

Speciale
Et kvantitativt perspektiv på en kvalitativ
undersøgelse

Martin Andreas Andersen*
Kandidatstuderende, pædagogisk sociologi
Aarhus Universitet[†]

11. Oktober 2020

*Stor tak til Katrine, der faciliterede at jeg fik arbejdsro til at fuldføre.

[†]Og tak til min vejleder, David Reimer, for god sparring undervejs.

Indhold

| | |
|--|------------|
| Indhold | i |
| Tabeller | ii |
| Figurer | iii |
| | |
| I Indledning | 4 |
| 1 Indledning | 5 |
| 1.1 At beskrive kontra at forklare | 5 |
| 2 Problemfelt | 7 |
| 2.1 Sortering til ungdomsuddannelserne i dag | 7 |
| 3 Problemformulering | 9 |
| 3.1 Undersøgelsesspørgsmål | 9 |
| 4 Specialets opbygning | 10 |
| | |
| II Baggrund og kontekst | 11 |
| 5 Markedskræfter og medborgerskab | 12 |
| 6 En gennemgang af forskningslandskabet | 13 |
| 6.1 (Ungdoms)uddannelse og social reproduktion | 13 |
| 6.2 Parlamentariske undersøgelser | 14 |
| 6.3 Algoritmisk sociologi | 15 |
| 7 Spændinger og tendenser i erhvervsuddannelserne i Danmark | 16 |
| 7.1 Et historisk oprids | 16 |
| 7.2 Spændinger og modsatrettede tendenser | 17 |
| | |
| III Metodiske og analytiske overvejelser | 19 |
| 8 Den algoritmiske sociologis muligheder og begrænsninger | 20 |
| 8.1 Hvad er computerbehandling af tekst? | 20 |
| 8.2 Begrænsninger og faldgruber | 21 |
| 8.3 Muligheder | 21 |
| 9 Dataindsamling og analyse | 23 |
| 9.1 Analyseredskaber og kode | 23 |
| 9.2 Forhåndsbearbejdning af data | 24 |

| | |
|---|-----------|
| 10 Analysestrategi | 26 |
| 10.1 Overordnet plan for analysen | 26 |
| 11 En politologisk undersøgelse i en pædagogisk-sociologisk kontekst | 30 |
| | |
| IV Analyse | 31 |
| 12 At beskrive den sociale virkelighed | 32 |
| 13 Indskrænkning af dokumenter til undersøgelse | 33 |
| 13.1 Talerne indbyrdes relationer | 36 |
| 14 Analyse af de udvalgte dokumenter | 38 |
| 14.1 Emner indenfor uddannelse | 38 |
| 14.2 Tendenser indenfor uddannelse | 38 |
| 14.3 Uddannelsesemner ligner hinanden | 40 |
| 15 På vej ned i detaljen | 43 |
| 15.1 Partipolitiske positioner omkring uddannelse | 43 |
| 15.2 Hvordan taler politikere om uddannelse? | 50 |
| | |
| V Konklusion | 60 |
| 16 Store data - store indsigter? | 61 |
| 16.1 Tekst som produkt - og som data | 61 |
| 16.2 Værdien af en beskrivende sociologi | 62 |
| 16.3 Beskrive - på baggrund af hvad? | 62 |
| 17 Fokus for videre forskning | 63 |
| 17.1 Metodologiske greb | 63 |
| 17.2 Andre fokusområder | 63 |
| 17.3 Anvendelighed på andre datasæt | 64 |
| 17.4 Indholdsanalysen er død. Længe leve indholdsanalysen! | 64 |
| | |
| Bibliografi | 65 |

Tabeller

| | |
|--|----|
| 15.1 Fordeling af partier i politiske blokke | 44 |
| 15.2 Oversigt over antal observationer for hver periode, i taler omhandlende uddannelse | 45 |
| 15.3 Oversigt over begrebsvægtning efter en <i>Wordfish</i> beregning af talerne i analyseperi- oderne. | 50 |

Figurer

| | | |
|-------|---|----|
| 13.1 | Beregning af optimalt antal emner for videre analyse; bigrams og ingen stopord. . . | 34 |
| 13.2 | Oversigt over udvalgte emner for perioden 2001 – 14, med tilhørende begreber. . . | 35 |
| 13.3 | LDA-modellens emner for perioden 2014 – 20, efter indbyrdes relationer. Emner repræsenteres med deres mest fremtrædende begreb. | 37 |
| 14.1 | Beregning af optimalt antal emner for dokumentsamlingen vedrørende uddannelse. . . | 39 |
| 14.2 | Oversigt over udvalgte emner indenfor uddannelse for perioden 2001 – 14, med tilhørende begreber. | 41 |
| 14.3 | Grendiagram over emner indenfor uddannelse. | 42 |
| 15.1 | Politiske blokkes politiske positioner over analyseperioderne | 45 |
| 15.2 | Politiske partiers politiske positioner over analyseperioderne | 46 |
| 15.3 | Blokkes politiske positioner over regeringsperioder fra 1978 til 2020 | 47 |
| 15.4 | Sammenligning af positioner omkring uddannelse i partiprogrammer og den seneste regeringsperiode. | 48 |
| 15.5 | Ordfordeling og vægtning for de politiske blokke, set som helhed over analyseperioder. | 49 |
| 15.6 | Antal positive og negative sætninger i partiernes taler | 51 |
| 15.7 | Taler fra (ungdoms)uddannelses korpus, spredning mellem sætningernes yderligste holdningsværdi. | 52 |
| 15.8 | Taler fra (ungdoms)uddannelses korpus, spredning mellem sætningernes gennemsnitlige holdningsværdi. | 55 |
| 15.9 | Taler fra (ungdoms)uddannelses korpus, spredning mellem talernes gennemsnitlige holdningsværdi. | 57 |
| 15.10 | Spredning over talernes gennemsnitlige holdningsniveau. Taler er organiseret efter mødedato, og er udvalgt på baggrund af min oversigt over politiske blokke. | 59 |

Resumé

Text mining educational policy discourse in a Danish context

In this thesis, I apply the tools of *computational sociology* to glean insights into the political discourse on educational policy in Denmark. Within the field of educational policy generally; I aim to focus on secondary education¹; and specifically the tension between the vocational and academic tracks in the Danish educational system. This thesis, in addition to introducing computational sociology into Danish educational sociology, also draws heavily on political science for perspectives and analytical techniques.

I am basing my research on transcripts from the floor of the Danish Parliament; from the fall of 1978 through the end of 2019. This is an extremely large corpus; even after trawling for relevant documents. The computer's capacity for inspecting and wrangling this large dataset is essential for the viability of my research. This is a particular strength in extending qualitative research with a computational approach (J. A. Evans & Aceves, 2016).

Inequality and inequity in Danish educational outcomes

A cohort study on the implications of tracking in Denmark indicates that there is a fair bit of correlation between socioeconomic status and lifelong earnings and which track in the educational system one is placed (Fels Birkelund m.fl., In progress). Furthermore, there is also a correlation on the socioeconomic status of ones parents and which track one ends up pursuing.

A general trend of higher general levels of education in Denmark over the past decades notwithstanding, the topic of equality and equity in educational outcomes remains a topic of intense political discussion.

Danish educational policy has had a stated goal of eliminating the so-called "remainder cohort"² since at least the 1990s; where then Minister for Education Ole Vig Jensen presented "Education for all" in 1993. This was further expanded to include a focus on higher education in the early 2000s. These policies seem to have had knock-on effects for the vocational educations. In 2016 the leaders of two of the largest unions in Denmark went so far as warning about Denmark becoming an "over-educated society"; projecting a severe shortage of skilled tradesmen.

Tensions in Danish secondary education

There is a tendency among Danes to devalue vocational education. According to Danmarks Statistik a scant majority had a positive impression of Danish vocational education; (Danmarks Statistik, 2019), with a 10% declining rate of enrollment after completed compulsory education from 2001 to 2013. The declining enrollment rate among vocational education is compounded by a markedly higher ratio of dropouts compared to the academic tracks (Danske Gymnasier, 2019).

In addition to the tensions between the academic and vocational tracks, the vocational educations in Denmark have their own internal tensions:

- a lack of apprenticeships available (Danmarks Statistik, 2019, s. 10); with remedial school education seen as less valuable

¹High school, more or less; what in Denmark is called "Gymnasium" or "Erhvervsuddannelse"

²The 8—or—so percent of a given youth cohort with only compulsory education and no gainful employment

- a change in the demographic among the vocational schools; as the more academically gifted youths gravitate toward the academic tracks, the vocational institutions are “saddled” with increasingly more academically—challenged students (Aarkrog, 2003, s. 365f)
- somewhat at odds with the previous, there is a long tradition for vocational education to be a path toward higher education (Bøndergaard, 2014, s. 47ff),

Knowledge gaps

I hope to gain insights into a broad array of policy dimensions; such as:

- period—specific discourse around (vocational) education policies
- per—party policy dimensions
- level of political agreement over time

Methodology — a quantitative perspective on qualitative research

I will be performing comparative time—series analysis on the parliamentary speeches, as well as sentiment analysis, where I move closer to selected texts. These are then fed through the unsupervised classification algorithm LDA (Blei m.fl., 2003); to discern speeches pertaining to educational policy. I follow up with applying the *Wordfish* algorithm by Slapin og Proksch (2008) to determine political positioning on a left—right axis; as well as sentiment analysis facilitated by the *SENTIDA* library for R (Lauridsen m.fl., 2019, 7. februar/2020).

Although I use methods that treats the text quantitatively; I maintain a largely qualitative focus. The parameters given and results obtained need to be evaluated by the researcher; and understood through their biases. The sentiment analysis, in particular, is supported with contextual reading of the surrounding text.

Insights gathered

Historical trends

The source material reflects the historical record as it is being written. As such, the generated topic models clearly reflect discussions of reforms and initiatives that have been influential upon Danish vocational and academic secondary education. The recognition of these historical trends also informs data extraction, and allows me to subset a corpora of over 700'000 parliamentary speeches to a much more manageable hundreds to low thousands range.

Party—wise policy dimensions

There is a clear right-to-left skew in the relative positions of both political parties and political blocs (as pertains to (secondary) education), mapping more or less to an instinctual assumption of these parties. Interestingly; graphing Danish political blocs to government periods shows quite a lot of variance over time. This may be reflective of a lack of sufficient data for s observation; however, I propose an alternate hypothesis for further investigation: To what extent is political positioning and posturing in Parliament in opposition to power, and how much is it a reflection of stated political principles? Controlling the most recent speeches against the current party programs, there is a rather large amount of variance, that also would warrant further investigation.

Politicians gonna politick

While doing my sentiment analysis, it became clear that politicians are playing a political game. Responding to political proposals with an exhortation of the impeccability of the policies of ones own party, cheap shots and sarcastic remarks: it is all fair game. However, it is also clear that politicians are not blind to the struggles and challenges facing the educational system; even if they do not always agree on framing paradigms and proposed solutions.

Going forward

This is a thesis firmly on the descriptive side of sociological research. I hope to inspire and inform further work utilizing computational sociology to wrangle the ever-ubiquitous large data that has become available in recent years.

A particularly interesting avenue for further research is investigating the connections between educational policy and other social issues. Speeches pertaining to affordable housing, adolescent crime and substance abuse, for instance, were consistently intertwined with topics related to education in my models.

Del I

Indledning

De unge stiller i dag store krav og forventninger om udviklingsmuligheder. Det er samfundets opgave at reagere imødekommende på disse behov. Hver enkelt af de unge skal have reelle muligheder for at opnå de bedst mulige uddannelsesmæssige forudsætninger for et rigt voksenliv. Det er i hele samfundets interesse at gøre en forstærket indsats. At efterlade en stor del af en ungdomsgeneration uden de nødvendige kvalifikationer er at lade dem i stikken i en fremtid, der stiller stadig større krav til uddannelse. En stor gruppe unge uden uddannelse er måske frem for alt et problem for demokratiet.

Undervisningsminister Ole Vig Jensen, Redegørelse til folketinget om uddannelse til alle, 11. november 1993

Vi skal have uddannelser i verdensklasse. Målet er, at eleverne i folkeskolen bliver blandt verdens bedste til læsning, matematik og naturfag. At mindst 95 pct. af de unge får en ungdomsuddannelse i 2015. Og at mindst 50 pct. af en ungdomsårgang får en videregående uddannelse i 2015.

Statsminister Anders Fogh Rasmussen, tale ved folketingets åbning, 24. februar 2005

1 Indledning

Hvorvidt der er fokus på, at klare sig i den internationale konkurrence; eller, at få "alle" med; så er der få politikere, der kan undlade, at have en holdning til uddannelse.

Daværende undervisningsminister Ole Vig Jensen funderede for eksempel over, hvorvidt de u(ud)dannede unge kan udgøre en trussel mod demokratiet. Dette er ikke en ny tanke. Siden 1970'erne har den såkaldte "restgruppe" vakt politisk bekymring (Hansen, 2004). Hvad gør man lige med den (forholdsvis stabile) andel af en ungdomsårgang, der ikke har kompetencegivende uddannelse, i et arbejdsmarked der har stadig mere rigide krav om formelle kvalifikationer (DPU, 2001)?

Vel et årti senere lover statsminister Anders Fogh Rasmussen uddannelse "i verdensklasse". "Restgruppeproblemet" er langt fra løst, men nu skal der oprustes til den globale konkurrence. Her skal Danmark hævde sig, og i bedste human capital stil, skal de unges potentialer forløses.

Ungdomsuddannelserne danner rammen for en af de vigtigste overgange i et menneskes liv. Man lægger barndommen fra sig i folkeskolen; og tager fat på det særligt moderne fænomen *ungdom* på vejen mod voksenlivet. Juul beskriver hvordan modernitetens ungdom er i refleksiv (re)konstruktion som individer. Samtidigt er ungdommen underlagt sociale strukturer, der begrænser deres reelle muligheder for at realisere et selvbiografisk projekt (2013, s. 11).

I „Diskurser om ungdom og arbejdsliv – fra velfærdsstat til konkurrencestat“ beskæftiger Juul sig med ungdomspolitik i det seneste århundrede, og de diskurser om ungdomslivet man kan udlede heraf. Hun henviser bl.a til at man i 70'erne havde fokus på, hvordan statens institutioner kunne have indflydelse på de unges liv; hvor fokuset i 90'erne var flyttet over mod overgangen til arbejdslivet. Eller, en sociologisk inspireret ungdomsdiskurs bliver underordnet en (samfunds)økonomisk forståelsesramme.

Jeg vil i dette speciale grave dybere ned i ungdomsuddannelserne og de politiske holdninger herfor, og hvordan disse eventuelt har ændret karakter over tid siden 1978. I min undersøgelse af ungdomsuddannelserne, vil jeg have et særligt fokus på erhvervsuddannelserne, og hvordan disse vægtes i forhold til de akademisk rettede ungdomsuddannelser.

For at kunne undersøge dette felt, vil jeg tage referater fra møderne i folketingsalen i brug. Da dette er et overvældende stort kildemateriale¹, at analysere og klassificere for egen hånd, vil jeg tage computerbaserede tekstanalyseværktøjer i brug. Dette speciale vil dermed også fremstå som en afprøvning af tekstmining og lignende analyseværktøjer anvendelighed i dansk pædagogisk-sociologisk sammenhæng. I det jeg undersøger det politiske felt, vil jeg også drage inspiration fra den politiske sociologi i forhold til analytiske metodologiske og perspektiver.

1.1 At beskrive kontra at forklare

Dette speciale lægger sig helt explicit i en beskrivende sociologisk tradition. Jeg håber, at kunne påvise nogle mønstre i dansk uddannelsespolitik, men jeg vil ikke umiddelbart efterstræbe en forklaringsmodel; eller forsøge at påvise kausalitet ud fra mine observationer². Mit formål er snarere, at bruge et nyt perspektiv til at udforske et forholdsvist velkendt landskab, og udpege

¹ Omkring 740.000 folketingstaler som udgangspunkt

² Implicit i denne vinkling er et ontologisk og epistemologisk perspektiv, der tilsiger at der er en social verden at beskrive – og at en sådan beskrivelse kan afspejle en social virkelighed

nye og spændende landemærker. Dermed er det gjort klar, til at den næste opdagelsesrejsende går lige på en mere dybdegående undersøgelse.

Der er dog en vis spænding mellem det deskriptive og det kausale. Allerede i 1951 så Sjøberg det nødvendigt, at tage den beskrivende sociologi i forsvar (Sjøberg, 1951). En renskåret afgrænsning mellem beskrivelse og forklaring kan endda anfægtes; i det dybdegående og systematiske beskrivelser vil kunne have en forklarende kraft (Sjøberg, 1951, s. 252).

I nærmere fortid kan nævnes Savage, der i 2009 argumenterede for, at det deskriptive er en central del af megen moderne sociologi, bredt antaget (Savage, 2009). Dette især, hvor det empiriske materiale sociologer beskæftiger sig med ikke nødvendigvis opstår som del af sirligt planlagte spørgeskemaer eller længerevarende feltstudier (Savage, 2009, s. 157) — som for eksempel er gældende for dette speciale.

Jeg orienterer mig både indenfor den pædagogiske sociologi, den politiske sociologi og en nyere "algoritmisk" sociologisk tilgang. Både Sjøberg og Savage ser en værdi i, at beskrivelser kan indeholdes i flere metodologiske referencerammer (Sjøberg, 1951, s. 256) eller endda opfordrer til en gentænkning af klassiske kategoriskeringer af moderne sociologisk gerning (Savage, 2009, s. 170). Jeg vil bestræbe mig på, at agere brobygger mellem politologien og den pædagogiske sociologi. I mit arbejde i dette krydsfelt trækker jeg kraftigt på politisk forskning, både for inspiration til dataindsamling og undersøgelsesfelt, og også som en kilde til metodologiske værktøj. Savage mener videre at se, at denne nye dragning til det deskriptive fører til en orientering mod det visuelle når resultater skal formidles (Savage, 2009, s.169). I arbejdet med at præsentere mine data og metoder, kan jeg anekdotisk tilføje, at denne observation holder for denne opgaves vedkommende.

2 Problemfelt

De indledende citater for specialets befinder sig begge i et overordnet neoliberalt paradigme for samfundets opbygning; om end med forskelligt fokus. Det er antaget skadeligt for hele samfundet, hvis vi ikke alle har mulighed til, at yde deres ypperste.

Der ligger dog mere under overfladen, når man taler om "uddannelse" end en maskine, hvor man kan skrue op for "integration" og "innovation". Det (post)moderne samfund er tiltagende specialiseret og arbejdsdelt (Bauman, 2000). Ud fra dette perspektiv giver dette god mening, at der skal etableres nogle rammer for det selvrealiseringsprojekt der blev henvist til af Juul ovenfor. Vi kan ikke alle være historikere, læger eller sociologer — samfundet har også behov for pædagoger, salgsassistenter og murere. Uddannelsessystemet fungerer i denne sammenhæng de facto som en sorteringsmekanisme; der sluser de unge mod de funktioner, de kan tjene samfundet i.

Hvis uddannelsessystemet har som funktion, at sortere borgermassen i differentierede arbejdsområder, vil der også være elementer af, en legitimering af strukturelle forskelle. Meget af sproget i og omkring ungdomsuddannelserne handler om, at imødekomme de unges (antagne) forudsætninger og interesser. En opdeling af de unge i forskellige spor — på engelsk "tracking" — er ofte set som en vej til, at opnå disse (formodede) fordele. Dermed kan de "bogligt interesserede" søge mod de akademiske ungdomsuddannelser; og de mere "praktisk orienterede" har muligheder indenfor erhvervsuddannelserne. Dog viser internationale undersøgelser, at tildeling til de forskellige spor i vid udstrækning er stærkt associeret med øvrige klasseskæl i samfundet; og at elever i højere spor øger deres forspring set i et livslangt perspektiv (Gamoran, 2010, s. 3).

Denne grundlæggende forskelsbehandling i praksis er dog skjult bag en forståelse af Danmark som et forholdsvist fladt samfund — hvor man har meget få begrænsninger i udarbejdelsen af sin selvbiografi. "Alle" har (principielt) mulighed for at blive "alt", til trods for, at der dokumenterede forskelle i uddannelsesudfald på baggrund af socioøkonomisk udgangspunkt (Munk, 2014, se for eksempel s 195ff i). Som Munk beskriver, er uddannelsesmobiliteten godt nok steget i de seneste år. Familiebaggrund betyder nu mindre end før, men viser sig blandt andet i hvilke videregående uddannelser unge fra (historisk) mindre privilegerede samfundsgrupper søger mod (2014, s. 199).

2.1 Sortering til ungdomsuddannelserne i dag

Der er en faldende tilslutning til erhvervsuddannelserne blandt de unge. I følge Danmarks Statistik, var antallet elever på erhvervsuddannelserne i oktober 2018 det laveste siden 1990erne (Danmarks Statistik, 2019, s. 10). Tal fra Børne- og undervisningsministeriet viser, at andelen unge, der søger erhvervsuddannelser efter endt grundskole er faldet jævnt fra godt 31 procent i 2001, til knap 20 procent 2013, hvor den har været nogenlunde stabil (Børne- og undervisningsministeriet, nodate, s. 6). Der har været en tilsvarende stigning i ansøgere til de gymnasiale uddannelser, der så en stigning fra knap 60 procent til omkring 74 procent i samme periode (Undervisningsministeriet, 2017, s 5f). Holdt sammen med, at frafaldet er markant højere på erhvervsuddannelserne end de gymnasiale uddannelser (Danske Gymnasier, 2019), er politiske målsætninger om, at flere gennemfører en ungdomsuddannelse tilsyneladende langt fra at være opnåede. Der er blandt andet bekymringer om, man kan samle "restgruppen" op. Andelen unge uden kompetencegivende uddannelse eller tilknytning til arbejdslivet har således ligget stabilt på omkring 7–8 procent i perioden 2012–2017 (T. B. Andersen, 2019, s. 9).

Qua de strukturelle uligheder der er nævnt tidligere, er der også forskelle i prestige og social agtelse i forskellige uddannelser. Denne manglende sociale agtelse underbygges af en opgørelse fra Danmarks Statistik, hvor kun 60 procent af befolkningen finder, at de danske erhvervsuddannelser har et positivt omdømme (2019, s. 7).

Det er til gengæld heller ikke alle, der nødvendigvis ser frem til 5–10 års yderligere skolegang efter grundskolen, hvor en mere direkte vej til arbejdsmarkedet er oplagt. Men også her er der en social skævvridning. Som samme opgørelse fra Danmarks Statistik viser, har for eksempel 19 procent af eleverne på erhvervsuddannelse forældre med en årlig bruttoindkomst på under 300 000, modsat kun 6 procent af eleverne på de gymnasiale uddannelser (2019, s. 6).

Denne forskel i social agtelse kan også ses, i hvordan formålsparagrafferne for tre forskellige ungdomsuddannelser italesætter de unge. Ved at se på erhvervsgrunduddannelsen (Uddannelsesministeriet, 2016, § 1), erhvervsuddannelserne (Uddannelsesministeriet, 2020, §1, stk.2) og de gymnasiale uddannelser (Uddannelsesministeriet, 2019, §1) viser, særskilte forestillinger af forskellige slags ungdom.

EGU-eleven skal opnå (eller sågar tilføres) kvaliteter, denne dermed må have som mangelvare; EUD-eleven skal understøttes i, at følge sit eget initiativ; og gymnasie-eleven fremstår enestående som demokratisk samfundsborger *in spe* — der skal blot lidt understøttet (selv)udvikling til. Disse fremstillinger af dansk ungdom skal henholdsvis i uddannelse eller i arbejde; uddannes til arbejdsmarkedet; og andre igen skal forberedes til universitetet.

Hvordan man helt præcis stiller sorteringsmekanismerne i uddannelse ind, kan man (blive ved med) at skrive meget om:

- Hvordan kan det være, at restgruppen ikke er blevet betydeligt minimeret de seneste år, trods ihærdige initiativer?
- Kan den meget tydelige kønssegregation i Danmarks uddannelsessystem og arbejdsmarked imødekommes¹.
- Hvordan forholder de unge sig til valget af ungdomsuddannelse?
- Er Danmark ved at være overuddannet, som formændene for Dansk Metal og HK udtalte for nogle år siden? (Simonsen & Jensen, 2016)

En dybdegående analyse af “Uddannelse til alle” i 90erne og fremefter ville forsåvidt være grundlag for et speciale i sig selv.

Jeg er dog mere interesseret i, politikeres udtalelser omkring uddannelse. Hvordan har de folkevalgte italesat deres uddannelsespolitiske positioner og holdninger gennem tiden? Til mit held, er referater fra møder i folketingsalen offentligt tilgængelige, til frit brug. Her kan man se hvordan politikere henvender sig til andre politikere — samtidig med, at de også taler i en semi offentlig kontekst. Disse tekster vil jeg foretage computerbaserede tekstanalyser på. Computerbaserede analyser er langt fra ufejlbarlige, men det er dog meget nemmere for en computer at se på over 700,000 taler end det er for mig at læse dem en efter en.

¹ Dette ser man i øvrigt også ved ansøgning til erhvervsuddannelserne — drengene søger overordnet ind på uddannelser som byggeri og teknik, og pigerne søger overvejende ind på sundhed og service (Danmarks Statistik, 2019, s. 48)

3 Problemformulering

Jeg vil i dette speciale se nærmere på, hvordan politikere har italesat ungdomsuddannelse i folketingsalen. Herunder især den politiske diskurs omkring de erhvervsrettede ungdomsuddannelser. Jeg vil også berøre de gymnasiale uddannelser, og hvordan debatten mellem “arbejdsmarkedsrelaterede kompetencer” og såkaldt “overuddannelse” arter sig. Min undersøgelse vil tage udgangspunkt i computeranalyser af politiske taler ved møder i folketingsalen fra 30. august 1978 til i dag.

Valget af 1978 som udgangspunkt er dels på baggrund af datatilgængelighed. Men 1978 er også anvendelig som et nedslagspunkt i dansk ungdomspolitik og uddannelseshistorie. U90 blev offentliggjort i 1978, med visioner om hvordan dansk uddannelse skulle se ud mod 1990 (Undervisningsministeriet, 1978) Derudover blev den erhvervsfaglige grunduddannelse vedtaget nationalt i 1977 (The Wikipedia Volunteers, 2019), og jeg kan dermed få diskurserne efterfølgende taget med. Den specifikke dato — 30. august 1978 — er den første dag for regeringen Anker Jørgensen III. Ved at holde mig til regeringsperioder, kan jeg holde nogle analysedimensioner åbne — er taleren i regering eller i opposition, for eksempel? — hvis dette skulle være relevant.

3.1 Undersøgelsesspørgsmål

- Hvordan genspejles tidens ånd i omtalen af (erhvervs)uddannelse for specifikke perioder?
- Kan de politiske udtalelser omkring uddannelse gøres op efter partipolitiske linjer?
- Hvor optræder der eventuel politisk enighed eller uenighed?
- Er der overensstemmelse mellem partiernes udtalelser i deres partiprogrammer og politikernes folketingstaler?
- Hvem ser politikerne som målgruppen for de erhvervsfaglige uddannelser?
- Hvordan positionerer politikerne sig og andre i kraft af deres sprogbrug?
- Hvordan forholder politikere sig til strukturelle forskelle og uligheder?

