ikke tydeligt, hvilken funding arkivet har, men umiddelbart tyder deres årsberetning fra 2020 på, at arkivets økonomiske grundlag er Næstved Kommune.

NæstvedArkivernes tekstarkiv - En undersøgelse af tilgængeligt webmateriale om besættelsestiden

Om fotoarkivet

NæstvedArkiverne indeholder mange års lokalhistorie. Vi vil nedenfor vise, hvilke veje vi har brugt til at finde kilder om modstandsbevægelsen i Næstved og undersøge de konkrete kilders metadata. NæstvedArkiverne har en kildesamling *Besættelsestiden i Næstved 1940-1945*. Denne kategorisering er i sig selv metadata, da det er en tematisk strukturering, som arkivet har foretaget sig i deres digitaliseret version af udvalgte kilder (Pomerantz 2015, 6).

Eksempel 1 på arkivets indhold om besættelsestiden

Kilde: NæstvedArkiverne

Direkte link: <https://online.flippingbook.com/view/133362/>

“Næstved - Mødrehjælpens retningslinjer for tyske sager 1942”

Fundet på denne måde:

<https://www.naestvedarkiverne.dk/> → BRUG SAMLINGERNE → KILDER PÅ NETTET Hverdag under besættelsen → Gå til kilderne → på rullemenuen Mødrehjælpens retningslinjer for tyske sager 1942

I navngivningen af kilden fremgår en præcis datering, og hvem udgiveren er: ikke hvem der konkret har skrevet teksten, men hvilken forening der officielt står bag. Der fremgår heller ikke af hvem eller hvornår den digitaliseret kilde, er publiceret. Desuden beskriver titlen dokumentets format som "*retningslinjer*".

I den scannede version af kilden, kan man i venstre hjørne se "jnr (red. journalnummer) 1995/75). Denne strukturelle metadata beskriver, hvor journalen kan findes i det analoge arkiv. Denne type metadata er en subkategori til betegnelsen administrativ metadata, som giver information om objektets oprindelse og vedligeholdelse (Pomerantz 2015, 17-18).

For at finde andet digitaliseret materiale af samme arkivskaber fører NæstvedArkiverne brugeren til en anden side, www.arkivfinder.dk. Søgeordet: *Næstved - Mødrehjælpens retningslinjer for tyske sager 1942* (overskriften på kilden) giver ingen resultater. Søger vi blot på *Mødrehjælpen* giver det 38 resultater og fortæller, at alle disse kommer fra en arkivskaber. Når vi denne vej igennem finder *retningslinjerne for tyske sager 1942*, får vi information om, hvilket arkiv og serie som teksten indgår i. Desuden er her information om indkomstnummer og bestillingsnummer, der gør det muligt at arbejde videre med den fysiske tekst hvis ønsket.

Det er ikke brugervenligt, at denne information ikke findes samme sted som den digitalt publicerede kilde, og NæstvedArkiverne lægger derfor ikke op til, at det er vigtig information for brugeren.

Kildekritik på de digitale kilder og deres metadata samt henvisningsprocessen

Ovenstående viser, at metadataen for de to arkiver sjældent er standardiseret og ofte mangler vigtig information. Det kan skyldes, at registreringen ofte er foretaget af frivillige. Den bygger derfor på en subjektiv tolkning og ikke en historisk metodisk faglighed. Det er tilstrækkeligt for at den brede befolkning kan opsøge kilder, men det er ikke rettet mod forskere. Når frivillige udvælger, hvilke tekster der skal digitaliseres, bliver det nemt bestemt ud fra personlig interesse. I forskningsøjemed vil det derfor være meget mere nuanceret og anvendeligt stadig at bruge de fysiske arkiver.

I arbejdet med originale analoge kilder og arkiver, henviser man til arkivsamlingens serie og navn. Denne metadata har til opgave at beskrive vores datapunkt. Metadata kan enten fremgå som en nedskrivelse af individuelle genstande og kategorisering af en genstand, men det væsentlige er, at denne type metadata giver oplysninger om noget, der er karakteristisk for

datapunktet, og som gør det muligt for os at finde frem til det ønskede objekt. Ved analoge kilder vil man altid kunne henvise til arkivsamlingens serie og navn, og dette kan derfor altid findes frem. Modsat er det, hvis man henviser til en digital kilde i form af en hjemmeside. I dette tilfælde ville man indsætte et link til den specifikke hjemmeside med oplysningerne: "Senest hentet eller vist d. dag-måned-år", men man vil ikke altid kunne genfinde det, og websitet kan ligeledes blive slettet eller ændret. Hvis en hjemmeside bliver ændret, er de gamle hjemmesider slettet med mindre man sørger for at gemme de andre versioner i et netaktiv som fx Internet Archive. Digitale kilder er derfor mere vanskelige at henvise til, da tilgængelsen af disse ikke altid er sikret og kræver en større forståelse af netarkiver end det er tilfældet med analoge kilder.