Disse spørgsmål bevæger sig fra det forholdsvis konkrete til det mere abstrakte. Jeg er noget nysgerrig omkring hvorvidt de redskaber jeg vil tage i brug kan bidrage til indsigter på det mere abstrakte niveau.

4 Specialets opbygning

I denne første del af specialet har jeg præsenteret mit undersøgelsesobjekt og dets relevans i en pædagogisk sociologisk kontekst, med en yderligere præsentation af min undersøgelsestilgang og mine underspørgsmål.

I specialets næste del vil jeg bestræbe mig på, at give baggrund og kontekst for min analyse. Jeg vil gennemgå et uddrag af senere tids forskning i ungdoms uddannelser, med et særligt øje for erhvervsuddannelserne. For erfaringer med computeranalyser af parlamentarisk materiale, drager jeg på et overvejende politisk–sociologisk forskningsarbejde. For at have en historisk kvalifikation for min analyse, og min inddeling i perioder til sammenligning, agter jeg derudover at give et kort oprids over tendenser i dansk uddannelses politik de seneste 4 årtier. Også her vil jeg særligt skele til ungdomsuddannelserne, herunder erhvervsuddannelserne.

I del III vil jeg drøfte mine metodiske og teoretiske overvejelser. Styrker og svagheder ved computerbaserede tekstanalyser vil blive gennemgået. Jeg vil derefter dykke ned i mine konkrete metoder og overvejelser set i forhold til det gældende dataset og mine problemstillinger.

En gennemgang af min analyse vil være fokus for del IV. Jeg vender tilbage til mine undersøgelsesspørgsmål, og vil demonstrere mit arbejde med at besvare disse. Her vil jeg illustrere og beskrive mine resultater (eller mangel på samme), og hvordan jeg er kommet frem til disse. Jeg vil også kort påpege uventede resultater og mulige veje fremad for videre forskning, efterhånden som det bliver relevant.

Jeg afslutter specialet i del V, med en diskussion af mine resultater, og hvad betydningen kan være for vores forståelse af uddannelsespolitik og ungdomsuddannelser. Jeg vil også drøfte implikationer for fremtidige undersøgelser i spændingsfeltet politisk og pædagogisk sociologi. På et overordnet niveau vil jeg også reflektere over brugen af *algoritmisk sociologi* som metodisk tilgang, og hvordan denne tilgang arter sig som et kvantitativt perspektiv på en kvalitativ problemstilling.

Del II

Baggrund og kontekst

Danmark mangler allerede nu kvalificeret arbejdskraft, og manglen vil vokse markant de kommende år. Tal fra Arbejderbevægelsens Erhvervsråd viser, at der de kommende ti år vil komme til at mangle omkring 70.000 faglærte. Alene jern- og metalindustriend vil mangle 30.000 faglærte, og tæller vi handels- og kontorområdet med, kan vi lægge yderligere 12.000 oven i. Udover faglærte vil der komme til at mangle flere end 60.000 personer med korte og mellemlange videregående uddannelser. Samtidig vil omkring 55.000 personer med en lang videregående uddannelse være i overskud i år 2030.

Claus Jensen, forbundsformand, Dansk Metal og Kim Simonsen, forbundsformand, HK, kronik i Berlingske Tidende, 3. april 2016

5 Markedskræfter og medborgerskab

Det er ikke kun politikere, der har noget de skulle sagt, om hvordan samfundets udvælgelsesprocesser og prioriteringsmekanismer er skruet sammen. Om man taler om en "overuddannet" befolkning, eller italesætter arbejdsløse akademikere som "dovne", skorter der ikke med skarp retorik omkring hvordan man aktuelt har indstillet sorteringsmaskinen.

Nu er dette næppe noget banebrydende. Uddannelse har en bred berøringsflade i samfundet, og påkalder sig dermed bred opmærksomhed. Og uddannelse af "ungdommen" — disse voksne under tilblivelse — især. Diskursen omkring ungdommens uddannelse er heller ikke en statisk størrelse, men ændrer sig med tiden og tidens ånd. I efterkrigstiden var fokus fx, at sikre de unge fra vilkårlighed ved at informere og vejlede (Juul, 2013, s 17ff). I 70erne ses uddannelse som en samfundsmæssig udligner, der ikke blot skal tjene erhvervslivet, men også hæve befolkningens kvalifikationer ud over det boglige (Juul, 2013, s 15ff). Og fra 90-tallet ser vi et fokus på uddannelse som vejen til at individet kan opfylde sin samfundsmæssige forpligtelser på arbejdsmarkedet (Juul, 2013, s. 18).

Disse skiftende holdninger til uddannelse vil jeg uddybe i denne del af specialet, for at kunne kontekstualisere de computerdrevne analyser jeg vil udarbejde.

Jeg starter med en kort gennemgang af nyere forskning, der berører mit emne og forskningsfelt. Derefter vil jeg drøfte diskursive tendenser i dansk (ungdoms)uddannelsespolitik fra 1970erne til i dag, med særligt fokus på erhvervsuddannelserne.

6 En gennemgang af forskningslandskabet

Jeg vil i det følgende gennemgå nyere forskning indenfor erhvervsuddannelserne specifikt, men også ungdomsuddannelser generelt. Jeg vil derefter kort nævne arbejde indenfor parlamentarisk forskning. Kapitlet afsluttes med udlægninger af, hvordan tekstmining kan anvendes i et sociologisk perspektiv.

6.1 (Ungdoms)uddannelse og social reproduktion

I det følgende præsenteres noget arbejde omkring erhvervs- og ungdomsuddannelserne. Jeg holder mig primært til forskning i dansk kontekst i forhold til erhvervsuddannelserne. Der er lavet nogle komparative analyser af dansk EUD i både skandinavisk og europæisk perspektiv, men dette forekommer mig at falde udenfor specialets fokus. Der er dog noget arbejde der ser på sorteringsmekanismerne i uddannelsespolitik jeg vil nævne.

Juul beskriver erhvervsuddannelserne som underundersøgte i dansk sammenhæng (Juul, 2004). Hun har forsøgt at afhjælpe dette, og har forsket i de danske erhvervsuddannelser i omkring 20 år („Ida Juul - Publikationer - Forskning - Aarhus Universitet“, nodate). Som eksempel kan nævnes „Diskurser om ungdom og arbejdsliv – fra velfærdsstat til konkurrencestat“, der tager et historisk blik på de diskurser om ungdomsliv der afspejles i policy-tekster om ungdomsuddannelser. Til forskel fra mit perspektiv, ser Juul på teksterne som konsensuspapirer, og foretager en dokumentanalyse af disse (Juul, 2013). Hvor Juul specifikt ikke ser på hvordan denne type konsensuspapirer opstår, er det netop dette jeg ønsker at undersøge.

I temanummeret *Erhvervsuddannelser - mellem politiske ambitioner og praktisk virkelighed* af Dansk pædagogisk tidsskrift drøftes virkningerne af reformen af erhvervsuddannelserne i 2014. Denne sættes i forhold til en generel reformiver omkring uddannelser i dag, hvor også tidsintervallerne mellem reformerne forkortes (Canger m.fl., 2016, s. 2).

Jeg hæfter mig især ved artiklen „Reformen af erhvervsuddannelserne“, hvor Jørgensen påpeger, at reformen af 2014 i høj grad handler om at imødekomme utilsigtede konsekvenser af tidligere reformer. De utilsigtede konsekvenser omkring erhvervsuddannelserne er mange, da disse i senere tid har haft berøringsflader til sociale såvel som arbejdsmæssige politiske arenaer.

Forskningsprojektet „Nord-VET“ ser på historiske kontekster for nuværende og fremtidige udfordringer ved erhvervsuddannelser i de nordiske lande. Projektet har især fokus på erhvervsuddannelsernes tiltagende dobbeltrolle, hvor de samtidigt med at uddanne til arbejdsmarkedet, også er adgangsgivende for højere uddannelse.

Der undersøges en del omkring de unges forhold til ungdomsuddannelserne. CFU, VIVE, EVA og Undervisningsministeriet foretager jævnligt undersøgelser omkring hvilke ungdomsuddannelser de unge vælger, og hvorfor (se fx. Børne- og undervisningsministeriet, nodate; Danmarks Statistik, 2019; Undervisningsministeriet, 2017). Dette har muligvis en nytteværdi for beslutningstagere, men mit fokus er ikke på de unges grundlag for at vælge uddannelser; men politikeres omtale af uddannelserne.

Jeg har i indledningen refereret til en opgørelse lavet af Danmarks Statistik. Denne udmærker sig ved at give en kort historisk kontekst for fortolkning af tallene (2019, s. 8ff). Der understreges også, at eleverne på erhvervsuddannelserne ikke nødvendigvis er “unge”. Gennemsnitsalderen for elever på erhvervsuddannelserne er knap 24 år, med variationer indenfor de forskellige uddannelser (Danmarks Statistik, 2019, s. 14).

Chanceulighed og social reproduktion er der set en del på i dansk kontekst.

Jeg henviste i specialets første del til Munk, der beskriver Danmark som et land med overordnet positiv udvikling; men at forældres baggrund stadigvæk er indikerende for, hvilke uddannelser man tager (Munk, 2014). Dette underbygges af Fels Birkelund m.fl., der ser på tracking i en dansk kontekst. Ved at gennemgå danske registerdata, viser de at de internationale erfaringer i også er gældende i Danmark. Er ens forældre højtuddannede, er det meget større sandsynlighed for, at blive universitetsstuderende, og man har også højere social status og disponibel indkomst over tid (Fels Birkelund m.fl., In progress).

I rapporten *Uddannelsesmobilitet i Danmark*, udgivet af SFI, viser det sig også, at på trods en generel stigning i uddannelsesmobilitet, er der stadig store forskelle i hvem der tager hvilke uddannelser.

For en drøftelse af sorteringsmekanismer til uddannelse, har jeg tidligere refereret til *The Structure, Causes, and Consequences of Tracking in the Danish Education System*; hvor Fels Birkelund m.fl. forsøger at afdække korrelationer mellem sociale klasseforskelle og uddannelsesudfald i en dansk kontekst. Det er et særligt fokus på de langsigtede konsekvenser af hvilket uddannelsesspor man følger, med udgangspunkt i årgangen født i 1975. De finder, at der er klart flertal af studerende med akademisk uddannede forældre der selv gennemfører en akademisk uddannelse. Videre viser det sig, at børn hvis forældre har en ungdomsuddannelse, i markant større grad gennemfører en videregående uddannelse end børn der kun har forældre med obligatorisk skolegang. Dog er der omtrentlig lige stor andel af unge fra begge disse grupperinger der gennemfører en ungdomsuddannelse (Fels Birkelund m.fl., In progress, s 12). Den langsigtede analyse tyder også på, at opnåelse af en erhvervsuddannelse i høj grad er medvirkende til bedre arbejdsmarkeds- og uddannelsesudfald (Fels Birkelund m.fl., In progress, s. 24).

Det tydelige sammenfald mellem socioøkonomisk baggrund og sortering til specifikke uddannelsesspor er underbygget i en tysk kontekst af Schnepf, der ser samme variation i optagelse ved tyske uddannelsesspor; med stor ulighed i uddannelsesudfald til følge; forstærket af den tidlige fordeling af uddannelsesspor i Tyskland (2002).

6.2 Parlamentariske undersøgelser

En stor del af inspirationen til min forskningsinteresse kommer fra en præsentation ved IMC, Århus Universitet (Interacting Minds Centre, Aarhus University, 2019). Jan Kostkan og Malte Lau Pedersen præsenterede analyser af ideers nyhedsværdi i en tjekkisk parlamentarisk kontekst¹.

Flere undersøgelser i dansk regi der foretager en massebehandling af folketingstaler har dog været svære at finde. Et eksempel er Hjorth, der i „Establishment Responses to Populist Challenges: Evidence from Legislative Speech“ undersøger hvordan danske politikere forholdt sig til, at Dansk Folkeparti blev en del af det politiske landskab. Med udgangspunkt i en tekst-som-data tilgang blev folketingstaler fra 1997 til 2004 analyseret ud fra en superviseret machine learning metodologi² (Hjorth, 2018).

Der er til gengæld foretaget en del indholdsanalyser af specifikke emner i dansk regi.

I „Tænk Ikke På En Ghetto“ (2012) viser M. Andersen og Madsen hvordan statsminister Løkke Rasmussens udtalelser omkring „ghettoer“ er medvirkende til, at skabe et verdens- og selvbillede for beboere i de omtalte områder. Deres analyse tager udgangspunkt i en nærlæsning af enkelte af statsministerens taler.

Internationalt er der dog foretaget flere undersøgelser, hvor flere samtidigt undersøger og udvikler de metodiske redskaber.

¹ Det var også her jeg blev bevidst, at folketingstalerne var tilgængelige for analyse

² Jeg uddyber de tekniske begreber i mit metodeafsnit

I „Position Taking in European Parliament Speeches“ ser forfatterne på politiske positioner blandt medlemmerne af Europaparlamentet. Ved hjælp af ordfrekvensanalyser udledes der niveauer af politisk uenighed, der forsøges gjort op på en højre/venstre politisk akse. Det viser sig dog, at de politiske konflikter i større grad er baseret på EU-samarbejdets karakter end over tilhørighed til bestemte politiske fløje (Proksch & Slapin, 2010).

I „Extracting Policy Positions from Political Texts Using Words as Data“ undersøger Laver m.fl. muligheder for at benytte tekst-som-data som en tilgang til at udlede politiske positioner. De ser endvidere på mulighederne i, at foretage computerbaserede kodninger af tekstmaterialet kontra kodning via mennesketimer, for dermed at gøre fremtidigt analysearbejde mindre omstændigt og tidkrævende (Laver m.fl., 2003).

„How to Analyze Political Attention with Minimal Assumptions and Costs“ er en metodologisk beskrivelse af forfatternes arbejde ved at udvikle en *topic model* for deres analyse af taler i det amerikanske senatet. Deres tilgang udleder både de ord, der er indikerende for specifikke emner; samt antal af emner i en samling af tekst (Quinn m.fl., 2010).

Slapin og Proksch beskriver en måde, hvorpå man kan arbejde med (parti)politiske positioner ved computeranalyser af tekst. De ser på partiprogrammer fra Tyskland i perioden 1990–2005, og kortlægger partiernes politiske positioner over tid. Deres tilgang adskiller sig fra Laver m.fl. idet, det ikke er nødvendigt at have et set veldefinerede tekster at beregne referenceværdier ud fra (Slapin & Proksch, 2008).

Der er også historiske undersøgelser der gør brug af computeranalyser af tekst. For eksempel har Barron m.fl. undersøgt, hvordan det første franske parlament efter den franske revolution forholdt sig til nye ideer. Gennem analysen viser de, at de store talere fanget folket; men meget af den faktiske indflydelse på lovgivningen faldt til dem, der mestrede komite-arbejdet i de politiske baglokaler (Barron m.fl., 2018).

6.3 Algoritmisk sociologi

De ovenstående eksempler drager i høj grad på computerbaserede undersøgelsesmetoder; men der findes også arbejde, der ser på den algoritmiske sociologi mere generelt.

J. A. Evans og Aceves, 2016 beskriver, hvad computerbaserede tekstanalyser kan tilføre sociologien, med et kvalitativt perspektiv. De ser på både superviserede og ikke-superviserede tilgange til computerbaseret tekstanalyse, i tre sociale dimensioner der kan undersøges ud fra tekstmaterialer.

Molina og Garip, 2019 giver et bredere overblik, over hvad såkaldt “machine learning” kan bidrage til i den sociologiske forskning. Der er særligt fokus på den kvantitative sociologi, og hvordan man både kan bruge computeranalyser til at krydsvalidere resultater, og til at udforske data for at udarbejde nye hypoteser.

I „Computation and the Sociological Imagination“ henviser forfatterne til, at den algoritmiske sociologi kun er i vækst; og det nu er mere aktuelt end nogen sinde, at også undervise i R og Stan på sociologiske uddannelser, for at kunne gribe de nye muligheder til at udforske den sociologiske fantasi (J. Evans & Foster, 2019).

7 Spændinger og tendenser i erhvervsuddannelserne i Danmark

Med et billede af den senere forskning jeg bygger på i dette speciale, vil jeg nu ridse op hvordan vi kom frem til der, hvor vi er i dag. Da mit fokus for specialet er erhvervsuddannelserne set i forhold til de gymnasiale uddannelser, vil denne udlægning også tage udgangspunkt i erhvervsuddannelserne. Derefter vil jeg beskrive nogle af interne spændinger der har præget — og stadig præger — erhvervsuddannelserne.

7.1 Et historisk oprids

Det følgende vil tage afsæt i Bøndergaard, 2014 og Juul, 2013 for en grov inddeling af perioder indenfor dansk erhvervsuddannelse.

Jeg starter med en (meget) overordnet gennemgang af tendenserne op til hvor min analyse starter; for derefter at gå mere i dybden med mine analyseperioder.

Historien før min analyse

Grundlaget for den erhvervsuddannelse vi ser i dag bliver lagt allerede før 1945, med et bredt samarbejde mellem arbejdstagere, arbejdsgivere og staten. De tekniske skoler en stor vækst, men gymnasierne er forbeholdt de mere privilegerede unge (Bøndergaard, 2014, s. 9). Det er også her, en formel skolebaseret undervisning i stigende grad bliver en del af erhvervsuddannelsen (Bøndergaard, 2014, s. 15ff).

Med store ungdomsårgange efter 2. verdenskrig, samt en stor vækst indenfor industriel produktion, er der nu en delvis svækkelse af tidligere restriktioner for antal lærlinge. Det er også nu vi ser dagskoler blive en integreret og obligatorisk del af undervisningen, der også bliver mere specialiseret. Undervisningen ivaretages også i højere grad af centralstyrede statslige skoler (Bøndergaard, 2014, s. 35).

Nu er "ungdom" også et fænomen der i mindre grad forbeholdes de mere velhavende samfundslag. Dermed ses det også som en vigtig politisk opgave at vejlede de blivende borgere i, at foretage fornuftige — forstået som i tråd med middelklassens værdier — valg, både kort- og langsigtet (Juul, 2013, s. 13ff).

1978 — 1990: Uddannelse i fokus

Grundlaget for hvad der bliver til den erhvervsfaglige grunduddannelse (EFG), lægges allerede fra sidst i 60'erne. Den implementeres i den anden halvdel af 70erne. Udarbejdet som en afløser til mesterlæren, hvor indgangene til de erhvervsrettede ungdomsuddannelser tilrettelægges under et samlet grundforløb, ender EFG dog med at køre sideløbende med den. Dette afspejler en grundlæggende konflikt omkring udformningen af erhvervsuddannelserne, mellem en mere skole-baseret og tilnærmet gymnasierne; og en mere praksis-orienteret tilgang (Bøndergaard, 2014, s. 49ff, 57).

I 1978, hvor jeg starter mine analyser, udgives *U 90: samlet uddannelsesplanlægning frem til 90'erne*. der ser for sig et samfund hvor uddannelse får en voksende betydning (Undervisningsministeriet, 1978). Denne rapport kommer også med kritik overfor erhvervslivets store indflydelse

på uddannelserne. Der bør være et mere helhedlig fokus, understreges der; med mere almene kvalifikationer fra uddannelse — og et generelt højere uddannelsesniveau i samfundet (Juul, 2013, s. 18f). Hovedtanken i den uddannelsespolitik der blev fremført af den socialdemokratiske regering, var at sikre lighed gennem uddannelse.

1990 — 2001: Målstyring og selvstyre, under markedsvilkår

Erhvervsuddannelserne skifter karakter i løbet af 90'erne. I 1990 udgives *U91: det nye mønster i dansk uddannelses- og forskningspolitik*, der er et kraftigt opgør med tankerne fra U90. I 1991 trådte der en EUD-reform i kraft, der afskaffede mesterlæren, gjorde erhvervsuddannelserne selvejede, indførte taxameterordninger og andre markedsorienterede tiltag. Dette for at komme væk fra central detaljstyring, der forhindrer nyskabelse og innovation. Uddannelsens primære fokus var også, at føre de unge hen til arbejdsmarkedet (Juul, 2013, s. 19).

2001 — 2014: Ansvar for egen læring, med individet i fokus

Nu er uddannelse vores forsvar mod globaliseringens trusler. Det påhviler den enkelte, at tage ansvar for at blive ved med at dygtiggøre sig selv; og hvis man ikke har den påkrævede konkurrenceorientering, skal man til at finde den. Incitamenter til at skifte disse mentale gear kommer via støtteordninger — fx vejledning — og økonomiske sanktioner.

2014 — 2020: Reformernes tidsalder

Som redaktørene skriver i indledningen til *Erhvervsuddannelser - mellem politiske ambitioner og praktisk virkelighed*, er de seneste år præget af en voldsom reformiver omkring uddannelse. Ofte kommer nye reformer inden den forrige reform får tid til at bundfalde sig. De hyppige reformer fremstilles dels som et udtryk for, at en uddannet befolkning er en af velfærdsstatens forudsætninger; og dels i lyset af, at (ungdoms)uddannelse er blevet en central velfærdspolitisk kamparena. Som Jørgensen antyder, har reformerne også aspekter af at være gammel vin på nye flasker. "Nye" tiltag sættes ind for at forsøge at rette op på den foregående reforms utilsigtede konsekvenser — inden næste reform svinger pendulet tilbage igen (Jørgensen, 2016, s.9).

7.2 Spændinger og modsatrettede tendenser

Jeg indledte denne del af specialet med et udtryk for (det til tider anstrengte) forhold mellem arbejdsrettet uddannelse, der har til sigte at give konkrete erhvervsrettede kompetencer der fører til arbejde; og en mere boglig, skolelæret og generel uddannelse, der åbner for mere (over)uddannelse.

Denne spænding udspiller sig til en vis grad også indenfor erhvervsuddannelserne selv, der længe har haft en dobbeltrolle. Erhvervsuddannelserne har i stigende grad skulle give både erhvervsrettede kompetencer og adgang til højere uddannelse; hvilket tog til i efterkrigstiden (Bøndergaard, 2014, s. 47ff) og blev stadfæstet ved indførelse af EUX i 2010.

Samtidigt har 95%-målsætningen; hvor mindst 95% af en ungdomsårgang skulle gennemføre en ungdomsuddannelse, ført til en ændret demografi hos erhvervsuddannelserne. Dette er en central del af tankegangen bag "uddannelse for alle", hvor erhvervsuddannelserne har skulle imødekomme et bredere udsnit af de unge, ved hjælp af øget differentiering (Aarkrog, 2003, s. 365f). Aarkrog beskriver videre, hvordan denne differentiering i Erhvervsuddannelsesreform 2000, ved at tage udgangspunkt i den enkelte elev, nu giver eleven en stor del af, ja ansvar for egen læring, Undervisningen er endvidere forholdsvis abstrakt, og stiller store krav til refleksion 2003, s. 367f. De dominerende pædagogiske diskurser overføres til erhvervsuddannelserne, uden at der skeles

til, om de er bedst tjent med en mere direkte, anvendelsesorienteret tilgang (Aarkrog, 2003, s. 370f).

Derudover har den aktive socialpolitik, der i høj grad betinger sociale ydelser på arbejds- eller uddannelsesvillighed, medført at mange af erhvervsuddannelsens elever fra omkring årtusindeskiftet har været mere ressourcetsvage end tidligere. Dette har, jævnfør de førnævnte utilsigtede konsekvenser, medført et frafald fra ungdomsuddannelserne der følger kravene fra aktiveringspolitikken (Jørgensen, 2016, s.13).

Dette er også et udtryk for spændingen mellem erhvervsuddannelsernes karakterisation som "ungdomsuddannelse" og deres funktion i efteruddannelsen af voksne. Gennemsnitsalderen for elever på erhvervsuddannelserne var knap 24 år i 2018; med uddannelser indenfor området for omsorg, sundhed og pædagogik med den højeste gennemsnitsalder; på knap 30 år (Danmarks Statistik, 2019, s. 14). Dermed vil et ensidigt fokus på at rekruttere unge kunne gøre erhvervsuddannelserne mindre attraktive for voksne der søger opkvalificering, som beskrevet af Jørgensen (2016, s. 13).

Der er heller ikke samsvar mellem antallet udbudte praktikpladser i forhold til antallet optagede elever; ligesom der ikke er opfyldt de politiske mål for optagelse til erhvervsuddannelserne (Danmarks Statistik, 2019, s. 10).

Del III

Metodiske og analytiske overvejelser

“Når jeg anvender et ord,” sagde Klumpe-Dumpe temmelig bånligt, “så betyder det lige netop, hvad jeg vil have, det skal betyde — hverken mere eller mindre.”

“Men spørgsmålet er,” sagde Alice, “om du kan få ordene til at betyde vidt forskellige ting.”

“Spørgsmålet er,” sagde Klumpe-Dumpe, “hvem det er, der bestemmer — det er det, der er det afgørende.”

Lewis Carroll, *Bag spejlet, og hvad Alice fandt der* (1977)

8 Den algoritmiske sociologis muligheder og begrænsninger

Ord er bærende for meget af den mellemmenneskelige kommunikation. Nu kan de færreste af os tillade os en så egenrådig tilgang til ords betydning som Klumpe-Dumpe. Ord har “kun” den mening, vi i fællesskab lægger i dem. Denne mening er ikke nødvendigvis statisk, og nye fænomener nødvendiggør også til tider nye begreber.

Computerbehandling af tekst er netop et sådant nyere fænomen — og afføder begrebet *algoritmisk sociologi*¹. I dette kapitel vil jeg først gennemgå del af den algoritmiske sociologi der omhandler tekstbehandling i grove træk. Jeg vil derefter beskrive nogle faldgruber og begrænsninger der skal tages højde for i anvendelsen af denne tilgang. Det bliver fulgt op af hvilke muligheder denne tilgang til sociologien åbner op for.

8.1 Hvad er computerbehandling af tekst?

Dette fænomen går gerne under betegnelser som “Natural Language Processing” — eller NLP², eller tekstmining³. Muligheden for, at lade computerprogrammer og algoritmer tygge sig igennem store mængder tekst har ikke kun ført til, at Google og Facebook ved mere om dig en din familie gør. Den har også banet vej for, en ny måde at udføre sociologi på — den “algoritmiske sociologi” (eng. “computational sociology”) (J. Evans & Foster, 2019).

Og dagens tæt forbundne informationssamfund har gjort store mængder tekst tilgængelige for (blandt andet) sociologisk analyse — herunder data fra Folketingstidende.dk; hvor man kan hente folketingets udgivelser siden 1953, ganske elektronisk (Folketingstidende, nodate).

J. A. Evans og Aceves benytter sig af en spil-metafor for socialt samhandling og interaktion; hvor der er bred enighed om regler, spillere, hvordan der spilles og vindes. Her er tekst en form for abstraheret repræsentation af (dele af) dette spil — en form for aflejret socialt samspil. Ved at se på tekst som repræsentationer af social samhandling, kan man udlede aspekter af det sociale spil der har foregået — menneskers handlinger, holdninger, følelser og viden (J. A. Evans & Aceves, 2016, s. 22).

I den sociale verden — hvor individer samhandler i relation til sociale strukturer, kulturelle systemer og symboler, samt materielle ressourcer — udspiller der sig flere sociale spil (J. A. Evans & Aceves, 2016, s. 23). Jeg tager del i “spillet om uddannelse” — på metaniveau endda. De politikere, hvis taler kommer igennem møllen af algoritmer, er deltagende i det politiske spil.

J. Evans og Foster henviser til, at formen og genren af tekst har indflydelse på hvilke sociale træk aflejres i den, og at de elementer af det sociale spil der kan udledes, afhænger af, hvor tæt forbundne de sociale træk er til teksten (2016, s. 23f). En SMS-samtale vil kunne gå for at være repræsentativt, eller sågar dækkende, for den overordnede sociale kontekst; hvor en formel, akademisk udlægning til en vis grad holder forskerens personlighed udenfor indholdet.

¹Den algoritmiske sociologi er ikke begrænset til tekstanalyse. Den kan — og bliver — også brugt til at undersøge og forudsige mønstre i interaktioner på sociale medier, hvor der foreligger store mængder observationelle data (J. Evans & Foster, 2019)

²Ej at forveksle med den anden NLP; “NeuroLinguistic Programming”

³Der er semantiske forskelle i disse samlebegreber; men de er ikke gensidigt udelukkende

For dette speciales vedkommende, hvor teksten er referater af folketalingstaler, vil jeg antage, at afstanden til det spil, der foregår på den politiske arena er forholdsvis kort. Derved ser jeg gode muligheder for, at kunne danne indsigter i det politiske spil i min analyse.

Klumpe-Dumpe antydede, at det er centralt at være den der bestemmer, for at kunne give ord og begreber en særlig betydning. Nu er den grundlæggende betydning af ordet *uddannelse* nok ikke særlig flydende over de seneste 4 årtier. Den historiske gennemgang i del II viser dog, at de aspekter af begrebet der har haft størst vægt politisk, har ændret sig over tid. For at bygge spil-metaforen ud⁴; så har de aktører, der har haft ledende positioner i spillet, haft mulighed for at påvirke brættet og reglerne; og de andre spillere har fulgt med for ikke at falde fra.

8.2 Begrænsninger og faldgruber

Men computeroptællinger over tekst gør ikke en sociologisk undersøgelse i sig selv. Computeren har ikke samme sociale og historiske kontekst som beskrevet i Juul i „Diskurser om ungdom og arbejdsliv – fra velfærdsstat til konkurrencestat“; og kan ikke give en analyse. Dette er stadigvæk forskerens ansvar.

De indsigter computeren kan tilbringe, kan også være af en mere overfladisk karakter end hvad den trænede analytiker kan tilvejebringe.

Det er også forskerens ansvar, at ikke falde i faldgruben, hvor man anser algoritmens svar som endegyldige. Algoritmens resultater kan ikke være bedre end de data der bliver tilført den⁵. I tillæg er der til- og fravalg blandt udvælgelse af data og algoritmer der til en vis grad er overladt til forskerens skøn. Mine egne overvejelser herom vil følge løbende i denne del af specialet.

I forlængelse af, at computerens resultater er afhængige af god datakvalitet; kan der også forekomme fejlkilder i dataindsamling og -generering⁶. Der vil også kunne forekomme artefakter i data efter klargøring til behandling – for eksempel fejloversættelser fra PDF-erne og andre uregelmæssigheder. Dette skal der også tages højde for under databehandlingen.

Når tekst behandles som data, vil der også være noget information der går tabt. Computerbaserede analyser er blot en anden vej til at reducere kompleksiteten i det undersøgte. Konkret vil tekstmining typisk glatte ekstremerne ud, og undertrykke lavfrekvent information. Afhængig af metodologi og dataudvælgelse, kan man også bekræfte forudindtagelser lige så vel som at bidrage med ny information (Kwartler, 2017, s. 15). Tekstgenkendelsesalgoritmer er også kendte for, at gengive de strukturelle bias der findes i de tekster de er trænet på (Kiritchenko & Mohammad, 2018).

8.3 Muligheder

Strukturerede klassifikationsalgoritmer kan give pålidelige klassifikationer af massive samlinger af tekst, der langt overgår den enkelte forskers kapacitet at gennemgå. Og ustrukturerede algoritmer kan afdække strukturer i tekstmaterialet, der kan have sociologisk interesse. De seneste års udviklinger indenfor analyseteknologier kan også udlede mening af blandt andet sætningsstrukturer og kontekst, til en vis grad (J. A. Evans & Aceves, 2016, s. 22).

Derudover, så giver den store stigning i data i tekstform muligheder for, at inddrage stadig mere komplekse sociale spil i vores analyser. Dette kan, for eksempel, gennemføres ved at undersøge

⁴Denne udlægning meget frit inspireret af en Bourdieusk fremstilling af magtkampe i et givet felt (Bourdieu, 2004, kap. 1)

⁵den såkaldte “garbage in – garbage out” paradigme

⁶En tidlig gennemgang af datagrundlaget for dette speciale viste for eksempel, at der tydeligt var en mangel på tidlige taler i det indsamlede materiale, i forhold til hvad der er tilgængeligt på folketingstidende.dk

og sammenligne tendenser og ændringer over tid (J. A. Evans & Aceves, 2016). Og udover tekstbehandling kan computeranalyser af registerdata med videre generere nye forklaringsmodeller og analyseprincipper for vores sociale hverdag (J. Evans & Foster, 2019).

En yderligere styrke ved computerbaseret tekstanalyse er, at undersøgelserne er reproducerbare (Kwartler, 2017, s. 13). Koden der ligger bag analysen kan deles, og undersøgelser kan gentages og afprøves uafhængigt.

9 Dataindsamling og analyse

Jeg vil i det følgende beskrive mit datagrundlag, min behandling af dette; og hvilke betydninger dette har for min analyse. Jævnfør min diskussion i indledningen omkring det deskriptives værdi; vil jeg bestræbe mig på, at gå omhyggelig og systematisk til værks. Jeg undlader dog, for læsbarhedens skyld, at dykke for langt ned i de konkrete formler og sandsynlighedsberegninger der ligger bag ved de anvendte algoritmer.

J. A. Evans og Aceves beskriver tre niveauer indenfor tekstanalyse (J. A. Evans & Aceves, 2016, s. 34):

- Man kan se på tekstens indhold, overordnet set, og udlede kollektive tankemønstre og beslutningsprocesser. Hvordan bliver indholdet videresendt, og hvordan udvikler det sig? Hvordan udtrykkes mening?
- Man kan se på kommunikationsprocessen, og de mikrointeraktioner der forekommer. Magt, status, og kulturelle artefakter træder frem, og belyser de konstituerende sociale spil.
- Man kan se på signaler i kommunikationen. Hvilke træk foretages der? Hvilke følelser, ideologier, standpunkter og normer er på spil? Hvad kan der udledes af den underliggende sociale verden?

Jeg vil, som nævnt, foretage en gennemgang af folketingsstaler fra 1978 indtil ultimo 2019. Datasættet¹ er dannet ved, at hente mødereferaterne i PDF-form fra Folketingstidende, og er derefter blevet omdannet til tekstformat, og forsynet med metadata i form af mødets titel, dato, talerens navn og partiforhold med videre, ved hjælp af computerbiblioteket „Folketinget“ (Pedersen, 2019). Denne samling *dokumenter* består af ord, der skal udledes mening af. Ord grupperes som *tokens*, der kan udgøre ét ord; men kan også være to eller flere efterfølgende ord; eller ord der har n ords afstand mellem hinanden².

Min undersøgelse er primært på det første af de tre niveauer i tekstanalyse beskrevet ovenfor. Jeg ser på tekstens overordnede indhold, og forsøger at drage slutninger omkring kulturelle former for kommunikation. Herfra vil jeg undersøge hvilke temaer indenfor uddannelse ses som centrale, og hvordan de italesættes. Mit projekt er også primært at klassificere som *tekst-mining*, idet jeg forsøger mig på udledning af mening fra en samling tekst. Alle ord rystes sammen, uden hensyntagen til deres placering i dokumentet³; og der udarbejdes matematiske modeller over forholdet mellem ord og dokumenter.

9.1 Analyseredskaber og kode

Jeg programmerer mine bearbejdnings og analyser i R, der er et programmeringssprog specifikt udarbejdet til statistisk analyse (The R Foundation, nodate). R er en moden programmeringsplatform, med et stort pakkebibliotek tilgængeligt — også til tekstanalyse, samt redskaber til præsentation af data.

¹Typisk kaldet et *korpus* indenfor tekstbehandlingsverdenen

²*unigrams*, *bigrams*, *n-grams* og *skip-n-grams*, respektive

³Typisk kaldet en “Bag of Words” model (Kwartler, 2017, s. 28)

Reproducerbare computeranalyser har megen lidt værdi hvis de bagvedliggende mellemregninger ikke er tilgængelige. Jeg har derfor en kopi af min kode tilgængeligt på GitHub, til frit brug og offentlig anskuelse ([andersenMasterthesis2020](#)), i bedste Open Data-tradition⁴.

9.2 Forhåndsbearbejdning af data

Helt grundlæggende deler jeg mit store korpus dokumenter i flere mindre korpora, baseret på tidsperioder. Jeg bibeholder dog mit samlede korpus til en overordnet analyse, for sammenligning og kontrol.

Tekstmaterialet kan med fordel simplificeres inden analyse, både for at begrænse kompleksiteten i analysen og begrænse tab af information. For eksempel er ordene “stemme”, “stemte”, “stemmer” alle relaterede til konceptet at stemme. Men uden forhåndsbehandling vil disse blive regnet for at være tre forskellige begreber. Jeg har derfor brugt et NLP-bibliotek (Straka & Straková, 2020) til at foretage såkaldt *lemmatization* ved hjælp af syntaktiske analyser (Kwartler, 2017, s. 28). Computeren analyserer hvert ord, og udleder grundstammen — lemma — af det. I eksemplet ovenfor er grundformen “stem” — og vi kan fange alle omtaler af konceptet at stemme. Dog vil man her også muligvis tabe noget information — handlede det mon om “en der stemmer”, eller “vi har nu talt stemmer op”? “Stemmer” kan også være lyden af mennesker der taler; og ikke, som man kan formode i en folketingskontekst, en kollektiv beslutningsproces.

En anden vej til at reducere kompleksiteten i tekstmaterialet er, at udelukke ord med lav analyseværdi. Tal og tegnsætning kan ofte udelukkes — og er blevet det i dette konkrete tilfælde. Dermed er sidetal og så videre ikke medregnet; dog med tab af information i form af datoer og aldersgrupper til følge.

En meget anvendt teknik er at sortere såkaldte *stopord* fra. Der findes generelle lister over ord der, på baggrund af formodet lav analyseværdi, kan filtreres ud af dokumenterne inden analyse. Men disse lister vil ikke kunne fange domænespecifikke ord af denne karakter; og der er også risiko for, at udelukke mulige meningsfulde analyser på forhånd, (Manning m.fl., 2008, s. 27)⁵. I folketingstalerne omtales “*regeringen*” og ministre forholdsvis oftere end i daglig tale. I stedet for — eller i tillæg til — at jeg på egen hånd vurderer, hvilke ord der skal sorteres fra, kan jeg også her tage en computers evner til at tælle i brug. Den kan beregne hvilke ord der forekommer hyppigt eller sjældent; jeg skal derefter “blot” beslutte mig for en tærskelværdi jeg vil vælge ord indenfor.

Begreber der forekommer meget hyppigt eller meget sjældent vil kunne forvrænge en analyse; og gøre det svært at drage meningsfulde konklusioner. Derfor sorterer jeg de mest almindelige begreber fra; og også dem, der ikke forekommer ret tit. Det gør jeg efter en *tf-idf*⁶ vægtning af tokens pr tale, sorteret de tokens fra, der falder under gennemsnitsværdien af *tf-idf*-score⁷ for *tf-idf*. Dermed har jeg udeladt tokens der forekommer meget hyppigt i min samling korpora, og vil have lav værdi i forhold til at adskille dokumenter fra hinanden. For at udelukke ord der forekommer meget sjældent — typisk tekstkonverteringsfejl og lignende — har jeg derefter udeladt de tokens, der er i de højeste to promille for *tf-idf* af de resterende dokumenter.

Efter jeg har genereret *tf-idf* værdier for tokens i mine korpora, laver jeg et *document-term matrix* over disse. Et *document-term matrix* er en krydstabel over tokens og dokumenter; hvor

⁴Datafilerne jeg har bearbejdet er tildelt (meget) store; og dermed noget udfordrende at gemme på GitHub. Datafiler udleveres dog gerne ved forespørgsel.

⁵Jeg vender tilbage til stopord; og deres analytiske anvendelighed, i del IV.

⁶*text frequency* — *inverse document frequency* er en beregning der tilskriver tokens en værdi efter hvor hyppigt de forekommer i det samlede korpus. Tokens der forekommer i mange eller alle dokumenter vil have en lav *tf-idf*; tokens der findes i færre dokumenter vil have en højere *tf-idf*-score (Silge & Robinson, 2017, s. 29).

⁷Jeg kom frem til at bruge gennemsnitsværdien som nedre tærskel efter en afprøvning af forskellige værdier herfor; hvor andre tilgange overvejende tog for meget information væk.

der markeres, hvorvidt et token forefindes i et dokument. Også her kan datasættets kompleksitet reduceres, ved at fjerne de termer, der ikke findes i ret mange dokumenter.

Som nævnt ovenfor, er en konsekvens af brug af disse teknikker en udviskning af grænseværdier og tab af data i yderpositioner. En overdreven ihærdighed i anvendelse kan bidrage til, en upålidelig analyse i sidste ende.

Jeg vil derfor igen understrege, at denne forhåndsbearbejdning er og vil altid være baseret på forskerskøn i forhold til det konkrete datasæt. Det er agtpåliggende for forskeren, at navigere i de mange muligheder — for eksempel, ved at prøve forskellige grænseværdier og se, hvilke analytiske muligheder der åbnes op eller udelukkes. Forskellige undersøgelsesspørgsmål vil kræve forskellige tilpasninger. Det bliver ikke mere objektivt blot fordi en computer er sat til at arbejde på det!

Mit datasæts særegenheder

En gennemgående udtalelse indenfor databehandling er, at “data er rodete”⁸. For at gøre det ekstra spændende, er de forskellige rå data rodede på hver sin måde. For indeværende opgave er data overvejende pænt forberedt, og forsynet med et rigt sæt metadata. Metadata er nyttig for at lave klassifikationer og undergrupperinger for analyse. Jeg kan for eksempel se, hvem der taler, og hvilket parti denne tilhører, hvornår talen er fra, og hvilken regeringsperiode det er tale om.

Dog har jeg bemærket:

- Tekstkonverteringsfejl, for eksempel i form af ord der deles over linjeskift der ikke genkendes; ord der er indlæst med fejlstavelser (ofte til uigenkendelighed).
- Gentagne sidehoveder og sidefodder i talerne.
- Som tidligere nævnt, var en stor del af taler fra før 1975 ca enten ikke til stede; eller også manglede de relevant metadata.
- Der er meget formalia i et folketingsmøde, hvilket øger forekomsten af enkelte begreber (“vedtage beslutning”, for eksempel).
- Dette medfører også, at spredningen af termer over dokumenter er forskøvet ret kraftigt til venstre, med mange forholdsvis lave vægtninger for $tf-idf$ — det vil sige; der er mange begreber der forekommer i mange dokumenter.
- Denne forholdsvis store homogenitet betyder, at overvejelser omkring skæringspunkter og grænseværdier er svære at kvalificere programmatisk⁹.
- “Huller” og formateringsfejl i indsamling og generering af metadata, som for eksempel partitilhørighed.

⁸Se blandt andre Wickham, (2014) for en R-centreret tilgang til at imødekomme dette fænomen.

⁹Jeg uddyber dette med eksempler i mit analyseafsnit

10 Analysestrategi

Jeg vil i det følgende gennemgå min tilgang og strategi for at finde mening i mit (meget store) datamateriale. Først med en oversigt over de overordnede trin, og derefter en drøftelse af mine overvejelser for disse.

10.1 Overordnet plan for analysen

(Kwartler, 2017, s. 17) beskriver 6 trin for analyse af tekst¹:

1. Definer problemet, og de specifikke mål;
2. Identificer teksten som skal behandles;
3. Skab orden i tekstmaterialet;
4. Lav udtræk af relevante aspekter af data;
5. Analyser;
6. Opnå nye indsigter;

I løbet af min analyse vil jeg bevæge mig gennem disse trin flere gange, med forskellige formål og indhold. Flere af iterationerne vil være tiltagende specifikke, idet jeg kan bruge resultater af de foregående runder som grundlag for de næste. Jeg vil også bevæge mig i mellem de overordnede niveauer efter (J. A. Evans & Aceves, 2016) som er beskrevet ovenfor.

Jeg vil:

- dele data op i perioder
- for hver periode,
 - generere emner på baggrund af tekst-indholdet i de enkelte taler
 - udvælge emner der ser ud til at omhandle uddannelse generelt og (eventuelt) erhvervsuddannelse specifikt
 - i disse emner vil jeg,
 - * se på tendenser i perioderne
 - * generere nye emner med dette subset
 - * se efter sammenhænge på baggrund af analyser på tværs af partiforhold, regeringsmagt, talerens (formodede) intentioner, med videre.

¹Disse trin er i øvrigt udmærkede til de fleste slags undersøgelser

Tekst bliver data

Forskellige trin i denne plan vil stille forskellige krav til at skabe orden i tekstmaterialet, for at gøre den anvendelig til analyserbare data. For eksempel, så vil opdeling i grove kategorier til videre analyse kunne anvende en opdeling i enkelte ord, for at fange overordnede trends. En mere granuleret analyse derimod, vil måske hellere se på to eller tre ord sammen; eller inddrage sproglige analyser for at udlede metadata på sprogniveau.

Unigrams — et ord af gangen — giver et godt billede af frekvenser. Jeg benyttede denne tilgang i de første spadestik af min eksplorative undersøgelse. Dette glatter dog kontekst ud. “Uddannelse” nævnes måske en del gange — men er man kritisk eller positiv? Dette forekom mig som et noget spinkelt grundlag for en analyse.

For næste trin i analysen, hvor jeg vil finde frem til dokumenter der omhandler “uddannelse” på baggrund af de genererede emner fra første trin, foretager jeg analysen med to efterfølgende ord — *bigrams*². Dermed bibeholdes noget kontekst i de videre analyser. Jeg vil udforske mit datasæt både med og uden stopord, for at se hvilken tilgang er mest givende for mine videre analyser.

Perioder

Jeg ønsker at se og sammenligne udvikling i diskurser omkring erhvervsrettede og akademisk orienterede ungdomsuddannelser over tid. Derfor nødvendigt vælger jeg at dele data ind i perioder til sammenligning.

For at opsummere: jeg vil se på diskurser omkring erhvervsuddannelser siden 1978, inddelt i følgende perioder:

- **1978 — 1990:** Uddannelse i fokus
- **1990 — 2001:** Målstyring og selvstyre, under markedsvilkår
- **2001 — 2014:** Ansvar for egen læring, med individet i fokus
- **2014 — 2020** Reformernes tidsalder

Som nævnt ovenfor, viste de første gennemgange af data, at mængden af taler fra før 1978 var meget lille i forhold til de efterfølgende perioder — det vil sige, knap 9000 i perioden 1956–68 og omkring 130 000 taler fra 1968–78. Denne forskel i omfang vil gøre eventuelle statistiske sammenligninger meget tvivlsomme, i bedste fald. Jeg starter derfor mine undersøgelser fra 30. august 1978, hvilket er tiltrædelsesdato for regeringen Anker Jørgensen III, og kort efter indførelsen af EFG på landsplan.

Jeg tager udgangspunkt i centrale hændelser i uddannelseshistorisk sammenhæng; og justerer derudover periodernes grænser til at falde sammen med regeringsperioder. Dermed har jeg mulighed for, at gennemføre analyser på baggrund af regeringsmagt kontra opposition, for eksempel. Dette er ikke den eneste gyldige opdeling³. Man kunne også gerne delt data op i årtier — 59erne, 70erne og så videre. Dette ville give perioder af ensartet længde, samt at man ikke behøver at tage stilling til hvilke år perioderne skal tilhøre. Det ville dog blive sværere at afdække eventuelle diskursændringer omkring specifikke uddannelseshistoriske hændelser.

²Jeg påbegyndte en udforskning af tre efterfølgende ord (*trigrams*), men blev nødt til at springe fra grundet den voldsomme mængde computerressourcer dette krævede.

³Man kunne, for eksempel, også have sat et skæringspunkt omkring 2005–2009 i stedet for i 2014; og dermed opnået lidt mere homogene tidsperioder videre i analysen. Jeg er dog blevet “fanget” af tidlige beslutninger, til en vis grad; og blev ikke bevidst dette i tide til at sadle om

Udledning af emner i det samlede korpus

Jeg sidder nu med ca. 740,000 folketalingstaler, uden at (kunne) vide, hvorvidt en given tale er relevant for min analyse. Mit første skridt vil være, at foretage en grovsortering, for at kunne finde taler der kunne være nyttige at se nærmere på. Dette er en uoverskuelig opgave at gøre manuelt. Men computeren vil gerne lave denne slags beregninger.

LDA

Den algoritme jeg benytter mig af i denne del af processen er LDA⁴ der giver en beregning på, hvor sandsynligt der er, at et token er en del af et givet emne. Alle tokens tildeles alle emner, med forskellige scores per emne. Ved at se på, hvilke dokumenter der indeholder tokens med højest vægtning for et givet emne, kan man udarbejde dokumentklynger. Hvor mange tokens per emne man medregner vil, igen, være op til forskerskøn. Antallet emner til en LDA-algoritme skal dog specificeres på forhånd.⁵

Hvor mange emner er relevant?

Næste trin i grovsorteringen vil derfor være, at finde antallet emner at analysere ud fra. Dette kan biblioteket `ldatuning` til R hjælpe med, der kan sammenholde forskellige beregninger til udvælgelse af antal emner i et ellers ukendt korpus (Nikita, 2016).

Derefter udvælger jeg et antal emner der virker fornuftigt på baggrund af denne proces, og udforsker indholdet af de forskellige computergenererede emner⁶. Jeg noterer, hvilke emner der eventuelt ser ud til at omhandle uddannelse, og tager dokumenter fra disse fra for videre analyse.

Generation af emner indenfor uddannelse

Med en samling af dokumenter, der, ud fra ovennævnte proces, jeg antager at omhandle uddannelse på den ene eller anden måde, har jeg nu opfyldt målet om, at identificere tekst til videre analyse. Jeg gentager nu processen ovenfor med en mindre samling dokumenter, for at se om der er emner at udlede indenfor disse. Herfra vil jeg gentage processen med, at forsøge trække dokumenter fra, der har særlig relevans for mit fokusområde omkring erhvervs- og ungdomsuddannelse.

Dokumenterne indenfor uddannelse udforskes

Forudsat den ovenstående proces er vellykket, har jeg nu en lidt mere overskuelig samling dokumenter, der i større eller mindre grad omhandler (ungdoms)uddannelse. Dette mindre korpus vil jeg udforske videre; med henvisning til mine undersøgelsesspørgsmål fra specialets del I.

Min *topic model* giver mig en pejlepind til, hvilke folketalingstaler er særligt relevante. I denne proces vil jeg støtte mig til min historiske gennemgang og undersøge, i hvilket omfang, jeg kan genkende de historiske tendenser, og dermed have et kvalificeret grundlag til udvælgelse af emner, og folketalingstaler herunder.

⁴LDA står for *Latent Dirichlet Allocation*, og er beskrevet af Blei m.fl. (2003). Kort fortalt: Dokumenter består af en blanding emner — for eksempel "uddannelse", "indvandring", "sundhed"; og emner er en blanding tokens — for eksempel "skolevalg", "videregående", "ungdom" for emnet "uddannelse". Tokens kan godt gå igen i emner; men med forskellig vægtning og forskellige ord omkring sig (Silge & Robinson, 2017, s. 86)

⁵Dette er en såkaldt ikke-superviseret tilgang til tekstanalyse — andre tilgange baserer sig på, at computerprogrammet ved noget om strukturen af data, for eksempel ved tilførelse af annotationer inden analyse

⁶Til denne eksplorative proces tager jeg biblioteket `LDAvis` i brug. Dette program udarbejder en interaktiv webside, hvor man kan se, hvilke tokens er fremtrædende i hvilke emner (Sievert, 2014, 5. marts/2020). Den interesserede læser kan finde frem til mine generede visualisationer på min blog. (M. A. Andersen, 2020b)

En del af mine spørgsmål omhandler politiske holdninger over tid. Til at udforske denne dimension vil jeg anvende *Wordfish*-algoritmen efter Slapin og Proksch (2008; som implementeret i biblioteket *aust* i *n* for R (Lowe, 2019)). Denne algoritme, præsenteret for en samling dokumenter, tildeler hvert dokument en venstre — højre vægtning baseret på relative ordfrekvenser. Jeg vil afprøve, i hvilken grad denne algoritme (der er afprøvet på politiske partiprogrammer) arter sig i mødet med politiske taler. Derudover vil jeg sammenligne en *Wordfish*-beregning over enkelte partiers seneste politiske uddannelsesudspil, og deres taler omkring uddannelse i regeringen Mette Frederiksen.

Jeg ønsker også at gå dybere i min analyse, og drage indsigter ud over hvad fremgår af komplicerede metoder til at tælle ord på. Jeg vil gerne undersøge *tonen* i talerne. Hvordan forholder politikerne sig, når de taler om uddannelse? Dette kaldes *sentiment analysis*⁷; og jeg vil benytte biblioteket *SENTIDA*, der netop er udviklet⁸ til dansk holdningsanalyse (Lauridsen m.fl., 2019, 7. februar/2020). Jeg vil undersøge yderværdier for taler og sætningers holdningscore; og bruge disse som afsæt til en mere tekstnær analyse.

⁷En dansk oversættelse vil nok være "holdningsanalyse"

⁸Dette er også et produkt af studerende på Aarhus Universitet

11 En politologisk undersøgelse i en pædagogisk–sociologisk kontekst

Alt dette arbejde skal have et formål. Jeg vil sammenholde mine undersøgelsesspørgsmål med de trends der er beskrevet i del II, og se, i hvilken grad jeg kan følge med i, at historien bliver skabt, så at sige.

Jeg vil også bruge resten af dette speciale til, at udforske dette krydsfelt af politisk og pædagogisk sociologi jeg har bevæget mig ud i. Hvilke politiske fokusområder og problemstillinger er tæt forbundne med (ungdoms)uddannelse, og hvad kan det betyde, for hvordan man griber videre forskning an?

Jeg vil også gøre mig en ære i, at mine beskrivelser er fyldestgørende. En solid beskrivelse kan have en vis forklarende kraft i sig selv. Men for at kunne inspirere og underbygge fremtidigt pædagogisk-sociologisk forskning; både i krydsfeltet med den politiske sociologi, såvel som med *“big data”* og tekstmining som metodologisk udgangspunkt; er det agtpågivende, at jeg er grundig om omhyggelig i mine beskrivelser. Ikke mindst, der hvor jeg oplever, at andre metodologiske greb end dem jeg har taget vil kunne være fornuftige for fremtidige forskere.