Hos både Frihedsmusset og NæstvedArkiverne er der ikke angivet særlige informationer om dette. Det tætteste man kan komme på noget vejledning hos begge udbydere er under genvejen "Tilgængelighed", hvor der blive gennemgået, hvordan man skal forholde sig til "ikke-tilgængeligt indhold" på deres hjemmesider (begge refererer særligt til webtilgængelighedsloven).

DEL 3: diskussion af digitale kilde og processer i historikerens arbejde

Brugen af digitale kilder og digitaliseringsprocesser har i høj grad ændret betingelserne for historisk forskning. Før digitaliseringen var historikere nødsaget til at opsøge relevante kilder i fysisk format i deres forskning. I dag kan historikere lynhurtigt fremsøge et væld af digitaliseret kildemateriale, der ovenikøbet ofte er digitaliseret i en brugervenlig form, f.eks. tekster, der er transskriberet fra gotisk håndskrift til letlæselig maskinskrift. Derudover er tildeling af emneord og organisering af data i kategorier en væsentlig del af digitaliseringsprocessen, der letter søgningen efter relevante kilder til et bestemt spørgsmål.

Alt dette gør, at der er langt bedre betingelser for, at historikeren *hurtigt* kan fremsøge relevante kilder. Med denne hurtighed kommer dog et ansvar.

Historikere må i dag være endnu mere bevidste om vigtigheden af at benytte den kildekritiske metode i arbejdet med digitaliseret materiale: forskere i 2025 skal ikke blot forholde sig kritisk til kildens ophavssituation og indhold, men også til digitaliseringsprocessen. Nutidens historikere må forholde sig kritisk til årsagerne til, at den enkelte kilde er blevet digitaliseret og undersøge, hvilke politiske og økonomiske interesser, kildens digitalisering repræsenterer.

Eksempelvis viser vores arbejde med Frihedsmuseets formål (at formidle valget om at gå ind i modstand) og Frihedsmuseets Fotoarkiv tilblivelse (digitalisering drevet af frivillige kræfter) at Frihedsmuseet har en klar ramme for, hvilke kilder der er relevante for museet at digitalisere. Udvælgelsen af relevant information til digitalisering er ifølge Pomerantz (2015) altid en subjektiv beslutning, og i Frihedsmuseets Fotoarkivs tilfælde er det tydeligt, at det i høj grad er de frivilliges egne interesser, der afgør, hvad der digitaliseres. Det er således vigtigt, at historikeren kan arbejde kildekritisk med de frivilliges subjektive motivation til at arbejde med digitalisering af det bestemte materiale.

Litteraturliste

Andersons, Benedict

1991 *Imagined Communities*, Verso, London, 1-17

“Available CRAN Packages By Name”, https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-T (set 13.maj 2025)

Chege, Kirstine

2024 “Flygtningekrisen i 2015”. *Folketinget*. (set 15 maj 2025):
<https://www.eu.dk/da/temaer/asyl-og-migration/asyl-og-flygtninge-i-eu/flygtningekrisen-2015>

Chege, Kirstine

2024 “Schengensamarbejdet og grænsekontrol”. *Folketinget*. (Set 15 maj 2025):
<https://www.eu.dk/da/temaer/asyl-og-migration/schengen>

Jensen, Helle Strandgaard

2020 “Digital Archival Literacy for (All) Historians” i *Media History* 27:2, 251-265

Lange, Lasse

2016 “Her er det nye politiske kompas”
<https://www.altinget.dk/christiansborg/artikel/det-nye-politiske-kompas> (set 15.05.2025)

Lauridsen, Gustav Aarup, Dalsgaard, Jacob Aarup & Svendsen, Lars Kjartan Bacher Svendsen

2019 *SENTIDA: A New Tool for Sentiment Analysis in Danish* i Journal of language works
vol. 4, no. 1, 38-53

“*Library Carpentry: Introduction to Regular Expressions*”,

<https://librarycarpentry.github.io/lc-data-intro/01-regular-expressions.html> (set 12.05.2025)