Del IV

Analyse

"Jo, hvis dét skal kaldes Fakta, saa benægter a Fakta!"

Folketingsmedlem Søren Kjær, i debat med Carl Steen Andersen Bille
*Politik skal ikke videnskabeliggøres. Der findes ikke noget facit i politik – kun følelser
og holdninger. Begreber som sandt og falsk eller godt og ondt har ganske enkelt ikke
hjemme i det politiske rum.*

Peter Skaarup, i et ugebrev for Dansk Folkeparti, 2017

12 At beskrive den sociale virkelighed

Denne opgave tager udgangspunkt i, at der antages at være en intersubjektiv social virkelighed, og det er muligt at undersøge denne. Jeg læner mig videre op af et socialkonstruktionistisk verdensbillede; hvor denne sociale verden er i kontinuerlig tilblivelse i en kollektiv samskabelsesproces (Gergen & Mellon, 2017). Et centralt element i denne samskabelse af en fælles social virkelighed er *samtalen* (Gergen & Mellon, 2017, s. 15f). Ord skaber og former virkeligheden. Ikke dermed sagt, at der altid er enighed omkring konturene af denne samskabte sociale virkelighed — heller ikke i politik. Dette understreges, måske lidt kækt, af Søren Kjær på titelbladet for denne del af specialet. Ikke dermed forstået, at man blankt skal afvise påstande man er uenig i; men at man gerne skal have belæg for hvad, man fremlægger som “fakta”. Peter Skaarup argumenterer derimod med et moralistisk udgangspunkt. Man ved hvad der er “rigtigt”; og har ikke behov for beviser, for at det rigtige også skal være “sandt”. Der er et spil der skal spilles; og præmien er en vis form for kontrol over, hvad der fremstår som “rigtigt”¹.

Jeg vil i det følgende bestræbe mig på, at både præsentere og underbygge mine fakta, i kraft af at ikke have en brik med i det politiske spil. Dette gør jeg med udgangspunkt i “trappen” fra del III. Undervejs vil jeg belyse mine undersøgelsesspørgsmål, efterhånden som det falder naturligt. Jeg vil også illustrere de særegenheder jeg måtte finde i mit datasæt undervejs; da dette har stor betydning for, hvordan jeg griber både forarbejdet og selve analysen an.

¹Spil-metaforen, kort drøftet i del II, kan også overføres til en socialkonstruktionistisk ramme: spillets regler udgør et bagtæppe for deltagerens verdensbillede. Dette fordrer specifikke sproglige ritualer og giver specifikke ord særlige betydninger (Gergen & Mellon, 2017, s. 24f).

13 Indskrækning af dokumenter til undersøgelse

Jeg begynder med, at udarbejde et *tf-idf* objekt for hver analyseperiode. En analyseperiode udgør således et korpus, hvor teksten er blevet forhåndsbehandlet og klargjort. For at øge analyseværdien har jeg udeladt ord der falder under gennemsnitsværdien for *tf-idf* for dermed at sortere de aller mest hyppigt forekommende ord fra. For at undgå, at de aller mest sjældne ord forvrænger analysen, udelukker jeg herefter ord der har en *tf-idf* i de yderste to promille.

Derefter foretager jeg en sammenligning af forskellige modeller for beregning af det optimale emner for videre analyse.

Figur 13.1 (side 34) viser en sammenligning af 4 forskellige modeller for udvælgelse af et "optimalt" antal emner for en LDA-baseret topic model. Ved at se på kurvene fra disse sammenligninger, konvergerer kurvene omkring de 25–65 emner, inden de bevæger sig opad igen¹. Jeg går videre med 45 emner, da det er her; der er flest perioder med en "bund" i sine kurve. Et nærmere blik på kurverne giver også nogle indsigter. Perioden 2001 – 14 flader ud helt op til ca 75 emner, i lighed med "kontrolperioden"² *all*. Dette er formentlig et produkt af, at disse grupperinger er på top i antal dokumenter at gennemgå; som funktion af deres forholdsvis lange tidsperioder. Dermed er jeg igen nået til en subjektiv vurdering: Skulle man give hver periode forskelligt antal emner? Jeg vælger at ensrette antal emner. Dette både for sammenligningen mellem perioders skyld; men også af tidshensyn.

En LDA beregning giver ikke navngivne emner, som sådan. I kraft af at være en ikke-superviseret algoritme, bliver dokumenterne fordelt ud over det anviste antal emner³, i det omfang (algoritmen mener) de ligner hinanden. Hvorvidt der er grupperinger der giver mening for mennesker i en social kontekst, og hvilken mening de eventuelt skulle have, er op til forskeren at vurdere. Som eksempel præsenteres et udvalg af emnerne genereret for perioden 2001 til 2014 i figur 13.2 (side 35). Emnerne er repræsenteret af de 15 hyppigste termer pr emne, rangeret efter begrebernes vægning for det pågældende emne.

Emne 4 er et eksempel på, de mange emner der genereres omkring "administrivia". Folketingsmøder har et særligt, formelt sprog, og dette genspejles i rigtigt mange emner, der omhandler denne slags formalia⁴

Der er dog alligevel meget klart, at nogle emner har (pædagogisk) sociologisk relevans. **Emne 5** omhandler tilsyneladende finanspolitik⁵. Man kan ane den førnævnte aktive socialpolitik i

¹Med undtagelse af (Deveaud m.fl., 2014); der ikke ser ud til at være en særlig hjælpsom algoritme i denne sammenhæng. Muligvis en konsekvens af mit noget specielle datasæt.

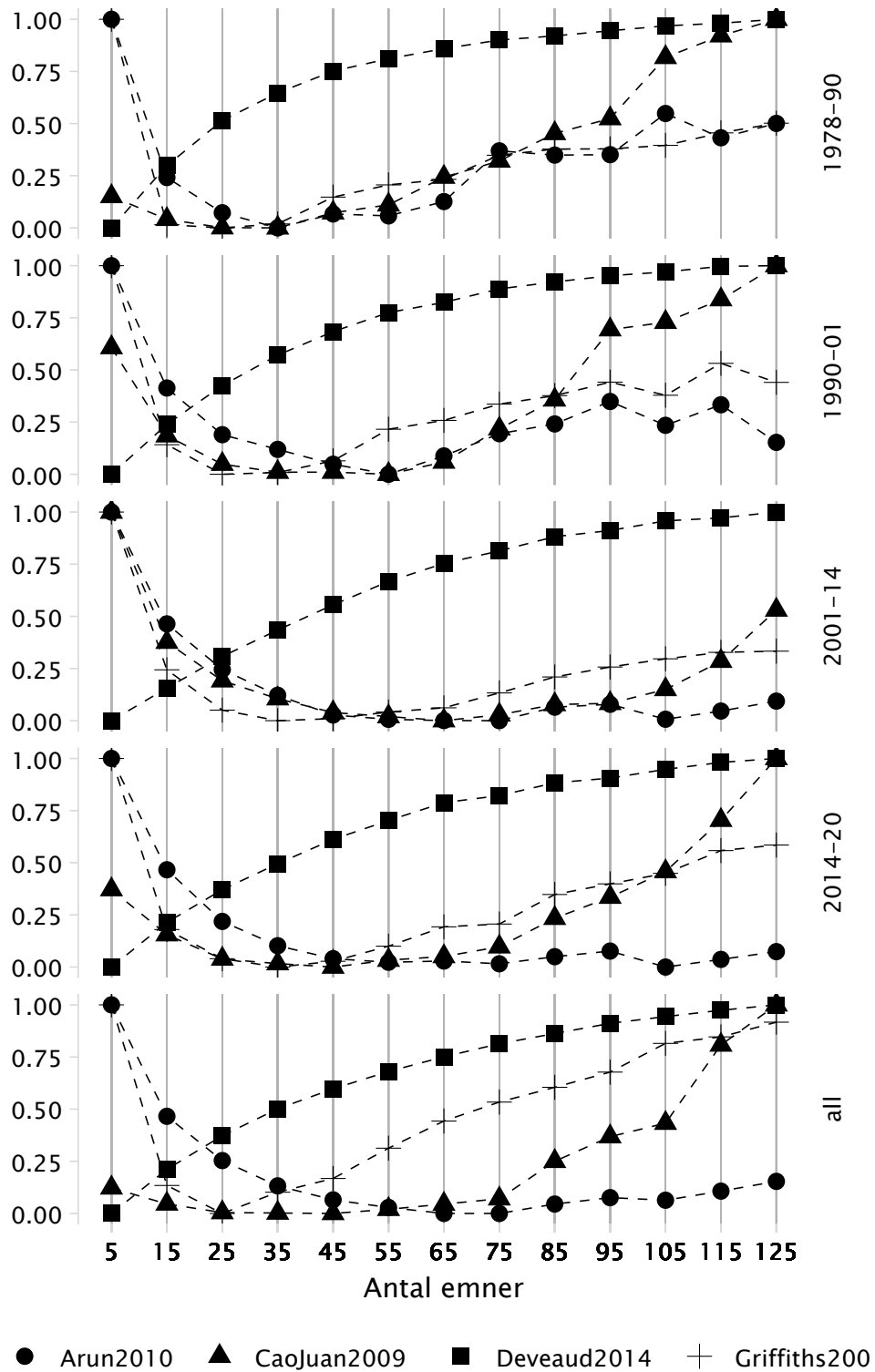
²De emner (og de tilhørende dokumenter) vil dog ikke være de samme som begreberne for de individuelle perioder lagt sammen; da modellen laver beregningen på baggrund af alle dokumenterne på en gang. Jævnfør "sækken med ord" fra min metodiskussion, er der nu langt flere ord i sækken; og de skal fordeles over flere dokumenter.

³LDA tildeler dokumenterne en score γ mellem 0 og 1 for deres bidrag til et givet emne. Emner er distributioner over begreber; hvor hvert begreb tildeles en score β mellem 0 og 1 for at tilhøre et givet emne.

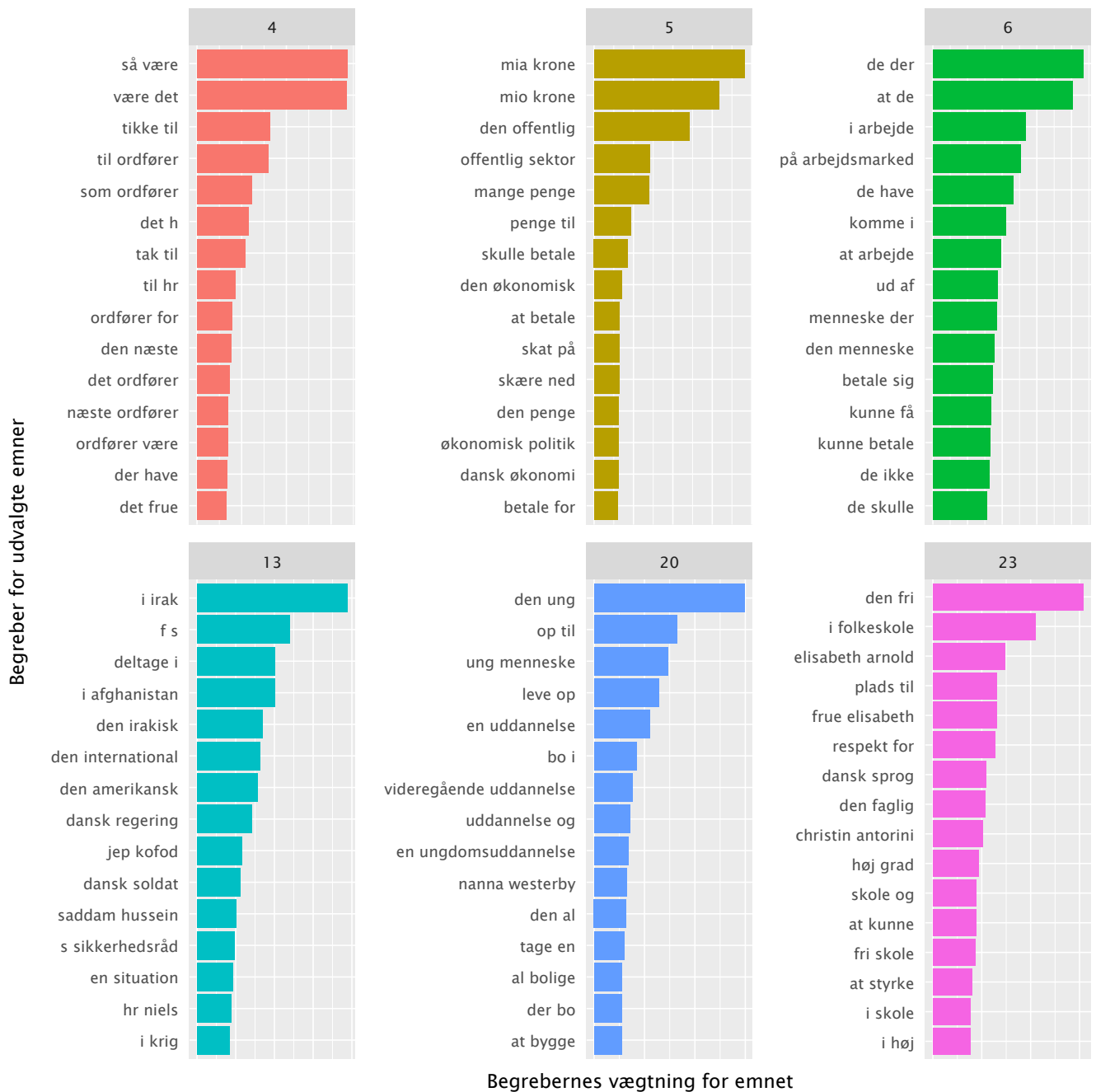
Hvert emne ϕ_k er en multinomiel distribution over ordforrådet W , og hvert dokument θ_d er en multinomiel distribution over K emner. Dermed er sandsynligheden for at et givet dokument (d) bidrager til emnet k $\theta_{d,k}$; og sandsynligheden for at ord w tilhører et emne $\phi_{k,w}$. De multinominale distributioner for emnet genereres af en konjugat Dirichlet prior ((FIXME: dansk begreb?)) $\vec{\beta}$, og distributionerne for dokumenter på baggrund af en Dirichlet prior $\vec{\alpha}$. Deraf navnet "*Latent Dirichlet Allocation*" (Deveaud m.fl., 2014, s.65f)

⁴En grov optælling tilsiger ca. 10–12 af denne slags emner for hver periode; med "kontrolperioden" op over 20.

⁵Et lignende emne, med mange af de samme begreber, er at finde i alle perioder



Figur 13.1: Beregning af optimalt antal emner for videre analyse; bigrams og ingen stopord.



Figur 13.2: Oversigt over udvalgte emner for perioden 2001 – 14, med tilhørende begreber.

emne 6, og krigerne i Irak og Afghanistan i emne 13. Men mest interessant for mit vedkommende er emne 20 og emne 23, der er repræsenteret af termerne “*ung menneske*”, “*videregående uddannelse*”, “*i folkeskole*” og “*dansk sprog*” blandt andre. Forhenværende børne- og undervisningsminister Christine Antorini gør også en optræden. Min udtalelse fra indledningen — det er få politikere der kan undgå at tale om uddannelse — er også underbygget; idet jeg kan finde emner der indeholder beslægtede begreber for hver analyseperiode.

Jeg vil gerne forsøge at øge mit forhold mellem signal og støj. Vil man kunne øge informationstætheden ved at udelukke de mange formalia-relaterede begreber? Dette er et oplagt scenarie til brug af *stopord*. Jeg bruger en forhåndsgenereret samling af generelle stopord (stopwords-iso, 2020); i tillæg til en liste domænespecifikke stopord⁶ jeg selv har udfærdiget. Dette viser sig dog at være en blindgyde. Kurverne er nu uden det pæne “buk”⁷, og svinder ned mod 0; uden der er en opadstigende kurve. Et uddrag af begreber, med antal emner sat til 45, giver tilsvarende utilfredsstillende resultater. At sortere stopordene fra har tilsyneladende medført for stort informationstab til at der er analyseværdi tilbage⁸. Muligvis ville det være muligt at tilpasse listen af stopord, men dette er blevet udeladt af tidshensyn⁹.

Jeg foretager dermed de videre analyser uden stopordslister. Men jeg kommer igen til et vurderingsspørgsmål: Det er nogle perioder, der har to emner der kan relatere til uddannelse bredt forstået; og andre der kun har en. Jeg vurderer, at mere data er mere godt i denne sammenhæng; og tager alle relevante emner med for hver periode.

Men inden jeg dykker ned i uddannelse specifikt, vil jeg gerne se på, hvordan mit datasæt ser ud i forhold til sig selv.

13.1 Talernes indbyrdes relationer

Som tidligere nævnt har hvert datasæt sine særegenheder. Jeg formodede i mit metodeafsnit, at der var en forholdsvis stor ensartethed i min samling af taler. Min hypotese er, at dette skyldes den noget kunstfærdige situation en tale i Folketinget befinder sig i. Talernes kontekst, lejret i det politiske spil, bringer en række rammer for talerne; der i sidste ende formodentlig får en tale til at ligne mange andre. Dette er illustreret ved figur 13.3; hvor emnernes indbyrdes relationer for perioden 2001 — 14 fremgår¹⁰. For at bringe denne diskussion tilbage til en sociologisk relevans, kan også visualisere hvilke emner, computeren mener er nært beslægtede¹¹ med emnerne omkring uddannelse.¹² Vi kan fange vores emner omkring uddannelse, repræsenteret af begreberne “*den ung*” og “*den fri*”.

Også i grendiagrammet er de tæt beslægtede. De ser videre ud til, at alle være en del af en stor klynge, der omhandler det almindelige politiske virke. Alle mine eksempelemner figurerer i denne klynge. Der er videre ikke stor forskel på bladens højde, i kraft af et forholdsvis homogent datasæt.

Næste trin i min analyse bliver nu, at se på indholdet i mine udvalgte emner.

⁶for den konkrete stopordliste se `lib/stopwords.txt` i specialets Github-repositorie (M. A. Andersen, 2020a).

⁷Igen, på nær (Deveaud m.fl., 2014)

⁸Det samme gjorde sig gældende ved andre tiltag for at reducere kompleksiteten i mit materiale — det være sig andre øvre eller nedre terskelværdier for filtrering i *tf-idf*; eller en reduktion af sjældne termer i de genererede *Document-Term Matrices*.

⁹Det tager flere dage til over en uge at sammenligne modellerne

¹⁰Graferne for de andre perioder kan ses på min blog (M. A. Andersen, 2020b)

¹¹Dette er en proces kaldet *hierarchical clustering*. Alle emner tildeles en klynge; hvorpå de klynger der er tættest på hinanden slås sammen. Jeg benytter Ward’s algoritme til dette, hvor klynger med mindst indbyrdes variation grupperes (Ward, 1963).

¹²Jeg har indikeret de emner, jeg trak ud i forrige figur med farve.



Figur 13.3: LDA-modellens emner for perioden 2014 – 20, efter indbyrdes relationer. Emner repræsenteres med deres mest fremtrædende begreb.

Jeg fortsætter, som nævnt, uden stopord; og undersøger de forskellige korpora hver især, og trækker de emner ud¹³, der ser ud til at omhandle uddannelse¹⁴.

¹³Emner 7 og 17 for 1978 – 90; 17 for 1990 – 01; 20 og 23 for 2001 – 14; og 25 for kontrolgruppen; hvis den interesserede læser vil følge med i den interaktive visualisation.

¹⁴Den interaktive visualisation jeg benytter til at udforske disse LDA-modeller kan (også) findes på min blog (M. A. Andersen, 2020b)

14 Analyse af de udvalgte dokumenter

Jeg har nu en samling emner, der ser ud til, at kunne omhandle uddannelse. De dokumenter der har højest sandsynlighed for at tilhøre disse emner skal nu trækkes ud til videre analyse, idet en LDA-model, som nævnt, vil tildele alle dokumenter en sandsynlighed for at tilhøre et bestemt emne. Jeg vælger de taler, der falder i de øverste 2% af sandsynlighed for, at tilhøre mine udvalgte emner. Dette giver en minimumsværdi for sandsynligheden for at dokumenterne tilhører mine udvalgte emner (γ) på omkring 0.1 . . . – 0.2. Dette er givetvis ikke særlig højt; men en lavere andel gav rigtigt få dokumenter at arbejde med¹. Med grænsen sat ved de højeste 2%, er der omkring 300 til 1300 antal dokumenter i de periodemæssige korpora.

Med mine dokumenter trukket ud er der tid til en videre analyse.

Først vil jeg se, hvilke emner der er gennemgående i dette subset af dokumenter. Derefter vil jeg undersøge, hvorvidt der er emner at trække ud indenfor disse dokumenter; da jeg ikke kun er interesseret i uddannelse generelt, men har en specifik interesse i ungdomsuddannelserne. Jeg vil både se nærmere på dette eventuelt forskningsmæssigt interessant subset af data; og også holde øje med, om der er dokumenter der tydeligt ikke omhandler uddannelse inden jeg fortsætter min analyse.

Jeg vil derefter forsøge at at afdække partipolitiske positioner over dette subset.

14.1 Emner indenfor uddannelse

Inden jeg kan analysere på mit subset af emner der (formentlig) omhandler uddannelse, foretager jeg en ny søgen efter et fornuftigt antal emner til gennemgang. Resultaterne er vist på figur 14.1. Denne gang er resultaterne knapt så entydige som de var med den store samling taler. Deveaud m.fl. (2014) er stadigvæk ikke særlig egnet; og har nu fået selskab af Cao Juan m.fl. (2009). Men både Arun m.fl. (2010) og Griffiths og Steyvers (2004) kommer med fornuftige bidrag. Arun m.fl. peger dog konsekvent på flere emner, samtidig som Griffiths og Steyvers er mere stabil i antal foreslåede emner på tværs af mine analyseperioder. For at ikke fare vild i en alt for stor samling evner går jeg derfor efter Griffiths og Steyvers sin algoritme, og lander på 25 emner² for hver periode.

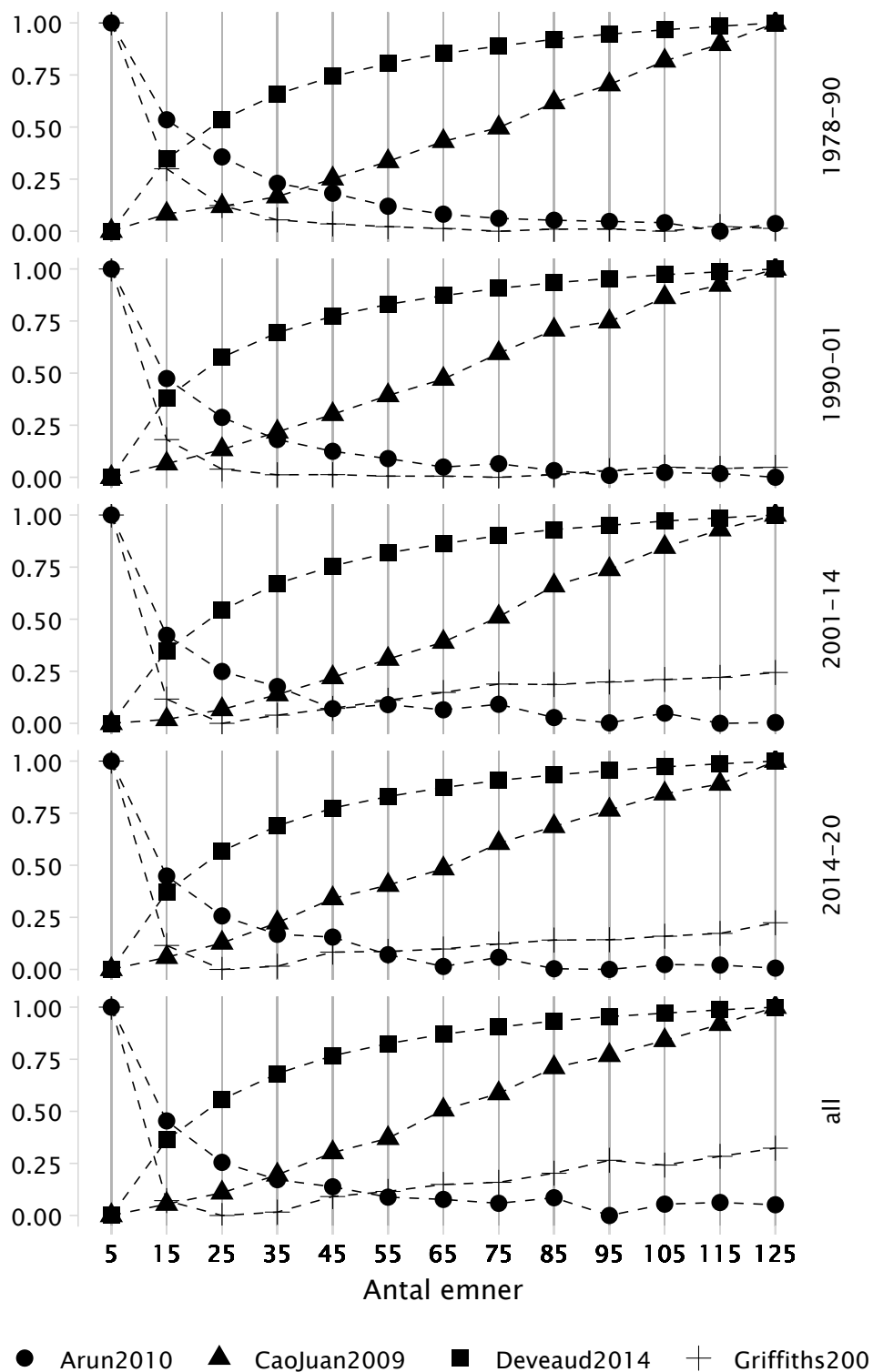
14.2 Tendenser indenfor uddannelse

Jeg gentager min udforskning af min *topic model* med LDAvis-biblioteket (Sievrt, 2014, 5. marts/2020). De genererede emner handler, ganske efter hensigten, overvejende om uddannelse. Som eksempel henvises til figur 14.2 (s. 41), hvor jeg vender tilbage til perioden 2001 til 2014 med udvalgte emner og tilhørende begreber.

Det er nu nemt, at genkende de tendenser jeg drøftede i del II. For eksempel i **emne 5**, hvor vi ser arven efter “*uddannelse for alle*”, hvor det efterstræbes at unge gennemfører en uddannelse. I **emne 8** omtales erhvervsuddannelserne i et lidt andet lys. Her er der antydninger til den aktive socialpolitik, der har sat sit præg på erhvervsuddannelserne i det 21. århundrede, med begreber

¹Jeg prøvede også med en minimumsværdi for γ ; men denne skulle være meget lav før jeg fik en nævneværdig samling dokumenter retur. Derudover var der nogen variation i højeste værdi for γ i de forskellige perioder; med tilsvarende variation i antal dokumenter for en given minimumsværdi.

²Dette er lidt udenfor det “optimale” for 1978 – 90; men det må jeg leve med i ensretningens navn.