Nielsen, Morten

2018 “*Nyrup mente ikke, at DF var stuerene - i dag siger han, at 'hver tid har sine svar'*”

<https://nyheder.tv2.dk/politik/2018-09-22-nyrup-mente-ikke-at-df-var-stuerene-i-dag-siger-han-at-hver-tid-har-sine-svar> (set 15.05.2025)

Pomerantz, Jeffery

2015 *Metadata*, MIT Press, 2-18

Schriver, Astrid Ølgaard Christensen & Jensen, Helle Strandgaard

2022 *Arkivets digitalisering En ny udfordring til historisk metode?*, 5-27

Sosis, Richard & Alcorta, Candace

2003 “Signaling, Solidarity, and the Sacred: The Evolution of Religious Behavior” i
Evolutionary Anthropology 12, 264-274

Wiesner-Hanks et al.

2024 *A History of World Societies* 12. udgave, 995

Data

DANSK FOLKEPARTI: Kristian Thulesen Dahls tale ved DF's årsmøde 19. september 2015, Herning: <https://www.dansketaler.dk/tale/kristian-thulesen-dahls-tale-paa-dansk-folkepartis-aarsmoede-2015> (set 12. maj 2025)

DET KONSERVATIVE FOLKEPARTI: Søren Pape Poulsens tale ved Folketingets åbning, 6. oktober 2015, København: <https://www.altinget.dk/artikel/aabningsdebat-her-er-soeren-papes-tale> (set 12 maj 2025)

ENHEDSLISTEN : Johanne Schmidt-Nielsens 1. maj-tale 1. maj 2015, ukendt sted: <https://www.dansketaler.dk/tale/johanne-schmidt-nielsens-1-maj-tale-2015> (set 12 maj 2025)

RADIKALE VENSTRE: Morten Østergaards tale på Folkemødet, 11. juni 2015, Allinge: <https://www.dansketaler.dk/tale/morten-oestergaards-tale-til-folkemoedet-2015> (set 12 maj 2025)

SOCIALDEMOKRATIET: Mette Frederiksen tale ved Socialdemokraternes partikongres 19. september 2015, Aalborg: <https://www.dansketaler.dk/tale/mette-frederiksens-tale-til-socialdemokratiets-kongres-2015> (set 12 maj 2025)

SOCIALISTISK FOLKEPARTI (SF): Pia Olsen Dyhrs tale ved Socialistisk Folkepartis landsmøde 19. april 2015, København: <https://www.dansketaler.dk/tale/pia-olsen-dyhrs-tale-ved-sfs-landsmoede-2015> (set 12. maj 2025)

VENSTRE: Statsminister Lars Løkke Rasmussens tale v. Folketingets åbning, 6. oktober 2015, København: <https://stm.dk/statsministeren/taler/statsminister-lars-loekke-rasmussens-tale-ved-folketingets-aabning-den-6-oktober-2015/> (set 12. maj 2025)

Kongehuset

“Kongerækken”,

<https://www.kongehuset.dk/monarkiet-i-danmark/kongerakken> (set 12.05.2025)

Næstvedarkiverne

<https://www.naestvedarkiverne.dk/> (set 12.05.2025)

Nationalmuseet

<https://modstand.natmus.dk/> (set 12.05.2025)

Deklaration for anvendt Generativ Kunstig Intelligens (GAI)

– gældende version fra 20. maj 2025.