Figur 14.1: Beregning af optimalt antal emner for dokumentsamlingen vedrørende uddannelse.

som *“social myndighed”* og *“gennemføre en”* (formodentligt efterfulgt af *“erhvervsuddannelse”*). Dette bliver fulgt til døren i **emne 12**, hvor det gennemgående tema er, at man skal have en *“kompetencegivende uddannelse”*.

Og der er også mønstre at undersøge i andre grænseflader mellem uddannelsespolitikken og den pædagogiske sociologi. I **emne 10** skal særlige behov ikke være en hindring; de studerende skal helst fuldføre på *“normeret tid”* i **emne 13**; og i **emne 20** træder *“nationale tests”* frem.

Kigger vi på resten af analyseperioderne, udspliller de historiske tendenser i erhvervsuddannelse sig også der. I 1978 – 90 var den erhvervfaglige grunduddannelse et centralt element, i delvis konflikt med arbejdsmarkedets parter. Derudover vil man sikre lighed og frihed gennem en almen uddannelse. Dette kan man se i mine emner for perioden³ med begreber som *“frihed til”*, *“folketingsbeslutning om”*, *“arbejdsmarked part”*, *“en praktikplads”*, *“langvarig ledighed”*, og *“kompetencegivende uddannelse”*. De sidste begreber hentyder til en af de centrale spændingsforhold i erhvervsuddannelserne, der ikke kun er ungdomsuddannelser. De har også en central rolle i op- og omkvalificering af voksne.

I 90erne er fokuset på markedsvilkår og arbejdsmarkedet – for både uddannelsesinstitutioner og elever. Der er begreber som *“åben marked”*, *“erhvervs-gymnasiale uddannelser”* (der er eksplicit rettet mod merkantile eller teknologiske videregående uddannelser) og *“gymnasiale uddannelser”*, *“behov for”* og *“måtte lukke”*.

Perioden efter årtusindeskiftet har vi været inde på. Vi skal alle være og blive dygtigere – og dem der ikke kan forvalte dette ansvar selv, skal skubbes i gang. Ud over de begreber nævnt ovenfor, vil jeg fremhæve *“en uddannelse”*, *“skulle gennemføre”*, og (gentagne gange) variationer over *“venstre”*. Videre kan vi også her se spændingen mellem ungdomsuddannelse *“og efteruddannelse”* udspille sig.

I 2014 – 20 er der reform efter reform. Centrale begreber er *“forberede grunduddannelse”*, *“stu en”*, *“en virksomhed”*, og *“brug for”*. Der er ikke mange nye begreber omkring netop mit fokusområde på ungdomsuddannelserne⁴; hvilket kan underbygge min bemærkning fra del II om “ny vin på gamle flasker.”

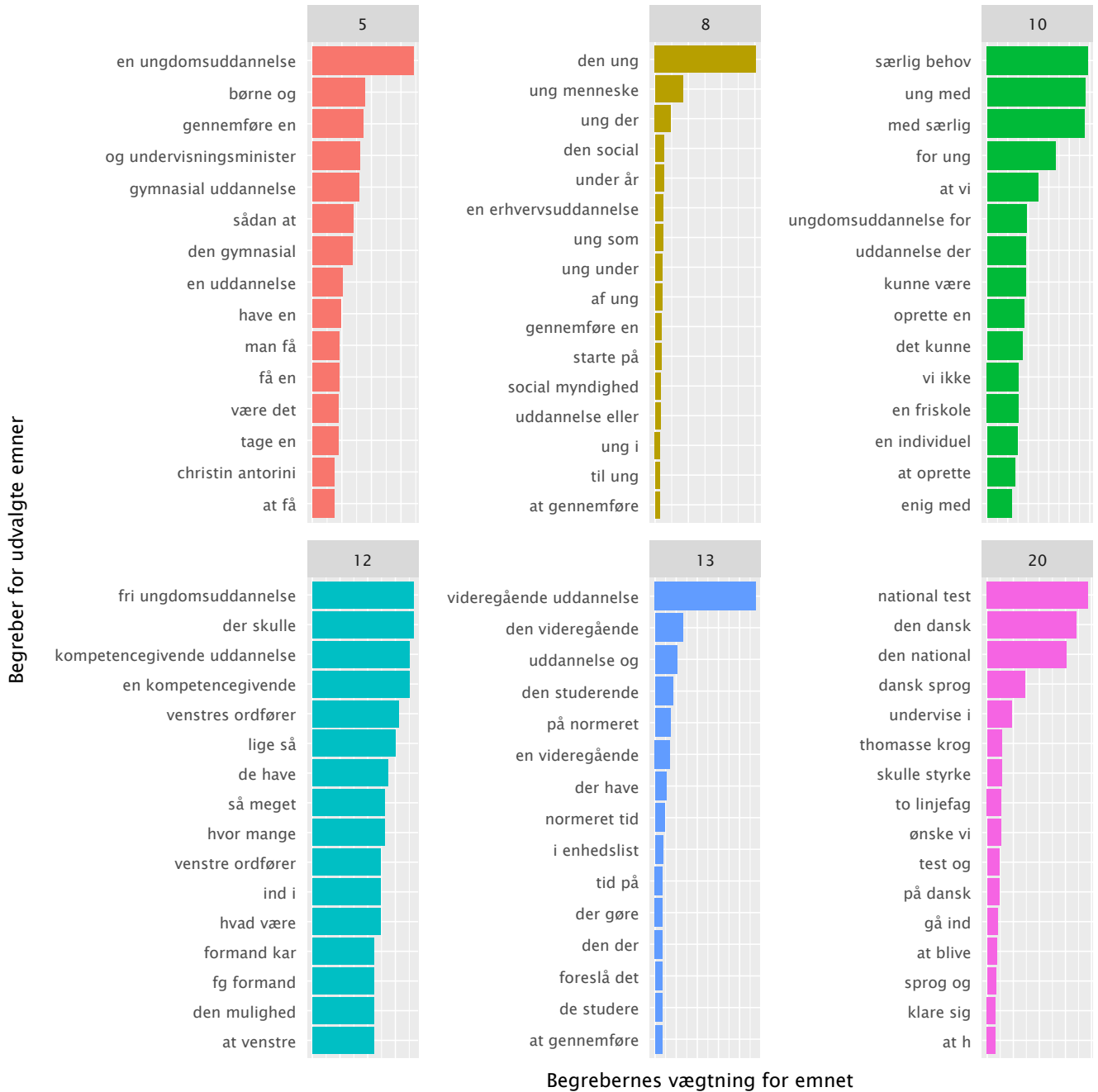
Når vi ser på perioderne sammenlagt, er det gennemgående fokus på *“en uddannelse”* for *“ung menneske”*; med adgang til *“videregående uddannelse”* uden at *“særlige behov”* står i vejen for udvikling af færdigheder.

14.3 Uddannelsesemner ligner hinanden

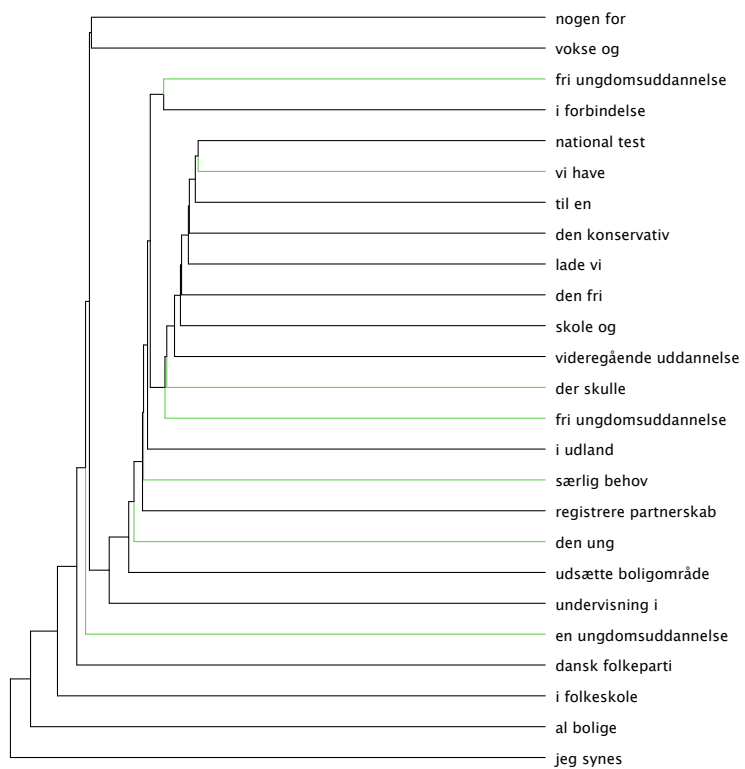
Det, at jeg har samlet dokumenter der overvejende handler om en ting – uddannelse – fremgår også når de præsenteres i et grendiagram. (14.3, s. 42). Emnerne er nogenlunde opdelt i 4 klynger; men der er ikke ret stor forskel på højden af de forskellige grene. Jeg bider mærke i, at der er et par emner, der ikke omhandler uddannelse; men der til gengæld omhandler andre sociale spørgsmål; repræsenteret af begreberne *“udsætte boligområde”* og *“registreret partnerskab”*. Uddannelse er tilsyneladende tæt på andre emner vedrørende samfundsmæssige fordelingsudfordringer. Den sidste kan måske også forklares ved, at undervisning har historisk set været tæt forbundet med kirken.

³ Disse begreber er taget fra et udvalg af emner jeg vurderer har relation til *erhvervsuddannelse*, bredt fortolket.

⁴ men der er andre udviklinger indenfor uddannelse at spore – forskning og videregående uddannelser er mere fremtrædende, for eksempel



Figur 14.2: Oversigt over udvalgte emner indenfor uddannelse for perioden 2001 – 14, med tilhørende begreber.



Figur 14.3: Grendiagram over emner indenfor uddannelse.

15 På vej ned i detaljen

Ordoptællingen ovenfor gav en del udbytte, i forhold til at undersøge hvilke begreber der var sigende for den enkelte periode. Det var dermed muligt at genkende trende og tendenser i dansk uddannelsespolitik over tid. Det er også en anden gevinst. Jeg kan sigte endnu mere skarpt på mit undersøgelsesområde: ungdomsuddannelser bredt forstået; med særligt fokus på erhvervsuddannelserne. Jeg foretager nok en sortering i emner og tokens; og trækker ud en håndfuld emner¹ fra mit uddannelsesubset. Disse er udvalgt på baggrund af en vurdering om, at de i større end mindre grad omhandler netop mit undersøgelsesområde. Jeg vælger denne gang at beholde de øverste 10% af dokumenter der kunne omhandle ungdomsuddannelser; og står tilbage med mellem 100 og 600 dokumenter for hver analyseperiode.

Nu kan jeg udforske mine undersøgelsesspørgsmål mere i dybden; med et specifikt fokus. Jeg lægger ud med et forsøg på, at gøre de politiske positioner omkring uddannelse op, over partipolitiske linjer. Derefter vil jeg lave en række holdningsanalyser, for at forsøge at afdække noget om, *hvordan* politikere taler om uddannelse.

15.1 Partipolitiske positioner omkring uddannelse

Jeg går fra én avanceret måde at lave ordoptællinger til en anden avanceret måde at lave ordoptællinger. Denne gang for at undersøge hvordan partiernes taler forholder sig til hinanden. Jeg vil benytte mig af den før omtalte *Wordfish*-algoritme (Slapin & Proksch, 2008).

Denne algoritme tager for sig en samling dokumenter der ideelt set omhandler nogenlunde det samme emne (hvorvidt de enkelte dokumenter omhandler det samme emne, og hvad der ingår i et bestemt emne, afhænger af forskerskøn)². Slapin og Proksch demonstrerer deres analyse på partiprogrammer - først i deres helhed; derefter delt op i politiske interesseområder.

For at se efter generelle tendenser, vil jeg analysere talerne der omhandler (ungdoms)uddannelse på baggrund af politiske fløje. Jeg vil også undersøge talerne på partiniveau, for til sidst at foretage en sammenligning af relative positioner fra partiprogrammer kontra politikernes taler.

I denne del af min undersøgelse tager jeg for mig min nye samling folketingstaler der skal analyseres videre; og tildeler dem en politisk fløj på baggrund af politisk parti³. Derudover kategoriseres Grønland og Færøerne for sig⁴. Min opdeling kan ses i tabel 15.1 på side 44. Jeg udelukker også nogle mindre partier fra den partispecifikke analyse. De partier der indgår i sammenligningen, er *Enbedslisten*, *Socialdemokratiet*, *Socialistisk Folkeparti*, *Dansk Folkeparti*, *Venstre*, *Konservative Folkeparti*, *Radikale Venstre*, *Liberal Alliance*, *Ny Alliance*, *Alternativet*, og *Fremskridspartiet*. For ikke at gøre de følgende figurer for svære at læse; er dog *Liberal Alliance*, *Ny Alliance*, *Alternativet*, og *Fremskridspartiet* udeladt fra disse.

¹Hvis man ønsker at undersøge disse i den interaktive model, er det emner 3, 7, 8, 9 og 24 for 1978–90; 14, 17, 18, 23 og 3 for 1990–01; 5, 8, 10, 12, 22 og 23 for 2001–14; 1, 3, 5, 6, 9 og 13 for 2014–20; og til sidst kontrolgruppen med emnerne 2, 7, 14 og 15.

²*Wordfish* (et ordspil over ord og den franske betydning af *poisson*) modellerer ordfrekvenser mod en Poisson-distribution; og tildeler en en-dimensionel værdi på baggrund af dette. Dette er også medvirkende til, at algoritmen har det bedst med dokumenter der falder indenfor samme emne; da man ellers risikerer en meget stor spredning i ordfrekvenser dokumenterne i mellem.

³Der er en mangel i mit datasæts metadata; hvor der ikke forefindes partitilhørighed på talere som "statministeren"; "formanden" og så videre. Disse er tildelt "ikke angivet" som blok. Det viste sig dog, at der ikke var nogen observationer i de data jeg havde trukket ud, der faldt i denne gruppe.

⁴Dette gør det muligt at udelukke disse i mine videre beregninger; for at undgå unødigt støj i mine data.

Tabel 15.1 : Fordeling af partier i politiske blokke

| Blå blok | Rød blok | Centrum | Grønland/Færøerne |
|-------------------------|------------------------------|-----------------------|---------------------|
| Dansk Folkeparti | Alternativet | Uden for partierne | Javnaðarflokkurin |
| Venstre | Enhedslisten | Radikale Venstre | Tjóðveldisflokkurin |
| Konservative Folkeparti | Socialdemokraterne | Kristeligt Folkeparti | Fólkaflokkurin |
| Fremskridtspartiet | Socialistisk Folkeparti | Centrum-Demokraterne | Sambandsflokkurin |
| De Uafhængige | Danmarks Kommunistiske Parti | Ny Alliance | Atássut |
| Nye Borgerlige | Venstresocialisterne | Liberal Alliance | Siumut |
| | Fælles Kurs | Retsforbundet | Inuit Ataqatigiit |
| | | Liberal Centrum | |

Wordfish-algoritmen beregner positioner fortolket som en venstre-højre akse ud fra dokumenternes ordfordeling sig i mellem. En negativ værdi trækker mod venstre; og en positiv værdi mod højre. Dokumenternes position holdes uafhængig af hinanden; således at et partis position i periode t ikke påvirkes af positionen i periode $t - 1$. Forbliver partiernes positioner forholdsvis uændrede, har partierne brugt lignende ord over tid. Er der derimod ændringer i hvilke ord der bruges over tid, vil dette vise sig i en ændret position over tid.

Jeg udarbejder et nyt korpus, hvor hvert korpus udgør talerne for hver blok eller parti. Disse taler slås sammen til ét dokument pr analyseperiode⁵. Ud fra dette tæller jeg op antallet begreber per sammenlagte taler, og laver beregninger ud fra dette. Dette gentager jeg for regeringsperioder; for at have en endnu mere detaljeret analyseposition. Til sammenligning laver jeg også en *Wordfish*-model, hvor jeg slår alle taler for de politiske blokke sammen fordelt over mine analyseperioder.

Antal observationer for hver inddeling fremgår af tabel 15.2. Der er en del variation i antal observationer; især hvad angår *Ny Alliance*. Jeg bibeholder den ene tale alligevel, som videreførelse af *Liberal Alliance*. Rød og Blå blok er, ikke uventet, vel repræsenteret. Hvad angår perioderne; er 1990 – 01 noget underrepræsenteret. Dette er værd at have in mente når vi går ind i de videre analyser. Spredningen for taler over regeringsperioderne er forholdsvis stabil; men kan deles op i et noget sparsomt 20. århundrede og en meget mere frugtbar tid efter 2001.

Politisk bevægelse over tid

I figur 15.1 fremgår der, hvordan *Rød Blok*, *Centrumpartierne* og *Blå Blok* har positioneret sig over analyseperioderne efter en *Wordfish*-beregning. Ud fra de udvalgte taler og partier, ser vi rød blok solidt til venstre i 1978 – 90, med centrumpartierne og blå blok positioneret forventeligt. Men så begynder en “sund fornufts-forventning” at blive vendt på hovedet. Rød blok er stadigvæk solidt til venstre i 1990 – 01; men skifter position med blå blok i 2001 – 14; for derefter at gentage bedriften i 2014 – 20. Centrumpartierne og blå blok har også skiftet position i 1990 – 01; hvorefter partierne i centrum holder sig solidt i midten. Dette noget underfundige resultat kan skyldes et for spinkelt datagrundlag. Jeg har dog et bud på en anden hypotese; som jeg vil vende tilbage til.

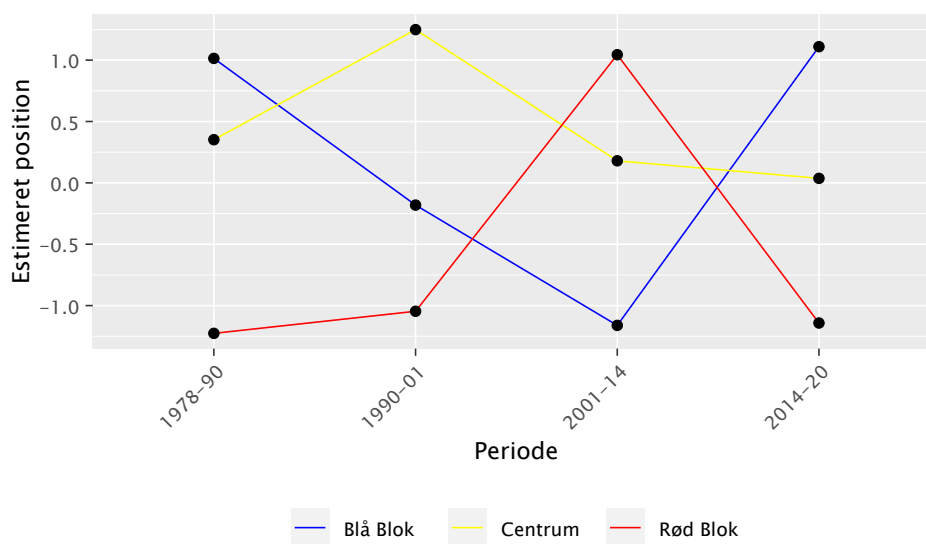
Her ser vi, hvordan blokkene har ændret position i hver analyseperiode. Der er nogen bevægelse i de forskellige blokke, men de relative positioner er forholdsvis stabile.

Figur 15.2 viser de politiske partiers bevægelser i løbet af mine analyseperioder. Man kan se *Socialdemokratiet* og *Socialistisk Folkeparti* bevæge sig mere mod højre, en tendens disse partier deler med blandt andre *Radikale Venstre* og, påfaldende nok, *Enhedslisten*; med *Dansk Folke-*

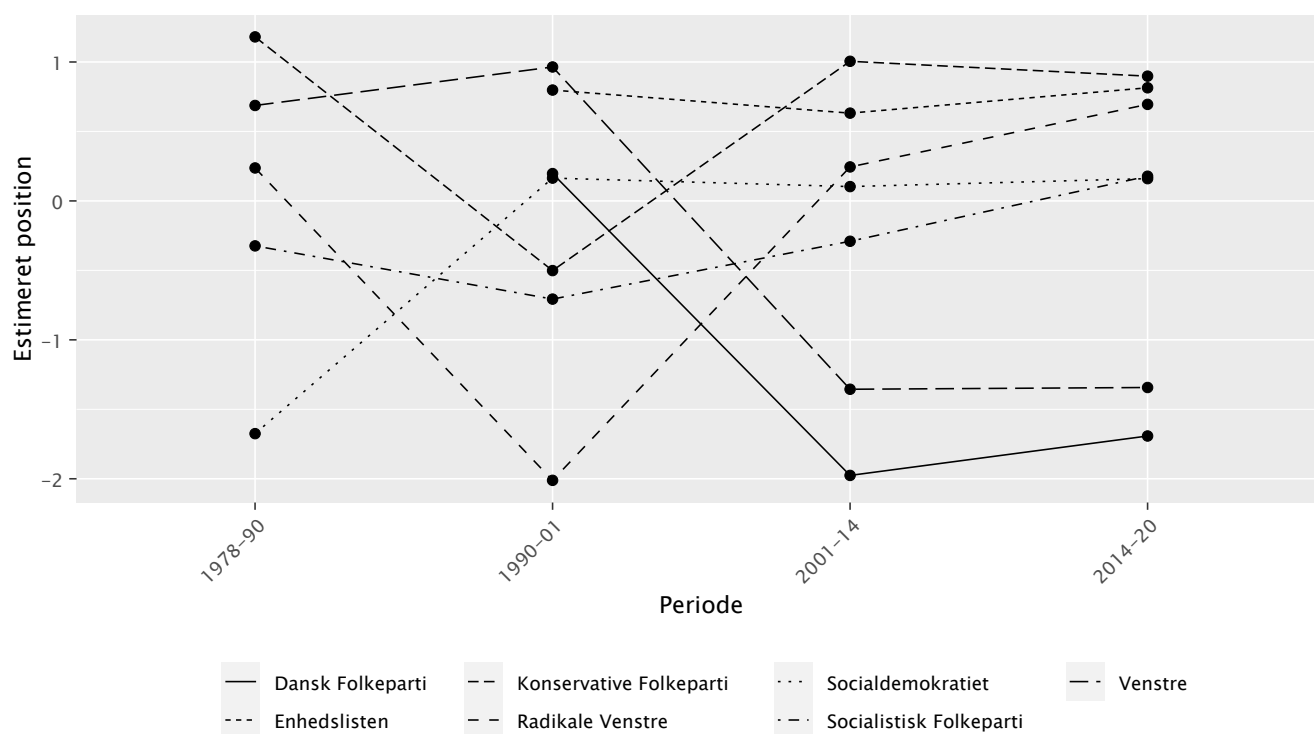
⁵Det vil sige, at partier og blokke der ikke fremgår af disse opgørelser er udeladt fra analyserne. Da *Wordfish*-algoritmen sammenligner alle de medtagne dokumenter i sin vægning, vil det risikere at forvride den relative præsentation af partiernes indbyrdes forhold

Tabel 15.2: Oversigt over antal observationer for hver periode, i taler omhandlende uddannelse

| Parti | Taler | Blok | Taler | Periode | Taler | Regering | Taler |
|-------------------------|-------|--------------|-------|-----------|-------|-------------|-------|
| Alternativet | 6 | Rød Blok | 305 | 1978 – 90 | 126 | 1977 – 1979 | 3 |
| Dansk Folkeparti | 87 | Centrum | 99 | 1990 – 01 | 69 | 1979 – 1981 | 20 |
| Enhedslisten | 49 | Blå Blok | 269 | 2001 – 14 | 365 | 1981 – 1984 | 18 |
| Konservative Folkeparti | 48 | ikke angivet | 48 | 2014 – 20 | 161 | 1984 – 1987 | 28 |
| Liberal Alliance | 22 | | | | | 1987 – 1988 | 11 |
| Ny Alliance | 1 | | | | | 1988 – 1990 | 25 |
| Radikale Venstre | 58 | | | | | 1990 – 1994 | 11 |
| Socialdemokratiet | 172 | | | | | 1994 – 1998 | 15 |
| Socialistisk Folkeparti | 66 | | | | | 1998 – 2001 | 40 |
| Venstre | 123 | | | | | 2001 – 2005 | 97 |
| | | | | | | 2005 – 2007 | 62 |
| | | | | | | 2007 – 2011 | 134 |
| | | | | | | 2011 – 2015 | 89 |
| | | | | | | 2015 – 2019 | 112 |
| | | | | | | 2019 – | 8 |



Figur 15.1: Politiske blokkes politiske positioner over analyseperioderne



Figur 15.2: Politiske partiers politiske positioner over analyseperioderne

parti i en klar bevægelse mod venstre; sammen med *Konservative Folkeparti*. Det overvejende indtryk af grafen er i sidste ende en generel samling mod højre, med et par afstikkere i bestemt opposition.

Når det er sagt, er denne graf ikke nødvendigvis særlig oplysende i min konkrete problemstilling; med et noget ujævnt datagrundlag som største udfordring.

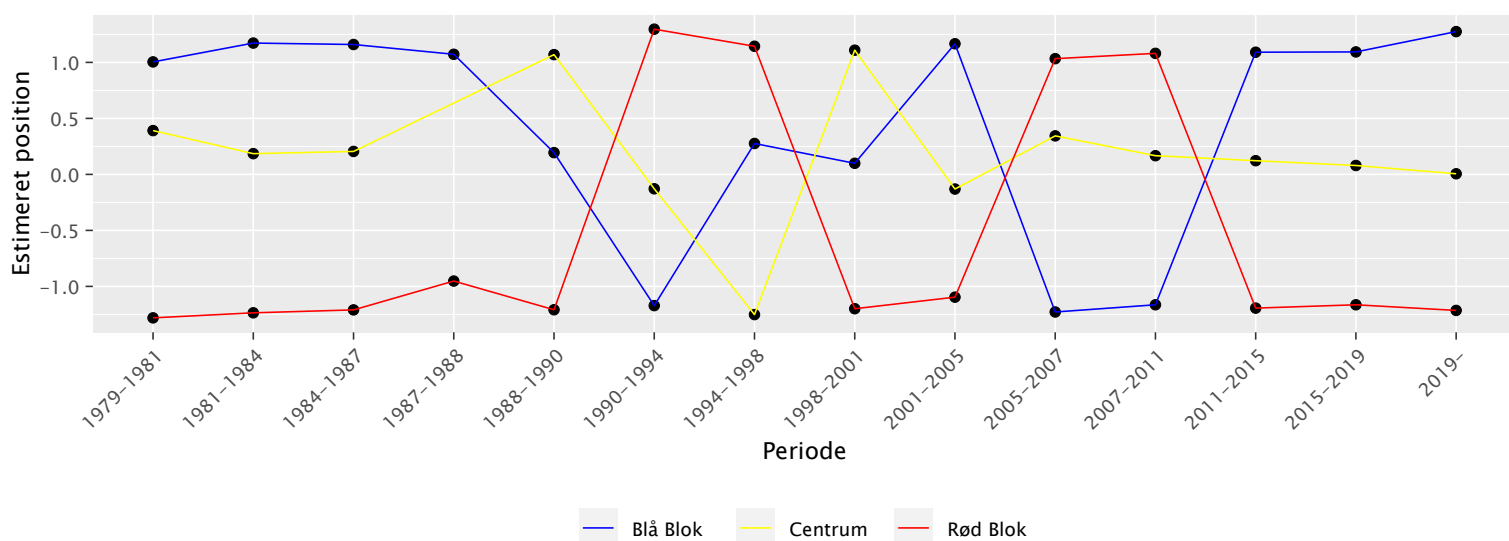
Når jeg tager de politiske blokke, og stiller deres politiske positioner op mod specifikke regeringsperioder, er det noget nemmere at se en udvikling. Meget af det mudrede billede i 90'erne kunne muligvis med rette tilskrives et spinkelt datagrundlag. Men min hypotese (som jeg lader ligge i denne omgang) er, hvorvidt og hvorledes politikeres udtalte politiske positioner påvirkes af at være i eller udenfor opposition til den siddende regering. En gennemgang af dette mulige mønster vil kræve et noget anderledes undersøgelsesdesign og strukturering af data. Det ville også være mere lige til for en politolog end for en sociologs.

Partiprogrammer kontra taler

Jeg har lige undersøgt, hvordan politikeres taler omkring uddannelse arter sig på en venstre-højre akse. Men er der samsvar mellem partiernes udtalte holdninger og deres udtalelser?

Inspireret af Slapin og Proksch (2008), har jeg gennemgået de nuværende partiprogrammer fra de større danske partier⁶. Jeg har trukket det materiale ud, der omhandlede uddannelse, bredt

⁶Data til dette lille eksperiment blev alle samlet ind fredag 9. oktober fra partiernes hjemmesider (Dansk Folkeparti, 2020a, 2020b; Det Konservative Folkeparti, 2020; Enhedslisten, 2020; Radikale Venstre, 2020; Socialdemokratiet, 2020a,



Figur 15.3: Blokkes politiske positioner over regeringsperioder fra 1978 til 2020

forstået, og lavet en ny *Wordfish*-analyse på dette materiale.

Dette sammenligner jeg med en *Wordfish* analyse over den seneste folketingsperiode⁷ I figur 15.4 (s. 48) kan vi se, at *Socialdemokratiet*, *Radikale Venstre* og *Dansk Folkeparti* har et godt samsvar med deres officielle position og deres politikeres folketingstaler. Det er også påfaldende, at *Socialdemokratiet* lægger sig lige til højre for midten. *Venstre* har derimod en meget stor spredning. Deres partiprogram er solidt til højre; men deres udtalelser er derimod kraftigt til venstre. *Enbedslisten* har også en lignende, om ikke lige stor, spredning, dog med modsat fortegn. Også *Socialistisk Folkeparti* og *Radikale Venstre* har en vis spredning; om end ikke helt så udtalt.

Overvejende er der overensstemmelse med positionerne til partiernes politiske programmer og en "sund fornuft" forventning. Den til tider store spredning mellem folketingstalerne og parti-programmerne inviterer til gengæld til en mere gennemgående undersøgelse, eventuelt med en større mængde data til sammenligning.

Hvilke begreber er kendetegnende for de politiske fløje over tid?

Man kan også undersøge de værdier, *Wordfish*-algoritmen har tildelt de enkelte begreber. Jeg laver en beregning over taler slået sammen for hver analyseperiode, og kan der dermed undersøge, hvilke begreber der trækker mod venstre og henholdsvis højre. Man kan også se, hvilke termer der forholder sig neutrale.

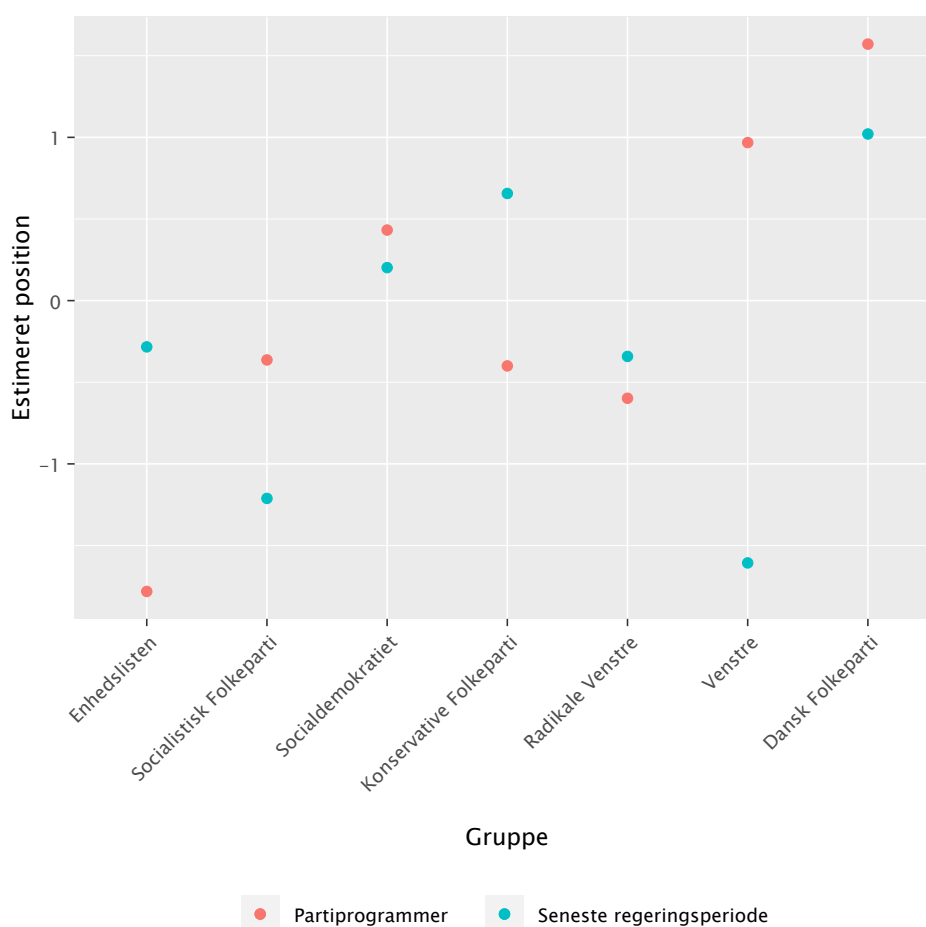
Ved at se på figur⁸ 15.5⁹, fremgår der, at fraser som "*det være*", "*være en*" er så almindelige, at de ikke indgår i tildeling om emnet. Ved "tårnets" base finder vi begreber der har mere vægt-

2020b, 2020c, 2020d; Socialistisk Folkeparti, 2020; Venstre, 2016).

⁷Jeg sammensatte mit datasæt i Januar 2020. Der er dermed taler med for ultimo 2019 i denne model.

⁸For indsigtens skyld har jeg sorteret interessante fyldord, gentagelser og navne fra i diagrammet.

⁹Den noget karakteristiske form ("Et Eiffel-tårn af ord", som beskrevet af Slapin og Proksch) kommer af, at politisk neutrale ord ikke får tildelt en positiv eller negativ vægtning; og vil ikke trække et parti eller en blok i en bestemt retning. Ord der bliver brugt ofte af alle parter, med andre ord, er i udgangspunktet politisk neutrale (Slapin & Proksch, 2008, s. 709)



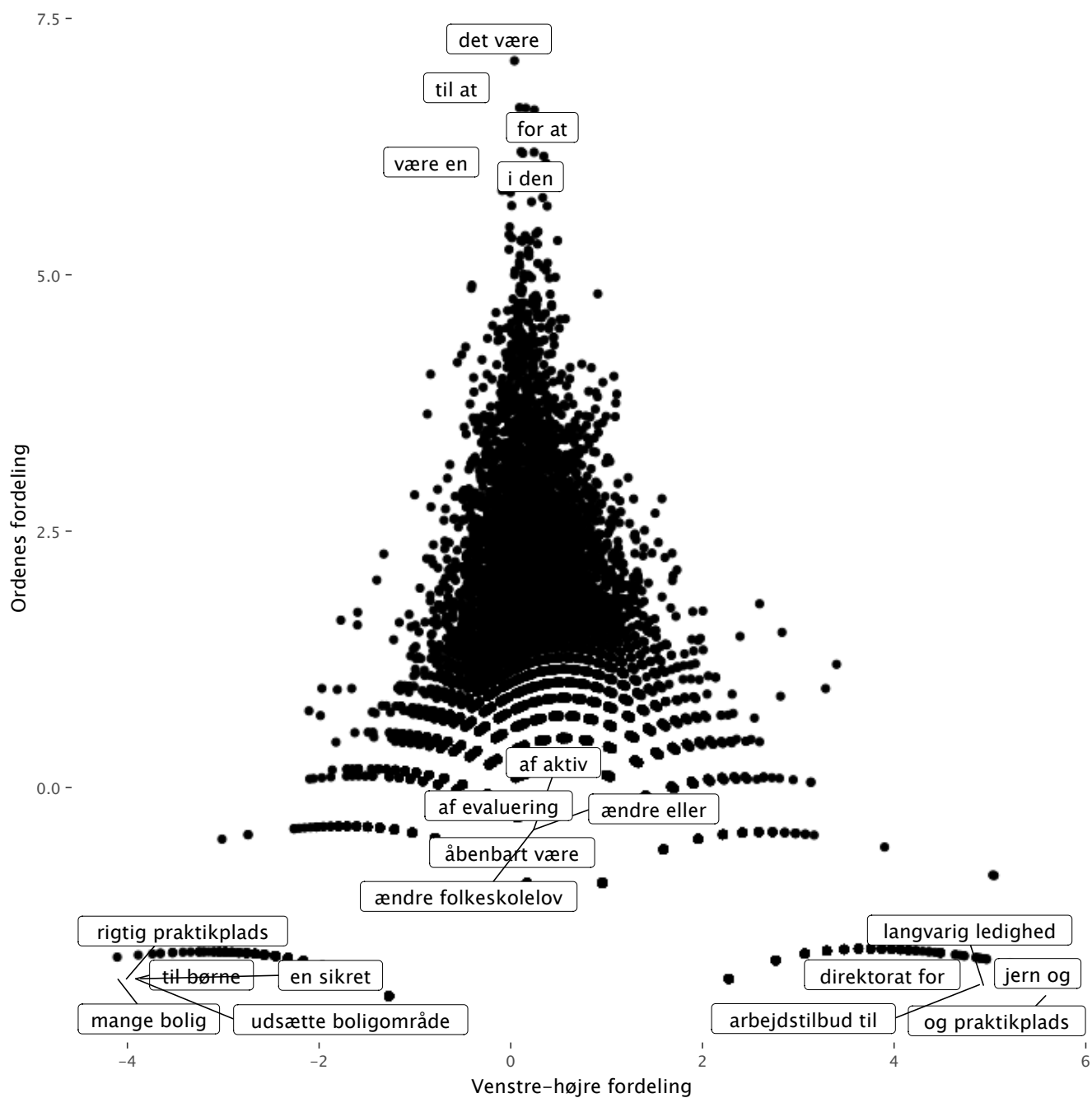
Figur 15.4: Sammenligning af positioner omkring uddannelse i partiprogrammer og den seneste regeringsperiode.

ning indenfor uddannelse; uden at være værdiladede. Politikere er tilsyneladende optagede “af evaluering” og at “ændre folkeskolelov”, uden at det er grundlag for stor uenighed.

Begreberne “rigtig praktikplads” og udsætte boligområde” indikerer en venstrerettet position. Boliger er ikke som sådan forbundet med uddannelse direkte. Dette tyder dog (igen) på, at uddannelsespolitik i høj grad også er situeret som socialpolitik. I de højreladede termer forholder politikerne sig til arbejdsmarkedet og industrien. De, der er i “langvarig ledighed” kan få “arbejdstilbud til”; og der skal og være praktikpladser. Dog ikke med fokus på, at den skal være rigtig.

Med et overblik over ordfordeling og vægtning for hele mit datagrundlag, kan vi nu udlede samme data for de enkelte analyseperioder. Ordfordelingen for hver periode kan aflæses i tabel 15.3, hvor de centrale begreber der giver venstre, højre, og neutral vægtning for hver periode fremgår.

Der er, også her, en del “administrivia” at se. Men der er alligevel tydelige (om ikke nødvendigvis overraskende) forskelle mellem venstre og højre at spore. Begreberne på venstre side vil



Figur 15.5: Ordfordeling og vægtning for de politiske blokke, set som helhed over mine analyseperioder.

Tabel 15.3: Oversigt over begrebsvægtning efter en *Wordfish* beregning af talerne i analyseperioderne.

| Periode | Venstreladede begreber | Værdineutrale begreber | Højreladede begreber |
|---------|--|--|---|
| 1978–90 | oprettelse af, bo kristen, fremgå af, hvis vi, og demokrati, som ordfører, bekræfte at, cirka cirka, elev med, give udtryk, i landsarbejdsnavn | det være, for at, til at, der være, for den, åbne op, åbning i, adgang for, ændring og, ærlig tale | kgl teater, den kgl, teknisk uddannelse, have han, en teknisk, teknisk gymnasium, jørgen ped, ped han, oplysnings forbund, til efteruddannelse |
| 1990–01 | god undervisningsmiljø, i socialdemokrati, til hr, på samsø, egu en, sfo være, danish crow, og mulighed, af nogen, en selvstændig, måtte lukke, socialdemokrati ville, studere have, studerende undervisningsmiljø | det være, for at, til at, der være, i den, abonnementskort med, ad frivillighed, af arbejdsulykke, af forhold, af lovgivning | den almen, egen virksomhed, i cd, en cvu, fysik og, og innovation, almen gymnasiale, evaluering af, faglig niveau, og kemi |
| 2001–14 | social myndighed, krame mikkelse, ung lovovertreder, en sikret, betingelse dom, anvendelse af, begå kriminalitet, tina peters, ung kriminel, vilkår om | det være, der være, til at, for at, at der, social myndighed, a la, adskillelse af, ændre den, ærgerligt at | h mad, freddie h, mest relevant, ny ungdomsboliger, kunne starte, på erhvervsskol, ville minister, der startede, en uddannelsesgaranti, operation dagsværk, ung fisker, vi kigge |
| 2014–20 | ny ramme, i alternativ, i enhedslist, variere undervisning, åbne op, i sf, jeg høre, lige mulighed, permanent ophold, betyde nogen, dac e, den variere, det den, helt afgørende, pege på, su en | det være, til at, der være, for at, for den, adgang frem, adgang ind, adgang på, adgang så, ændring som | af studerende, i efterværn, engelsksproget uddannelse, international studerende, uddannelsesinstitution skulle, få su, høj faglig, den kriminel, international skole, al studerende, år der, clau larsenjens, en erkendtlighed, til vandrende, udvikle en, ung mellem |

have “*eleven med*” til et “*godt undervisningsmiljø*” med “*varieret undervisning*”. Og igen trænger socialpolitikken sig på, med et fokus på “*social myndighed*”. Mod højre er fokuset i langt højere grad på produktion og innovation. En “*teknisk uddannelse*” vil kunne føre til en “*egen virksomhed*”. Vil man dygtiggøre sig, skal der være “*en uddannelsesgaranti*”. Og uddannelserne skal have et “*højt fagligt*” niveau.

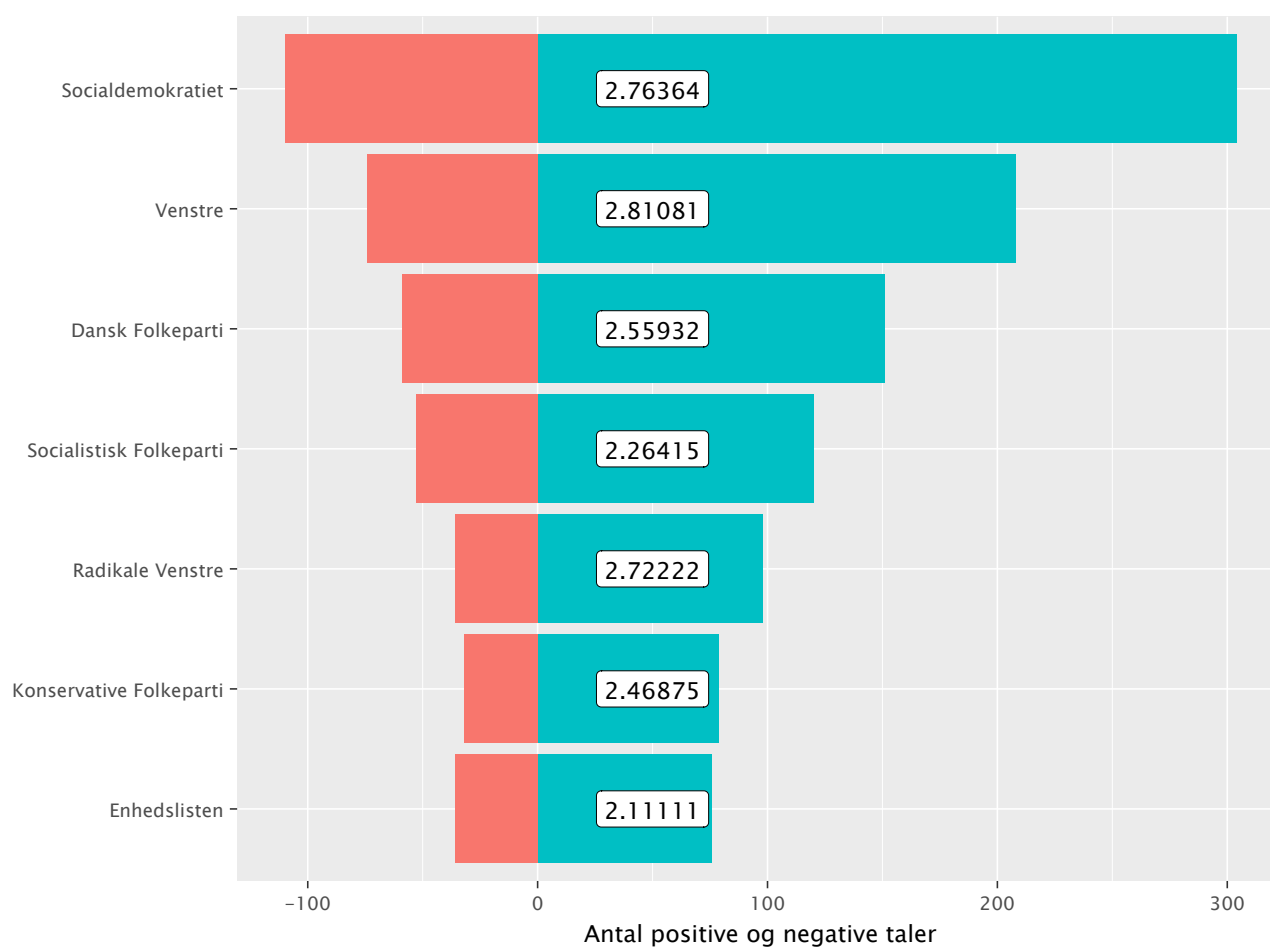
I de mere neutrale begreber ser vi, som forventet, flere almindelige fraser. De sociale problematikker er også med her; og der skal være “*adgang*” og muligheder.

15.2 Hvordan taler politikere om uddannelse?

Ovenfor har jeg givet en visualisering af de politiske partiers relative positioner omkring uddannelse. Der var måske ikke de store overraskelser, men jeg fandt det påfaldende, at se hvordan uddannelses- og socialpolitik fletter sig ind i hinanden.

Dette siger dog ikke noget om *tonen* i talerne. Ved hjælp af biblioteket *SENTIDA* udarbejder jeg en holdningsanalyse, hvor jeg tildeler talerne en holdningsværdi; både for talen i sin helhed og for de individuelle sætninger talene er bygget op af. Er talen eller sætningen overvejende positiv; får den en positiv værdi, og er den negativ, gives der en negativ værdi¹⁰. Dermed kan man undersøge, hvilke grupperinger i Folketinget taler overvejende positivt eller negativt. Jeg vil i det følgende se efter iøjnefaldende mønstre til videre undersøgelse, og derefter trække enkelte taler og sætninger ud til gennemgang.

¹⁰Der udarbejdes to værdier, for at være præcis: en gennemsnitlig holdningsscore for sætningen, og en score for sætningens totale holdningsværdi. Jeg har derudover beregnet minimums- og maksimumsværdier for de enkelte sætninger i talerne; samt en gennemsnits- og totalværdi for talernes holdningsscore som helhed.



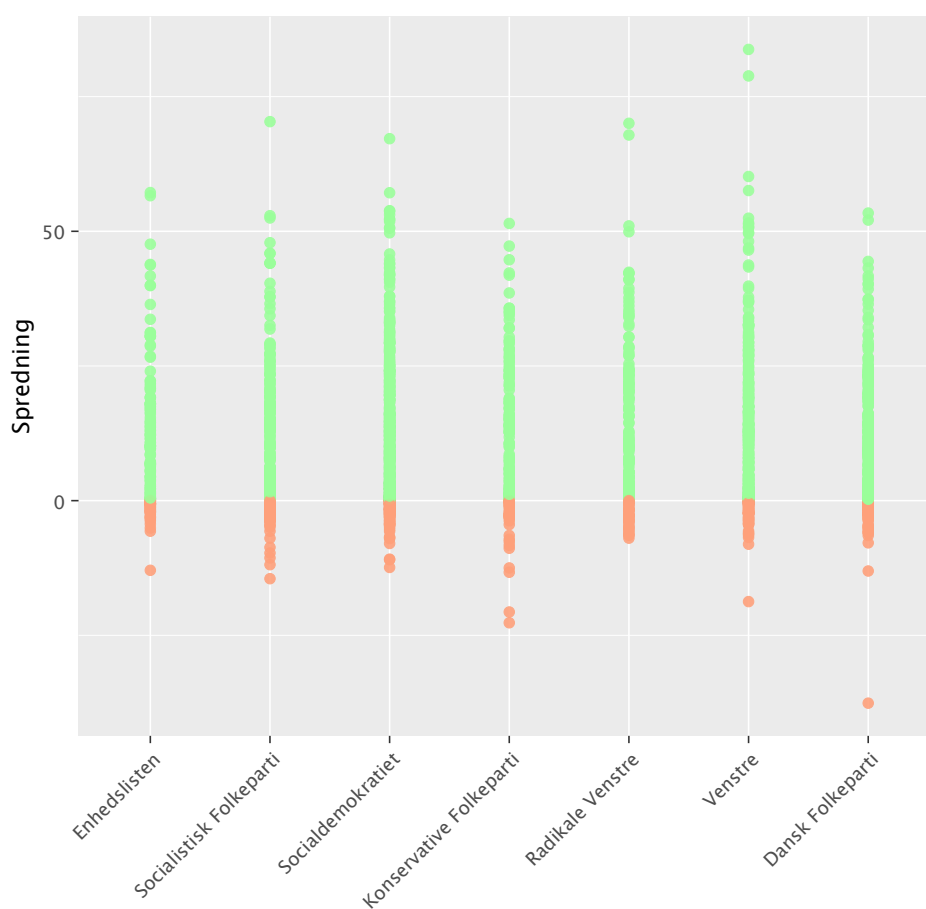
Figur 15.6: Antal positive og negative sætninger i partiernes taler

Jeg lægger ud med en gennemgang af antallet positive og negative sætninger i politikernes taler. Udgangspunktet for beregningerne er mit uddrag af data, der (formodentlig) har med ungdomsuddannelser at gøre.

Figur 15.6 er en optælling af antallet positive og negative sætninger i talerne i mit (ungdoms)uddannelses-datasæt. Her kan man også se, hvilke partier har haft mest på hjertet; da flere og længere taler ganske naturligt vil give flere sætninger. Men med *forholdet* mellem positive og negative taler med i figuren er der muligvis indsigter at høste. Partierne ligger alle omkring 2,5 i større eller mindre grad; med *Venstre* og *Socialdemokratiet* som dem, der har mest på hjertet om uddannelse (og relaterede emner) - både positivt og negativt; tæt fulgt af *Radikale Venstre*. *Enhedslisten* har til gengæld størst andel positive taler - men færrest taler i absolutte tal. For at se på, hvad der ligger bagved denne fordeling, er næste trin at se på holdningsværdierne, og det tekstmateriale der ligger til grund for dem.

I figur 15.7 arbejder jeg videre med mine holdningsanalyser. Figuren viser partiernes¹¹ spredning mellem lave og høje totalværdier for sætningernes holdningsscore, fordelt over mine analy-

¹¹Jeg holder mig til det udvalg af partier jeg har benyttet mig af tidligere



Figur 15.7: Taler fra (ungdoms)uddannelses korpus, spredning mellem sætningernes yderligste holdningsværdi.

seperioder som helhed.

Som udgangspunkt ser det ud til, at (ungdoms)uddannelse overvejende omtalt i en positiv kontekst. *Venstre* udmærker sig med den højeste totalværdi; og *Dansk Folkeparti* har den mest negative sætning omkring uddannelse.

For at undersøge, hvad der (tilsyneladende) virker polariserende i dansk politik, trækker jeg talerne ud der ligger i top tre for høj og lav total holdningsværdi¹² for hvert parti, sammen med den tale de er udledt af. Det viser sig igen, at uddannelsespolitik er tæt knyttet til socialpolitik, hvor flere af sætningerne er fra taler knyttet til kriminalitet eller arbejdsmarkedspolitik. Et blik på indholdet af udvalgte taler viser, at en positiv eller negativ holdningscore ofte er en respons på egne eller andres politikker¹³. Dette vil jeg nu uddybe med nogle eksempler, med positivt ladede

¹²Nogle af talerne har en meget høj total score; hvor næsten hele talen er blevet regnet med i beregningen. Dette er tydeligvis en begrænsning i algoritmen, der skal tages højde for hvis disse metoder skal anvendes i anden sammenhæng. Jeg ser dog stadigvæk god analyseværdi i resultaterne; og bringer dem med i videre analyser.

¹³Og også responser til *andre politikere*; og i mindre grad indikerende for en holdning til uddannelse.

sætninger først.

...nøglen til at øge den sociale mobilitet er bl.a., at børn og unge har mulighed for at tilegne sig de færdigheder, der skal til for at påbegynde en videregående uddannelse...

Venstre 78.84

Venstre har, i uddrag, en tale hvor de starter med at være meget enige i et uddannelses-udspil. Men efterhånden som talen skrider frem, handler det lige meget om, at bruge taletiden til at fortælle om *Venstres uddannelsespolitik*. Begreber som "social mobilitet" og "arbejdsmarkedets behov" er gennemgående, alt i et "human resource" perspektiv: Ved at tilrettelægge muligheder til uddannelse og udvikling alle, tjener samfundet på det.

...vi skal have mange flere unge mennesker til at tage en erhvervsuddannelse, og det skal vi ikke kun, fordi det er vigtigt for vores samfund, at vi også har faglærte i fremtiden, men fordi det er rigtig fedt at tage en erhvervsuddannelse...

Socialistisk Folkeparti 52.90

Dette uddrag fra en længere tale fra *Socialistisk Folkeparti* understreger behovet for at flere søger erhvervsuddannelser. Der anerkendes, at samfundet har behov for faglærte; men hovedargumentet er, at gøre det *attraktivt* at tage en erhvervsuddannelse.

...hvad vil det nemlig sige, at man udvikler den unges personlige kompetencer...