<p>Vi har benyttet generativ kunstig intelligens til udfærdigelse af dette projekt (sæt kryds).</p> <p><i>Oplist, hvilke GAI-værktøjer der er benyttet (husk version):</i></p> <p>X Microsofts Copilot, version GPT -4</p> <p>X OpenAIs ChatGPT, version-4o-model</p>			
<p>Vi har brugt GAI-værktøjer på følgende vis:</p> <table border="1"><tr><td><p>Til sparring ifm. valg af teori og metode X</p><p><i>Vi har brugt GAI til hjælp med at overskue lange teoritekster, så vi ikke læste alt for meget forgæves.</i></p><p><i>Desuden har vi brugt det metodisk i kodningsprocessen, se nedenfor.</i></p></td></tr><tr><td><p>Som en hjælp i læseprocessen X</p><p><i>Til at overskue store tekstmængder for at finde relevant litteratur bl.a. i final project og opgave U15, som er de opgaver, der indeholder mest teorilitteratur.</i></p></td></tr><tr><td><p>Til programmeringsopgaver og dataanalyse X</p><p><i>Vi har løbende brugt GAI som sparring i programmering i R: som hjælp til at forstå værdien af diverse programpakker, til forslag til nye koder, når vi var gået i stå og til at forstå fejlkoder.</i></p></td></tr></table>	<p>Til sparring ifm. valg af teori og metode X</p> <p><i>Vi har brugt GAI til hjælp med at overskue lange teoritekster, så vi ikke læste alt for meget forgæves.</i></p> <p><i>Desuden har vi brugt det metodisk i kodningsprocessen, se nedenfor.</i></p>	<p>Som en hjælp i læseprocessen X</p> <p><i>Til at overskue store tekstmængder for at finde relevant litteratur bl.a. i final project og opgave U15, som er de opgaver, der indeholder mest teorilitteratur.</i></p>	<p>Til programmeringsopgaver og dataanalyse X</p> <p><i>Vi har løbende brugt GAI som sparring i programmering i R: som hjælp til at forstå værdien af diverse programpakker, til forslag til nye koder, når vi var gået i stå og til at forstå fejlkoder.</i></p>
<p>Til sparring ifm. valg af teori og metode X</p> <p><i>Vi har brugt GAI til hjælp med at overskue lange teoritekster, så vi ikke læste alt for meget forgæves.</i></p> <p><i>Desuden har vi brugt det metodisk i kodningsprocessen, se nedenfor.</i></p>			
<p>Som en hjælp i læseprocessen X</p> <p><i>Til at overskue store tekstmængder for at finde relevant litteratur bl.a. i final project og opgave U15, som er de opgaver, der indeholder mest teorilitteratur.</i></p>			
<p>Til programmeringsopgaver og dataanalyse X</p> <p><i>Vi har løbende brugt GAI som sparring i programmering i R: som hjælp til at forstå værdien af diverse programpakker, til forslag til nye koder, når vi var gået i stå og til at forstå fejlkoder.</i></p>			

Metadata

Table 1 – Software metadata

Software metadata description	
Permanent link to Github repository where you put your script(s), R project, and data	https://github.com/AnnaLeth/DAM_eksamen_F25_.git

<i>Software License</i>	<ul style="list-style-type: none"> - Creative Commons Zero v1.0 Universal - R-studios, version 2024.12.1+563 (2024.12.1+563). - Voyant-Tools.org
<i>Data License</i>	<p>www.dansketaler.dk</p> <p>This is a website driven by the self-governing institution, Danske Taler. The purpose of the website is to make historical and speeches available for the public. This website is free to use and founded by the danish government and privat funds.</p> <p>www.altinget.dk</p> <p>This is a commercial online newspaper about Danish politics. In the assignment we online use the public articles, which do not need a subscription. This website is a part of Alrow Media-koncernen</p> <p>www.kongehuset.dk</p> <p>this website is owned by the Danish government to give information to the Danish people about the Royal family in Denmark today and the royal history of Denmark.</p> <p>www.naestvedarkiverne.dk</p> <p>This digital archive is owned by Næstved Kommune to inform the people of Næstved, scientist and the media about the story of Næstved. See more information about this metadata in <i>Uge 15</i> assignment.</p> <p>https://modstand.natmus.dk/</p> <p>this website is originally a gift from the Danish resistance movement to the Danish state. Today it is driven by the Danish state to give the people more information about the Danish resistance movement under WW2. See more information about this metadata in <i>Uge 15</i> assignment.</p>
<i>Software versions, Installation requirements & dependencies for software not used in class</i>	<i>You need to have ImageMagick 3.8 installed</i>
<i>Support email for questions</i>	<p>Christine.2410@icloud.com Au723140@uni.au.dk</p> <p>Helene.sk@hotmail.dk Au723949@uni.au.dk</p>

Table 2 – Data metadata

<i>Metadata description</i>	
<p>"DAM_eksamen_F25" → "sentimentAnalysis 3" → "FP_work" → "data_taler"</p> <ul style="list-style-type: none"> - Dansk_Folkeparti_DF.pdf - Det_Konservative_Folkeparti_K.pdf - Enhedslisten_EL.pdf - Radikale_Venstre_R.pdf - Socialdemokratiet_S.pdf - Socialistisk_Folkeparti_SF.pdf - Venstre_V.pdf 	<p>Pdf-files of seven danish political speeces from april-september 2015. All the speeces are transskribed.</p> <p>The files are named after the name of the political party of the speaker.</p>