Dansk Folkeparti 44.41

Denne sætning fra *Dansk Folkeparti* er, trods den høje score, en længere tirade overfor konceptet "den fri ungdomsuddannelse". Der bliver leveret en række positivt ladede ord; men i en sarkastisk kontekst. En holdningsanalyse er dermed kun et skridt på vejen til indsigter i politikernes faktiske holdninger

Jeg går nu over til sætninger med høj negativ totalværdi.

...senest er det dokumenteret, at der er en klar sammenhæng eller i hvert fald et klart sammenfald imellem de elever, som har svage boglige forudsætninger og har fået dårlige karakterer ved 9.-klassens-prøven, og så de elever, som falder fra i erhvervsuddannelserne og aldrig får en uddannelse.

Venstre —8, 08

Venstre kommenterer klart og nøgternt, hvad der er af problemer at løse omkring ungdomsuddannelse. Problemerne — der i øvrigt, skal der påpeges, forhindrer tilknytning til arbejdsmarkedet — er, ikke overraskende, negativt ladet.

det forekommer også, at fanden læser bibelen bedre, end hr. frank dahlgaard læser undersøgelsesrapporten. f.eks. konkluderer rapporten, at der ikke er nogen sikker indikation af, at problemet er stigende

Enbedslisten —5, 67

Denne sætning fra *Enbedslisten* er et meget tydeligt eksempel på, at politikere bruger taletid på, at stille andre politikere i et dårligt lys.

vi lader regelsættet stå, indtil det giver problemer, og når det giver problemer, så afskaffer vi salmesang og fadervor.

Dansk Folkeparti —7, 83

Dansk Folkeparti har igen en sætning (og en tale) i en sarkastisk ånd, hvor der overhovedet ikke er problemer med *kristne* bønne til morgensamlinger. En muslimsk bøn, til gengæld, vil skulle betyde afskaffelse af kirkebesøg til højtider og konfirmationsundervisning i folkeskolen; ja, nærmest en aflysning af dansk kultur.

En totalværdi vil dog kunne give et delvist misvisende billede; da lange sætninger vil, som vist, have større mulighed for en højere score. Figur 15.8, der viser spredning i den gennemsnitlige holdningsværdi for talernes sætninger, ligner i første øjekast figur 15.7. Men når vi kigger nærmere på figuren, er der vigtige forskelle. Andelen taler der har en negativ holdningsværdi er nu markant højere, og der er flere taler der i større grad afviger fra klyngen i midten.

Det betyder dog ikke, at informationstætheden er blevet højere. Især ved høje positive gennemsnitsværdier bliver høflighedsfraser altoverskyggende, som i dette eksempel fra *Venstre*:

tak for det

Venstre 3.17

Vel er høflighed en dyd; men det siger givetvis ikke ret meget, når det er en fast indledning til de fleste taler. En analyse med et lille udvalg stopord ville muligvis være værdt at afprøve. I sætningerne med negativ gennemsnitlig værdi er der mere at komme efter. Der er, også her, flere andre sociale problematikker end uddannelse at se; men ikke dermed sagt, at uddannelse ikke bliver nævnt.

det gør vejen videre endnu sværere for dem, for så har vi påført dem nederlag.
Enhedslisten, -1, 69

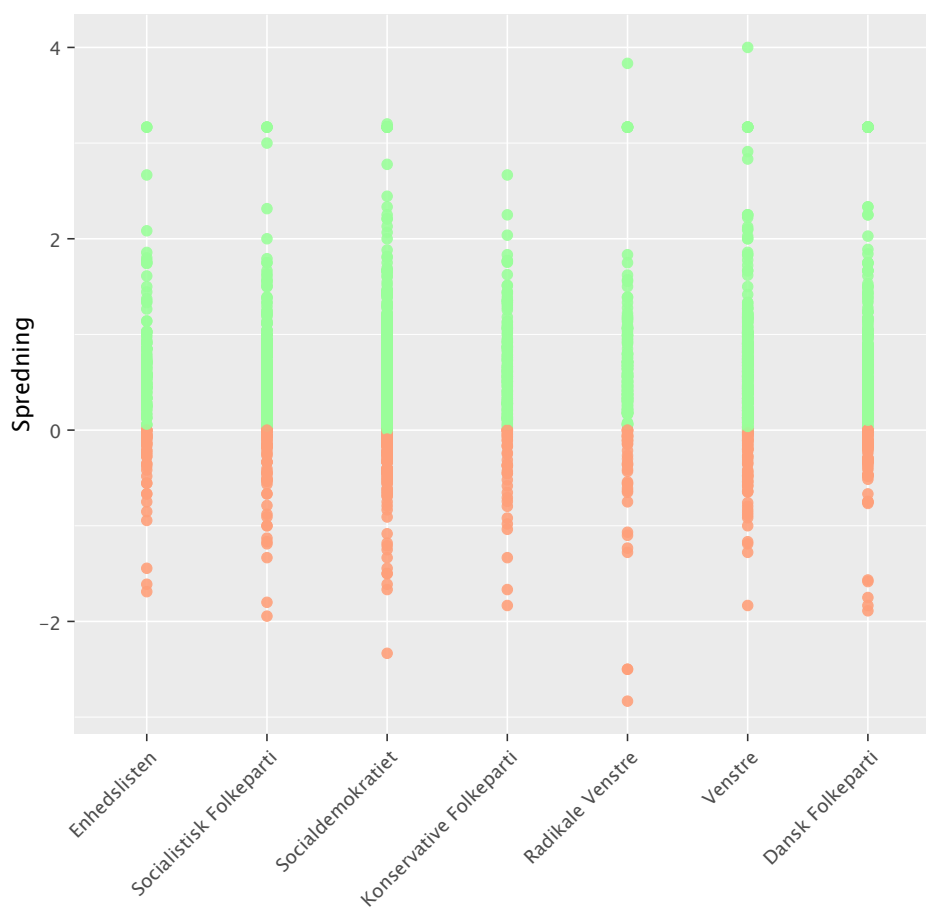
så sætter vi nogle strikse regler ind, og der er straf og sanktioner.
Socialistisk Folkeparti, -1, 96

Både *Enhedslisten* og *Socialistisk Folkeparti* har sætninger der er negativt ladede; i taler der kritiserer gældende regler for ungdomsuddannelserne. *Enhedslisten* advarer mod, at presse unge til en uddannelse de ikke er klar til; og *Socialistisk Folkeparti* harselerer over at sætte sanktioner ind i stedet for at skabe motiverende undervisning. Jeg nævnte også den aktive socialpolitik og et fokus på at få flere i uddannelse som nogle af erhvervsuddannelsernes udfordringer i senere tid; noget det lader til, ikke er gået alle politikere forbi.

Ved holdningsanalyserne på taleniveau vælger jeg ikke, at bruge talernes totale holdningsværdi; da den store mængde ord giver voldsomt høje tal. Jeg vil dog se nærmere på talerne med de højeste og laveste gennemsnitsværdier. Som figur 15.9 viser, har den store mængde ord forskøvet fordelingen af taler opefter; hvor meget få taler har en score under 0; og de fleste taler er klumpet sammen omkring 0, 2 – 0, 7. Der er dog nogle afstikkere; både positiv og i negativ retning; og jeg vil gennemgå et udvalg¹⁴ af disse nu.

det virker, som om det er lidt af en sygdom, der er udbredt i dag. ordføreren fik ikke svaret på det spørgsmål, jeg stillede, for jeg spurgte jo: bliver vi i danmark klogere eller dummere med det finanslovsforslag, der her er fremsat? vil den konservative ordfører ikke nok svare på det: bliver vi klogere eller dummere af, at vi nu skal skære massivt på voksen- og efteruddannelsen for de danskere, som har fået mindst uddannelse fra starten? bliver vi klogere eller dummere af, at vko nu tvinger kommunerne ud i voldsomme besparelser på vores børns folkeskole? bliver vi klogere eller dummere af, at man skærer 3,8 mia. kr. på uddannelse og forskning? bliver vi klogere eller dummere af, at der står 9.000 unge mennesker og mangler

¹⁴Jeg trækker teksten ud af mit (delvist - alle ord er med små bokstaver) ubearbejdede korpus; med småfejl som følge af PDF-til-tekst processen.



Figur 15.8: Taler fra (ungdoms)uddannelses korpus, spredning mellem sætningernes gennemsnitlige holdningsværdi.

en praktikplads, hvor regeringens svar er, at man nu skærer yderligere 1.500 skolepraktikpladser? bliver vi klogere eller dummere af det? spørger jeg hr. mike legarth. det er sådan set et ja- eller nejspørgsmål *Enhedslisten*, -0, 02

Enhedslisten går nok en gang i kødet på andre politikere. Der er besparelser på vej, og der søges svar på, om der er tænkt følgevirkninger og konsekvenser igennem. Spørgsmålene er alle stillet fra et perspektiv omkring social uligning. Der lader til at være en holdning om, at staten har et ansvar for at stille muligheder op for de borgere, der ikke har de bedste forudsætninger

vi afskaffer ikke skolepraktikordningen, medmindre vi har noget brugbart at sætte i stedet, og det er selvfølgelig, fordi vi under ingen omstændigheder vil lade de unge mennesker i stikken. de unge mennesker skal have en uddannelse. skolepraktikken er ikke den mest ideelle måde at give de unge mennesker en uddannelse på, men vi skal holde fast i, at de trods alt får en uddannelse, medens de er i et

skolepraktikforløb. med hensyn til hvad vi kan sætte i stedet, eller hvordan vi kan forbedre unges uddannelsesmuligheder, er vi som nævnt ved at se på det i forbindelse med handlingsplanen bedre uddannelse», og det vil jeg gerne have lejlighed til at vende tilbage til.

Venstre, 0, 87

Venstre vil gøre det bedste ud af en skidt situation, og vinder en ret positiv score på den baggrund. Det kan godt være; at en skolepraktikordning er som at komme på B-holdet kontra en "rigtig" praktikplads (jævnfør erhvervsuddannelsernes udfordringer med at stille nok praktikpladser til rådighed; og det langvarige spændingsforhold mellem skoleundervisning og mesterlære i erhvervsuddannelsernes historie), men en skolepraktikplads er trods alt bedre end slet ingen praktikplads. Ikke?

det konservative folkeparti støtter naturligvis også forslaget. det er meget vigtigt, at vi får hævet erhvervsuddannelsernes anseelse. det er vigtigt, at vi får tiltrukket flere talentfulde unge til erhvervsuddannelserne. det er afgørende både for dansk økonomi og for virksomhederne og også for de unges personlige udvikling, at de får valgt den rette hylde, hvor deres talenter kan udfolde sig. der er alt for få unge mennesker, som er opmærksomme på, at erhvervsuddannelserne er nogle rigtig gode uddannelser. for at vi kan tiltrække talenter, er det vigtigt, at erhvervsuddannelserne ikke bliver opfattet som en blindgyde, og at man kan læse videre fra en erhvervsuddannelse. det er klart, at hvis man er et meget talentfuldt og ambitiøst ungt menneske, vil man gerne have, at der er muligheder for at videreudanne sig, og at man ikke nødvendigvis skal gå over det almene gymnasium for at kunne læse videre. vi så alle sammen for nylig en undersøgelse, der viste, at der var en høj grad af dobbeltuddannelse, hvor der var unge mennesker, der valgte først at læse på gymnasiet og derefter tage en erhvervsuddannelse. det her forslag kan jo rette lidt op på det, fordi det vil betyde, at de unge mennesker ikke føler, at det er nødvendigt at læse på gymnasiet for at have en chance for at komme videre i uddannelsessystemet. derfor støtter vi meget varmt det her forslag. vi synes, det giver nogle rigtig gode muligheder. det giver mulighed for at få flere ambitiøse unge mennesker ind på erhvervsuddannelserne, og det har vi brug for. samtidig håber vi selvfølgelig også, at vi vil fortsætte det oplysende arbejde omkring erhvervsuddannelserne og åbne flere unge menneskers øjne for, at det er en rigtig god vej at vælge.

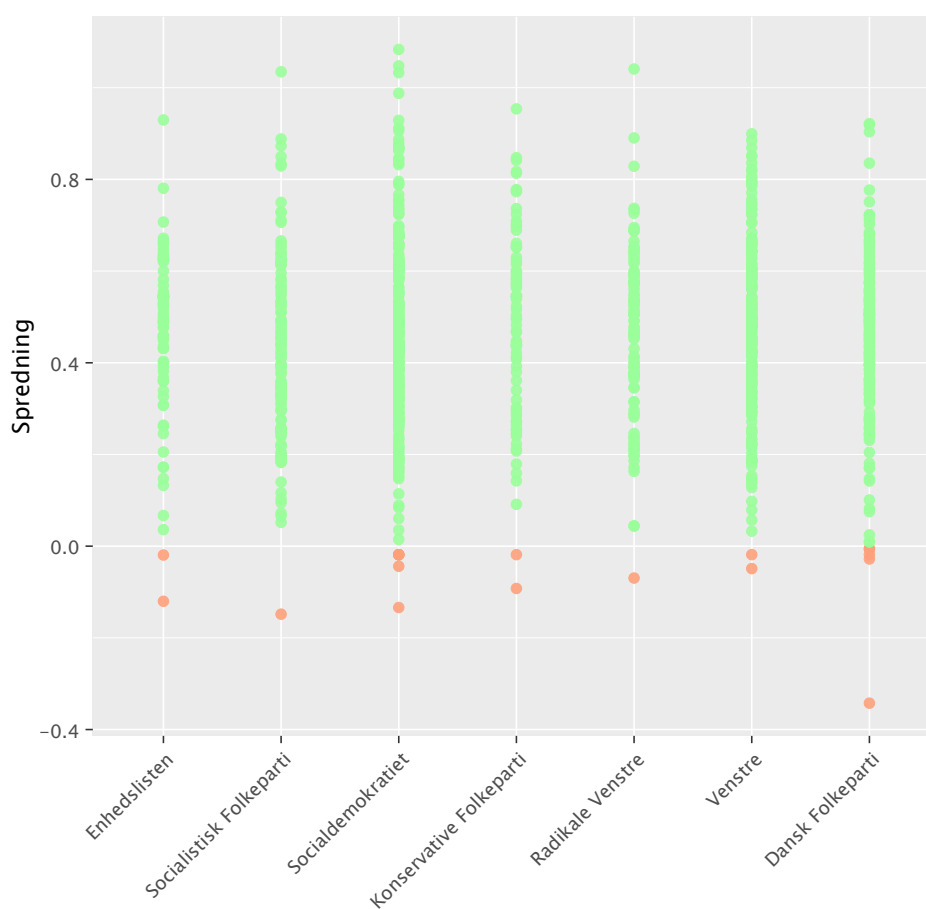
Det Konservative Folkeparti, 1, 03

Denne tale fra "Det Konservative Folkeparti" er et meget illustrerende eksempel på nogle af de udfordringer erhvervsuddannelserne har stået overfor i senere tid¹⁵. Man skal imødekomme arbejdsmarkedets behov for kvalificeret arbejdskraft. Og det skal være attraktivt for de "talentfulde unge" at søge ind på erhvervsuddannelserne; der ikke skal være for "en blindgyde" at regne. Dermed er det en god idé, at erhvervsuddannelserne kan give adgang til videregående uddannelser; således at man kan undgå at de unge dobbeltuddanner sig.

Men dog. Nu begynder politikere at sadle flere heste på en gang; hvorfor man får reform efter reform; for at afbøde den forrige reforms utilsigtede konsekvenser. Skal erhvervsuddannelserne ikke også være en vej til arbejdsmarkedet for de mindre ... bogligt engagerede ... unge? Og hvad med erhvervsuddannelsernes betydning for efter- og videreuddannelse af voksne?

Ud over disse eksempler, der er blevet fremhævet grundet deres relevans for mine undersøgelsesspørgsmål, er der de efterhånden velkendte, beslægtede sociale problemstillinger der er

¹⁵Talen er fra 2010.



Figur 15.9: Taler fra (ungdoms)uddannelses korpus, spredning mellem talernes gennemsnitlige holdningsværdi.

gennemgående. *Dansk Folkeparti* vil holde opsyn med de muslimske friskoler. Alment boligbyggeri er et tilbagevendende emne, ligesom arbejdsmarkedspolitik og offentlig understøttelse og (ungdoms)kriminalitet.

Politikerne spiller det politiske spil

Denne gennemgang af en holdningsanalyse over taler omkring uddannelse¹⁶ har givet god anledning til at besvare mine undersøgelsesspørgsmål. Der synes dog også være tydeligt at se, at politikerne ikke holder sig tilbage for, at spille sit politiske spil. Kan man bruge en kommentar til et politisk udspil til at fremhæve sit partis *aldeles fremragende* politiske ståsteder, lader man ikke øjeblikket smuldre mellem fingrene. Ligeledes, hvis man kan komme med et spidsfindigt verbalt sværdstik; eller kan fremhæve sin pointe med en sarkastisk bemærkning; er det klart den mulighed man griber.

¹⁶Og, igen, en række beslægtede socialpolitiske problemstillinger.

Politikeres holdninger til uddannelse over tid

Jeg vil nu undersøge, om der er information at hente i at se på udviklingen af politiske holdninger til uddannelse over tid. Figur 15.10 viser hvordan politikeres holdning til mit bredere uddannelseskorpus¹⁷ fordeler sig over tid. Trods mit valg af det bredere uddannelseskorpus, er der med det samme meget tydeligt, at der er en mangel på datapunkter i 1990 – 01. Et yderligere blik på diagrammet gør det også tydeligt, at langt de fleste ekstreme observationer ligger i perioderne 2001 – 14 og 2014 – 20.

Dette er selvfølgelig vigtigt at have med i overvejelserne, når man forholder sig til min gennemgang af holdningsanalyser ovenfor. Den store mængde datapunkter for de senere perioder overskygger i denne sammenhæng de tidligere; og giver et noget unuanceret billede for hvad der har været polariserende i uddannelsesdebatten. Uden at gentage hele min holdningsanalyse for hver periode, vil jeg nu lave udtræk af talerne med lavest og højest gennemsnitlig holdningsværdi¹⁸ for hver periode, og kort drøfte disse.

I 1978 – 90 er der især fokus på erhvervsuddannelserne. *Fremskridspartiet* tordner mod uansvarligt ført uddannelsespolitik, uden en *“balance mellem udgift til uddannelsesforbrug og de indtægte vores erhvervsliv har haft mulighed for at skabe”* i en tale med en gennemsnitlig holdningsscore på -0.21 . *Radikale Venstre* taler varmt (score 0, 69) om et fokus på HF, hvor der må *“være en vigtig målsætning at hf, der ikke mindst søges af kvinder, er en almen og studieforberedende uddannelse og en reel alternativ til studentereksamen”*

For perioden 1990 – 01 taler *Socialdemokratiet* imod *Venstre* i kontekst af religionsundervisning i folkeskolen: *“jeg synes man skulle læse grundtvig en gang mere i venstre hvis orde livsoplysning foran kristendomsundervisning kunne skabe så stor kvababbelse”*; med den laveste score for perioden på 0, 04. Dette er nok et eksempel på, at en lav holdningsscore nogle gange siger mere om konteksten for talerne end politikernes synspunkter omkring et emne. Blant de positive taler for perioderne er der repræsentanter fra flere partier der taler varmt om en udvikling af erhvervsuddannelserne; i kontekst af AMU-centre, den fri ungdomsuddannelse og erhvervsakademierne.

Når vi kommer til 2001 – 14, er der mange gengangere fra min første gennemgang af holdningsanalyser. I forlængelse af denne trend, fremlægger *Socialdemokratiet* at realkompetencer fra produktionsskoler burde anerkendes med et kompetencebevis.

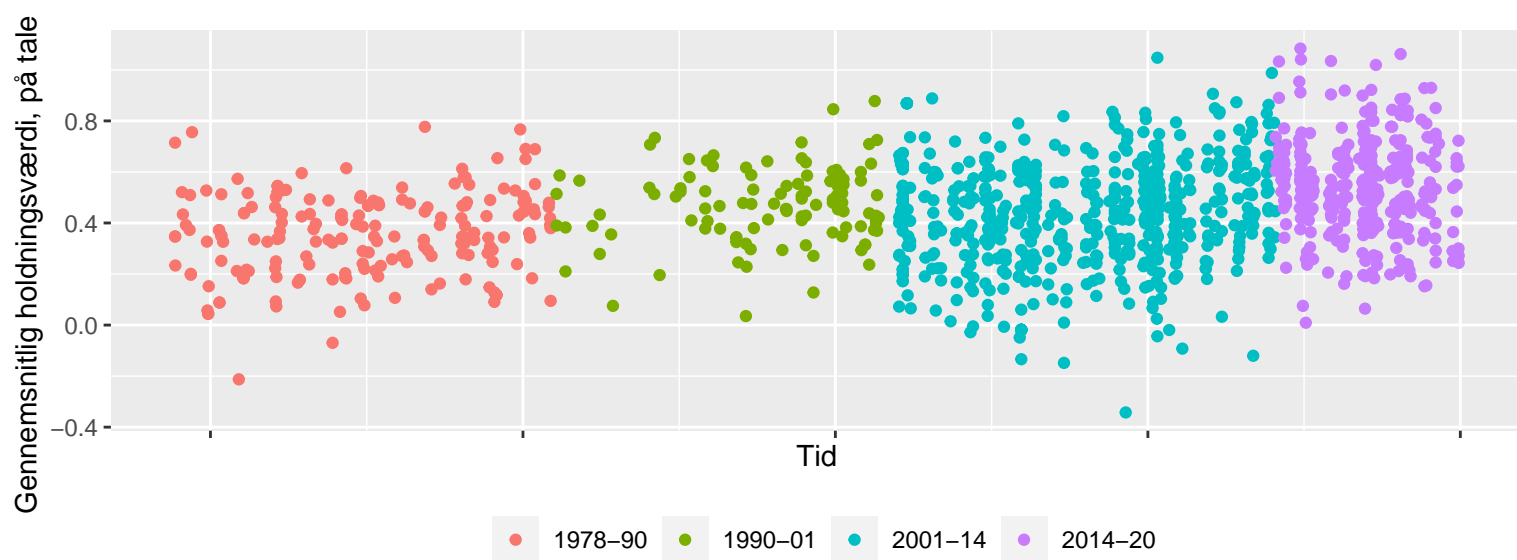
Afslutningsvis, i 2014 – 20, ser vi eksempelvis *Dansk Folkeparti* efterspørge et *“skærpe tilsyn vi have jo forsøge at få en høring op at stå om bla muslimsk friskole og friskole generelt”*, med en score for hele talen på 0, 01. Og blandt de positivt ladede taler, er *Liberal Alliance* og *Alternativet* begge varme fortalere for, et frit skolevalg i Danmark.

Og ja, igen er der en pæn samling af drøftelser omkring andre sociale problematikker — i alle perioderne.

Helt overordnet understreger denne lille øvelse, at det er vigtigt at undersøge sit datagrundlag over flere dimensioner. Denne historiske gennemgang af holdningspositioner giver et mere nuanceret og bredt billede, end en snæver gennemgang af partiernes yderpositioner.

¹⁷Jeg har valgt det mere generelle uddannelseskorpus til denne undersøgelse, for at imødekomme den noget ujævne fordeling af observationer der forekommer for de tidlige perioder.

¹⁸for at være helt præcis, vil jeg trække taler med en score der er indtil en 3. plads fra top og bund for gennemsnitlig holdningsværdi; for at sikre mig et relevant uddrag af data.



Figur 15.10: Spredning over talernes gennemsnitlige holdningsniveau. Taler er organiseret efter mødedato, og er udvalgt på baggrund af min oversigt over politiske blokke.

Del V

Konklusion

It is, I think, the political task of the social scientist who accepts the ideals of freedom and reason, to address his work to each of the other three types of men I have classified in terms of power and knowledge.

To those with power and with awareness of it, he imputes varying measures of responsibility for such structural consequences as he finds by his work to be decisively influenced by their decisions and their lack of decisions.

To those whose actions have such consequences, but who do not seem to be aware of them, he directs whatever he has found out about those consequences. He attempts to educate and then, again, he imputes responsibility.

To those who are regularly without such power and whose awareness is confined to their everyday milieu, he reveals by his work the meaning of structural trends and decisions for these milieu, the ways in which personal troubles are connected with public issues; in the course of these efforts, he states what he has found out concerning the actions of the more powerful.

C. Wright Mills, *The Sociological Imagination*

Computation allows us to model, measure, and modify both social structures and the texture of individual experience — and to do so on bigger and smaller scales than ever before. The sociological imagination has been blown wide open. We must not forget that the new possibilities unleashed by the digital age — formal worlds and digital observatories, intelligent surveys, virtual laboratories, and machine discovery — require the sociological imagination to achieve their full potential.

James Evans og Jacob G. Foster, „Computation and the Sociological Imagination“
(2019)

16 Store data – store indsigter?

Jeg har vist, hvordan politikere har forholdt sig til (erhvervs- og ungdoms)uddannelse har artet sig fra 1978 til 2019.

Jeg har beskrevet hvordan generelle tendenser i uddannelse, der fremgår af historiske opgørelser kan spores i politikernes taler. Arbejdsmarkedets parter forekommer sammen med erhvervsuddannelser i 80'erne. På 90-tallet tales der om forskning og naturvidenskab. I 00'erne er der nationale tests på banen. Fra 2014 er det et gennemgående fokus; at man starter på **og fuldfører** en uddannelse.

Jeg har gået dybere i detaljen, og undersøgt hvordan positioner i politisk diskurs omkring (ungdoms)uddannelser har bevæget sig over tid. Jeg har også set et mønster, der afføder en hypotese: forskelle i positioner *kan* også handle om, at man “taler magten imod”. Men, som sagt, så vil det kræve et andet greb på data og teori. Når blikket vendes mod indholdet i de politiske positioner, bliver billedet nærmest karikeret; hvor venstreladede begreber har rummelige konnotationer; og højreladede begreber holder fokus på at sætte innovation og vækst i gang. Hvis man holder partiernes politiske udspil fra deres partiprogrammer op mod deres udtalelser i folketingssalen, er billedet lidt mere mudret og uklart. Forholdet mellem partiprogrammer og deres politikeres taler, lader til at være en vej fremad for fremtidige frugtbare analyser i det politiske felt.

Jeg har også set på, hvordan politikerne *taler om* uddannelse. Gennem holdningsanalyser viser jeg, at politikerne tilsyneladende taler om uddannelse i overvejende positive vendinger. Når man graver lidt rundt i mudderet, er det dog ikke så entydigt. Politisk taletid vendes gerne til egen fordel; eller tages som en åbning for et skarpt retorisk sværdstik mod politiske opponenter.

Når jeg vender blikket mod erhvervs- og ungdomsuddannelserne, er der som sådan opmærksomhed på, at der er problemer og udfordringer at løse og overvinde. *Hvem* der står i problemer, og *hvordan* disse udfordringer skal overvindes, er der dog mindre enighed om. Er der tale om strukturelle uretfærdigheder, som det offentlige skal bygge stilladser således de er til at bestige? Eller handler det om at stille muligheder til rette for den enkelte, så man ikke spilder menneskelige ressourcer der ellers kunne komme samfundet til gode? Erhvervsuddannelserne skal tilgodeses til de ...mindre bogligt engagerede ..., der alligevel kan have “hænderne godt skruet på”, og dermed opleve, at de har sammenhæng og mestring i deres liv. Men samtidig skal der være plads til den engagerede og dygtige erhvervsskoleelev, der ikke skal risikere at komme i en uddannelsesmæssig blindgyde, men have muligheder for videre uddannelse. Dette er noget af en spaltet personlighed at imødekomme; for ikke at tale om videre- og efteruddannelse af voksne.

Når jeg inddrager et tidsperspektiv på min holdningsanalyse, bliver det tydeligt, at et ensidig fokus på én dimension – som ovenover, hvor jeg kun kiggede på absolutte yderværdier – kan medføre, at man går glip af vigtige nuancer og historiske kontekster.

16.1 Tekst som produkt – og som data

Dette har været en (delvist) kvantitativ tilgang til et kvalitativt udgangspunkt. Dette giver muligheder; primært i forhold til mængden af tekst jeg har kunnet forholde mig til. Computeren er også i mindre grad plaget af forudindtagelser og bias end jeg¹; og har dermed muligheder for at se mønstre og tendenser der kunne forbigå et sølle menneske.

¹ Eller rettere; det er mig som forsker, der tilfører algoritmerne eventuelle bias.

Det var for eksempel slående for mig, i hvor høj grad uddannelse er indlejret i socialpolitikken. Det være sig ungdomsboliger, langtidsledige eller unge kriminelle; er afstanden til uddannelse i datamaterialet ikke ret stor. Den logiske slutning er heller ikke svær at trække. Gennemgribende ulighedsforhold vil fordrer gennemgående indsatser, der skruer på flere parametre samtidigt. Om der er taget højde for alle mulige og umulige utilsigtede konsekvenser af en gennemgribende indsats er dog en anden diskussion.

Men, som nævnt i del III, er denne kvantificering af det kvalitative fordrende en homogenisering af materialet inden analyse. Tabet af ekstremerne er et muligt tab af information; der skal tages højde for i fortolkning af data.

Den tekst jeg har kværnet igennem min algoritmiske kødhakker har, inden det blev transkriberet, startet som politisk tale; med helt særlige institutionelle rammer. Det er, for eksempel, et specielt taleprodukt i forhold til hvem den henvender sig til. Når man taler om “uddannelse for alle” eller “uddannelse i verdensklasse”, for eksempel, henvender man sig samtidigt til andre politikere; sine vælgere; og offentligheden i almindelighed. Her kan ekstreme observationer være vigtige; da de peger på, hvad der eventuelt “skiller vandene” i den politiske debat.

16.2 Værdien af en beskrivende sociologi

Jeg refererede i min indledning til Savage (2009, der går fuldt ind for en deskriptiv sociologi. Disse bemærkninger er ikke gået ubemærkede hen; og Gane advarer mod at tabe dybden i den sociologiske forskning — forklaringer, analyser, kausalitetsspørgsmål — i en deskriptiv iver.

Jeg har tydeligt situeret dette speciale i en *beskrivende* sociologisk tradition. Dette — meget bevidste — valg giver mig en frihed til, at se på *bredden* i mit empiriske grundlag. Jeg kan påpege tendenser i politiske ståsteder; jeg kan udforske trends i holdning og tone blandt politikere; jeg kan udforske klynger af emner der optræder sammen; jeg kan se hvordan de politisk centrale emner ændrer karakter.

Dette ser jeg som et vigtigt og værdifuldt produkt i sig selv. Ikke mindst lige vigtigt er mulighederne et velbeskrevet felt giver for videre analyse. Men, som Mills understreger, er beskrivelser i sig selv ikke nok. Der er en politisk opgave, mener han, at oplyse både magthavere og de magteløse om, hvad kollektive beslutninger kan betyde for den enkelte (Mills, 2000, s. 185). Men beskrivelsen er ikke nok for Mills. Han understreger også, at det er agtpåliggende at få blotlagt de kausale forbindelser mellem personligt miljø og sociale strukturer (2000, s. 130). En beskrivelse kan ikke stå alene — der skal også gerne være en forklaring på et tidspunkt.

Her er der rigelige muligheder for at fortsætte i *dybden*, og uddybe forklaringerne. Som J. Evans og Foster beskriver; så har den algoritmiske sociologi slået dørene for den sociologiske fantasi åben på vidt gab; og det er bare at gå i gang.

16.3 Beskrive – på baggrund af hvad?

På et meta-niveau er det nyttigt for videnskaben, at facilitere og forhåndsbehandle data, og dele dette arbejde med verden. Jeg har selv nydt godt af andres arbejde i at gøre grovarbejdet i, at gøre Folketingstidendes datasæt tilgængeligt for videre bearbejdning. Jeg har dog, som bemærket flere steder i mit arbejde, også stødt på flere mangler og gråsoner i løbet af arbejdet med dette speciale. Jeg vil meget gerne selv give tilbage til verden, og bidrage til, at imødekomme dette. Godt forskningsarbejde er nemmere med et godt datasæt at arbejde med.

Min kode er også tilgængelige for offentligheden ([andersenNorseghostMasterThesis2020](#)); til frit skue og inspiration. Dermed håber jeg, at kunne bidrage til den kollektive viden og forståelse af vores fælles sociale verden.

17 Fokus for videre forskning

17.1 Metodologiske greb

Det vil være spændende, at fortsætte computeranalyserne med andre modeller for at udlede emner. LDA er en af de velkendte tilgange til *topic modeling*, men slet ikke det eneste mulige værktøj. At arbejde med en anden tokenisering vil også være værd at udforske — trigrams og skip-grams, for eksempel, der blev udelukket i denne omgang af tidshensyn. Og man kan også arbejde i højere grad med analyser på de enkelte sætninger.

Som tidligere nævnt, vil en videre udforskning af politiske positioner i og udenfor Folketinget også være spændende.

I holdningsanalyserne havde jeg til en vis grad en hemske i form af ufuldstændigt bearbejdet datamateriale. Jeg står ved mine analyser og mine overvejelser; men det ville formentligt være givende at gentage processen på et andet datagrundlag.

Det er også muligt at arbejde, vel, *algoritmisk*, på et andet niveau. Computeren er god til at fremhæve mønstre; og kan udarbejde modeller til at arbejde *forudsigende*¹. På baggrund af tidligere observationer (af computeralgoritmer eller af mennesker) kan man beregne sandsynligheden for at en givet udtalelse kom fra et specifikt parti eller en specifik politiker — eller omvendt; foreslå hvad politiske fløje/partier/politikere ville have af holdninger til bestemte emner.

17.2 Andre fokusområder

Som jeg var inde på i specialets forrige del, var der flere emner indenfor uddannelse man også kunne grave dybere i, ud over det gennemgående skæringspunkt med socialpolitik generelt. Forholdet til tosprogede elever, for eksempel; og hvordan det flugter med andre sociale indsatser. Eller noget så oplagt som folkeskolen, og omtalen af denne.

Det var også muligt at antyde emner, der kunne have interesse for andre forskningsfelter. Gennemgående emner for hver analyseperiode har for eksempel været

- Skat, erhverv, landbrug, skibsfart, (over tid grøn og vedvarende) energi
- Socialpolitik, kriminalitet, stofmisbrug
- Danmarks forhold til omverdenen: EU; Norden; mellemøsten; flygtninge og indvandrere

Her er der rigeligt med muligheder for at danne nye indsigter med lignende metoder. Der var, for eksempel, meget tydeligt at se hvordan omtalen af indvandrere og flygtninge ændrede karakter; fra begreber som *dansk flygtningehjælp*, *røde kors* i 1978-90 over i *international forpligtigelse*, *kriminell udlænding* i 2014-20. En let dramatisering giver en udvikling fra “der kommer nogen”, til “hvem er **det** for nogen; og hvad *vil* de egentlig her?”².

Der var også en række emner, der næsten kun bestod af navne. Jeg formoder, at dette er antydninger til politiske historier jeg mangler kontekst til at gennemskue. Denne formodning underbygges blandt andet af, at Christine Antorini dukker op i uddannelsesemner; og af at Joachim B. Olsen optrådte sammen med begreber omkring det private erhvervsliv.

¹Eng: *predictive algorithm*; begrebet oversat fremstår mere skråsikkert end mine intentioner.

²Dette er min fortolkning af de fremtrædende ord for emnerne. Det drejer sig om emnerne 26, 9, 31, 11 for de respektive analyseperioder.

17.3 Anvendelighed på andre datasæt

Man kunne også anvende den demonstrerede teknik til at indkredse gennemgående temaer i et stort datamateriale på et andet grundlag. Jeg har over 15 års erfaring indenfor specialpædagogikken; og som del af mit arbejdsliv udfærdiger mine kolleger og jeg utallige spaltmeter af rapporter fra dagligdagen. En analyse af dette datamateriale vil kunne give indsigter i, hvordan den professionelle diskurs omkring mennesker med kognitive handicap arter sig. Man kan også forestille sig, at optegnelser over adfærdsændringer blev fodret til en prediktiv algoritme. Dermed vil computeren kunne finde mønstre, der leder op til uønskede eller uhensigtsmæssige reaktioner hos borgeren³. Med denne viden, vil der være bedre muligheder for at tilpasse det pædagogiske miljø omkring borgeren for at undgå udløsende faktorer for den uhensigtsmæssige adfærd.

Man kunne også rette blikket mod lærebøger i folkeskolen; lovtækste; eller politirapporter. Jeg forestiller mig dog, at *datatilgang* vil være en udfordring i nogle sammenhænge.

17.4 Indholdsanalysen er død. Længe leve indholdsanalysen!

Selv med computeren til hjælp, er denne opgave overordnet en undersøgelse der kun lige kommer under overfladen af kildematerialet; min begrænsede tekstnære diskussion omkring mine holdningsanalyser til trods.

Et detaljeret øje for i hvilket omfang, der er samsvar med de uddannelsespolitiske beslutninger der fremgår af de historiske opgørelser, og de aktuelle politiske debatter har jeg ikke opnået. Men jeg har afdækket tendenser. Kan politikeres forhold til de strukturelle udfordringer i tilgængelighed til forskellige ungdomsuddannelser blive uddybet? Eller forsvinder dette perspektiv i en karikatur af Bernstein og Bourdieu; hvor den dominerende middelklasses perspektiver og levevis tages for givet — og som et selvfølgeligt ideal der skal efterstræbes? En detaljeret diskussion af ovennævnte postulat er (endnu en) opgave i sig selv. Her vil en mere detaljeret nærlæsning, med udgangspunkt i nogle af de taler min analyse peger på, måske kunne give et svar.

³Inspireret af Danielsen m.fl. (2019, hvor en *machine learning* algoritme kunne forudsige behov for fiksering hos psykiatriske patienter.

Bibliografi

- Andersen, M. A.
(2020a, 11. oktober).
Master's Thesis.
Hentet 16.06.2020, fra <https://github.com/norseghost/master-thesis>
- Andersen, M. A.
(2020b, 11. oktober).
Selected assignments at Aarhus University.
Hentet 04.10.2020, fra <https://martinadreasandersen.com/projects/au/>
- Andersen, M. & Madsen, C. S.
(2012).
Tænk Ikke På En Ghetto.
master's thesis, Department of Society and Globalization, Roskilde University.
- Andersen, T. B.
(2019 marts).
Unge uden uddannelse eller beskæftigelse Del I.
Social- og Indenrigsministeriets Benchmarkingenhed.
København.
- Arun, R., Suresh, V., Veni Madhavan, C. E. & Narasimha Murthy, M. N.
(2010).
On finding the natural number of topics with latent dirichlet allocation: Some observations.
I M. J. Zaki, J. X. Yu, B. Ravindran & V. Pudi (Red.),
Advances in knowledge discovery and data mining (s. 391–402).
Springer Berlin Heidelberg.
https://doi.org/10.1007/978-3-642-13657-3_43
- Barron, A. T. J., Huang, J., Spang, R. L. & DeDeo, S.
(2018).
Individuals, Institutions, and Innovation in the Debates of the French Revolution.
Proceedings of the National Academy of Sciences, 115(18),
4607–4612.
<https://doi.org/10/gfsqwt>
- Bauman, Z.
(2000).
Liquid Modernity.
Polity Press.
- Blei, D. M., Ng, A. Y. & Jordan, M. I.
(2003).
Latent Dirichlet Allocation.
J. Mach. Learn. Res., 3,
993–1022.
- Bourdieu, P.
(2004).
The Field of Cultural Production: Essays on Art and Literature.
Polity Press.

- Bøndergaard, G.
(2014).
The historical emergence of the key challenges for the future of VET in Denmark.
Nord-VET – The future of Vocational Education in the Nordic countries.
- Børne- og undervisningsministeriet.
(nodate).
Hvem og hvor mange vælger en gymnasial uddannelse efter 9. og 10. klasse?
Hentet 23.02.2020, fra <https://www.uvm.dk:443/statistik/gymnasiale-uddannelser/elever/soegning-til-de-gymnasiale-uddannelser>
- Canger, T., Hjort-Madsen, P., Kondrup, S. & Neubert, A. (Red.).
(2016 december).
Tema: Erhvervsuddannelser - mellem politiske ambitioner og praktisk virkelighed (Bd. 2016-4).
Foreningen bag udgivelsen af Dansk pædagogisk Tidsskrift.
https://dpt.dk/wp-content/uploads/2020/01/DpT_04_16_ISSUU-2.pdf
- Cao Juan, Xia Tian, Li Jintao, Zhang Yongdong & Tang Sheng.
(2009).
A Density-Based Method for Adaptive LDA Model Selection.
Neurocomputing – 16th European Symposium on Artificial Neural Networks 2008,
72(7–9),
1775–1781.
<https://doi.org/10/dvh9dh>
- Carroll, L.
(1977).
Alice i Eventyrland Og Bag Spejlet (K. Elfelt & M. J. Nissen, Overs.; 4. udg.).
Gyldendal.
- Danielsen, A. A., Fenger, M. H. J., Østergaard, S. D., Nielbo, K. L. & Mors, O.
(2019).
Predicting mechanical restraint of psychiatric inpatients by applying machine learning on electronic health data.
Acta Psychiatrica Scandinavica, 140(2),
147–157.
<https://doi.org/10/ghb4gc>
- Danmarks Statistik.
(2019 maj).
Erhvervsuddannelser i Danmark 2019.
Danmarks Statistik.
København.
Hentet 23.04.2020, fra <https://www.dst.dk/Site/Dst/Udgivelser/GetPubFile.aspx?id=32526&sid=erhvervsuddannelser%20i%20danmark%202019>
- Dansk Folkeparti.
(2020a, 9. oktober).
Dansk Folkepartis Principprogram - Læs Om Partiets Værdier.
Hentet 09.10.2020, fra <https://danskfolkeparti.dk/politik/principprogram/>
- Dansk Folkeparti.
(2020b, 9. oktober).
Uddannelse - Dansk Folkeparti.
Hentet 09.10.2020, fra <https://danskfolkeparti.dk/partiet/bestil-materiale/spoergsmaal-og-svar/12453-2/>
- Danske Gymnasier.

- (2019, 7. februar).
Fuldførelse og karakterer.
 Hentet 23.02.2020, fra <https://www.danskegymnasier.dk/frafald-og-karakterer/>
- Det Konservative Folkeparti.
 (2020, 9. oktober).
Uddannelse og skole.
 Hentet 09.10.2020, fra <https://konservative.dk/politik/politikomraader/uddannelse-og-skole/>
- Deveaud, R., SanJuan, É. & Bellot, P.
 (2014).
 Accurate and Effective Latent Concept Modeling for Ad Hoc Information Retrieval.
Document numérique, 17(1),
 61–84.
<https://doi.org/10/ggbbjh>
- DPU.
 (2001, 4. september).
Nyhedsbrev: Uddannelse kan ikke fjerne restgruppen.
 Hentet 22.02.2020, fra <https://dpu.au.dk/aktuelt/nyhed/artikel/nyhedsbrev-uddannelse-kan-ikke-fjerne-restgruppen/>
- Enhedslisten.
 (2020, 9. oktober).
Uddannelse og forskning > Enhedslisten.
 Hentet 09.10.2020, fra <https://enhedslisten.dk/politikomraade/uddannelse-og-forskning>
- Evans, J. & Foster, J. G.
 (2019).
 Computation and the Sociological Imagination.
Contexts, 18(4),
 10–15.
<https://doi.org/10/ggzxq>
- Evans, J. A. & Aceves, P.
 (2016).
 Machine Translation: Mining Text for Social Theory.
Annual Review of Sociology, 42(1),
 21–50.
<https://doi.org/10/gc6z mh>
- Fels Birkelund, J., Kristian, K. & Reimer, D.
 (In progress).
The structure, causes, and consequences of tracking in the danish education system
 (WorkingPaper Nr. 10).
 Dynamics of Inequality Across the Life-course (DIAL) / Dynamics of Inequality Across the Life-course (DIAL).
- Folketingstidende.
 (nodate).
Om Folketingstidende.
 Hentet 15.06.2020, fra <https://www.folketingstidende.dk/da/om-folketingstidende>
- Gamoran, A.
 (2010).
 Tracking and Inequality: New Directions for Research and Practice.
The Routledge International Handbook of the Sociology of Education.

- Gane, N.
(2020).
Against a descriptive turn.
The British Journal of Sociology, 71(1),
4–18.
<https://doi.org/10/ggd9hr>
- Gergen, K. J. & Mellon.
(2017).
Socialkonstruktionisme og uddannelse (1. udgave.).
Hans Reitzel.
- Griffiths, T. L. & Steyvers, M.
(2004).
Finding Scientific Topics.
Proceedings of the National Academy of Sciences, 101,
5228–5235.
<https://doi.org/10/adf>
- Hansen, E. J.
(2004, 23. januar).
Restgruppen.
Leksikon.Org.
Hentet 23.02.2020, fra <https://www.leksikon.org/art.php?n=4940>
- Hjorth, F.
(2018).
Establishment Responses to Populist Challenges: Evidence from Legislative Speech.
Ida Juul - Publikationer - Forskning - Aarhus Universitet.
(nodate).
Hentet 24.02.2020, fra <https://pure.au.dk/portal/da/persons/ida-juul/f6339469-6288-4113-b25d-f7a052b1f018/publications.html?filter=research>
- Interacting Minds Centre, Aarhus University.
(2019, 11. oktober).
NLP workshop @IMC, Fall 2019.
Hentet 10.03.2020, fra <https://interactingminds.au.dk/events/single-events/artikel/nlp-workshop-imc-fall-2019/>
- Juul, I.
(2004).
Erhvervsuddannelserne — et Forsømt Forskningsområde.
Tidsskrift for Arbejdsliv, 6(4),
11.
<https://doi.org/10/ggzxt>
- Juul, I.
(2013).
Diskurser om ungdom og arbejdsliv — fra velfærdsstat til konkurrencestat.
Tidsskrift for Arbejdsliv, 15(3),
9–23.
<https://doi.org/10/ggzxzs>
- Jørgensen, C. H.
(2016).
Reformen af erhvervsuddannelserne - En gang til, forfra og om igen (T. Canger, P. Hjort-Madsen, S. Kondrup & A. Neubert, Red.).
Dansk Pædagogisk Tidsskrift, 2016(4),

- 10–18.
https://dpt.dk/wp-content/uploads/2020/01/DpT_04_16_ISSUU-2.pdf
- Kiritchenko, S. & Mohammad, S.
(2018).
Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems.
Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics,
43–53.
<https://doi.org/10/ggwczt>
- Kwartler, T.
(2017, 31. juli).
Text Mining in Practice with R.
John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781119282105>
- Lauridsen, G. A., Svendsen, L. K. B. & Dalsgaard, J. A.
(2020, 9. september).
Sentida.
Hentet 30.09.2020, fra <https://github.com/Guscode/Sentida>
- Laver, M., Benoit, K. & Garry, J.
(2003).
Extracting Policy Positions from Political Texts Using Words as Data.
American Political Science Review, 97(02).
<https://doi.org/10/ctpz7>
- Lowe, W.
(2019, 2. maj).
Austin: A Package for Doing Things with Words (Vers. 0.2).
Hentet 27.08.2020, fra <https://rdr.io/rforge/austin/#vignettes>
- Manning, C. D., Raghavan, P. & Schütze, H.
(2008).
Introduction to Information Retrieval.
Cambridge University Press.
<https://doi.org/10.1017/CBO9780511809071>
- Mills, C. W.
(2000).
The Sociological Imagination.
Oxford University Press.
- Molina, M. & Garip, F.
(2019).
Machine Learning for Sociology.
Annual Review of Sociology, 45(1),
27–45.
<https://doi.org/10/gf3vm4>
- Munk, M. D.
(2014).
Social Ulyghed Og Uddannelse.
I H. Dorf & J. Rasmussen (Red.),
Pædagogisk Sociologi.
Hans Reitzels Forlag.
- Nikita, M.
(2016, 24. oktober).
Select number of topics for LDA model.

- Hentet 23.08.2020, fra <https://rpubs.com/siri/ldatuning>
Nord-VET.
(nodate).
The Future of Vocational Education – Learning from the Nordic Countries - Nord-VET.
Hentet 04.05.2020, fra <http://nord-vet.dk/the-future-of-vocational-education-learning-from-the-nordic-countries/>
- Pedersen, M. L.
(2019, 12. november).
Folketinget.
- Proksch, S.-O. & Slapin, J. B.
(2010).
Position Taking in European Parliament Speeches.
British Journal of Political Science, 40(3),
587–611.
<https://doi.org/10/d62v7k>
- Quinn, K. M., Monroe, B. L., Colaresi, M., Crespín, M. H. & Radev, D. R.
(2010).
How to Analyze Political Attention with Minimal Assumptions and Costs.
American Journal of Political Science, 54(1),
209–228.
<https://doi.org/10/cvfs44>
- Radikale Venstre.
(2020, 9. oktober).
Unge Og Uddannelse.
Hentet 09.10.2020, fra <https://valg.radikale.dk/det-mener-vi-om/unge-og-uddannelse/>
- Rasmussen, A. F.
(2005, 24. februar).
Statsminister Anders Fogh Rasmussens redegørelse i Folketinget 24. februar 2005 (Speech).
København.
Hentet 08.02.2020, fra http://www.stm.dk/_p_7489.html
- Redegørelse R3 1993
.
Hentet 09.02.2020, fra <https://www.folketingstidende.dk/samling/19931/redegoerelse/R3/index.htm>
- Savage, M.
(2009).
Contemporary Sociology and the Challenge of Descriptive Assemblage.
European Journal of Social Theory, 12(1),
155–174.
<https://doi.org/10/fcxh24>
- Schnepf, S. V.
(2002).
A Sorting Hat That Fails?: The Transition from Primary to Secondary School in Germany.
Citeseer.
- Sievert, C.
(2020, 10. august).
Cpsievert/LDAvis.
Hentet 23.08.2020, fra <https://github.com/cpsievert/LDAvis>

- Silge, J. & Robinson, D.
(2017).
Text Mining with R: A Tidy Approach.
" O'Reilly Media, Inc."
- Simonsen, K. & Jensen, C.
(2016).
Lad os gøre op med overuddannelsen [newspaper].
Berlingske.dk: Kommentatorer.
Hentet 08.02.2020, fra <https://www.berlingske.dk/content/item/107756>
- Sjoberg, G.
(1951).
A Rationale for Descriptive Sociology.
Social Forces, 29(3),
251–257.
<https://doi.org/10/ffz449>
- Skaarup, P.
(2017, 25. juli).
Politik Er Først Og Fremmest Følelser.
Hentet 04.09.2020, fra <https://danskfolkeparti.dk/politik-foerst-fremmest-foelelser/>
- Slapin, J. B. & Proksch, S.-O.
(2008).
A Scaling Model for Estimating Time-Series Party Positions from Texts.
American Journal of Political Science, 52(3),
705–722.
<https://doi.org/10/brh9q7>
- Socialdemokratiet.
(2020a, 9. oktober).
Erbvervsuddannelser | Socialdemokratiet.
Hentet 09.10.2020, fra <https://www.socialdemokratiet.dk/da/politik/erhvervsuddannelser/>
- Socialdemokratiet.
(2020b, 9. oktober).
Gymnasium | Socialdemokratiet.
Hentet 09.10.2020, fra <https://www.socialdemokratiet.dk/da/politik/gymnasium/>
- Socialdemokratiet.
(2020c, 9. oktober).
Uddannelse | Socialdemokratiets uddannelsespolitik.
Hentet 09.10.2020, fra <https://www.socialdemokratiet.dk/da/politik/uddannelse/>
- Socialdemokratiet.
(2020d, 9. oktober).
Videregående Uddannelse | Socialdemokratiet.
Hentet 09.10.2020, fra <https://www.socialdemokratiet.dk/da/politik/videregaende-uddannelse/>
- Socialistisk Folkeparti.
(2020, 9. oktober).
Skole, Uddannelse Og Forskning - Socialistisk Folkeparti.
Hentet 09.10.2020, fra <https://sf.dk/det-vil-vi/uddannelse-forskning/>
- stopwords-iso.
(2020).
Stopwords ISO.
<https://github.com/stopwords-iso/stopwords-iso>

- Straka, M. & Straková, J.
(2020).
UDPipe | ÚFAL.
Praha.
Hentet 15.06.2020, fra <http://ufal.mff.cuni.cz/udpipe>
- The R Foundation.
(nodate).
R: What Is R?
Hentet 24.08.2020, fra <https://www.r-project.org/about.html>
- The Wikipedia Volunteers.
(2019, 20. august).
EFG.
Wikipedia, den frie encyklopædi.
Hentet 08.02.2020, fra <https://da.wikipedia.org/w/index.php?title=EFG&oldid=10036360>
- Page Version ID: 10036360
- Thomsen, J.-P. & Andrade, S. B.
(2016 december).
Uddannelsesmobilitet i Danmark.
SFI – Det Nationale Forskningscenter for Velfærd.
København.
- Bekendtgørelse Af Lov Om Erhvervsgrunduddannelse m.v.
(2016, 20. juni).
Hentet 12.02.2020, fra <https://www.retsinformation.dk/Forms/R0710.aspx?id=181960>
- Bekendtgørelse Af Lov Om de Gymnasiale Uddannelser
(2019, 28. maj).
Hentet 12.02.2020, fra <https://www.retsinformation.dk/Forms/R0710.aspx?id=209370>
- Bekendtgørelse Af Lov Om Erhvervsuddannelser
(2020, 22. januar).
Hentet 12.02.2020, fra <https://www.retsinformation.dk/Forms/R0710.aspx?id=212462>
- Undervisningsministeriet.
(nodate).
U91: det nye mønster i dansk uddannelses- og forskningspolitik (1. udgave, 2. oplag.).
Undervisningsministeriet.
- Undervisningsministeriet.
(1978).
U 90: samlet uddannelsesplanlægning frem til 90'erne.
Undervisningsministeriet.
- Undervisningsministeriet.
(2017, 19. marts).
9. Og 10. Klasseelevernes Tilmeldinger Til Ung- Domsuddannelserne Og 10. Klasse Mv. 2017.
Undervisningsministeriet, Styrelsen for IT og læring.
København.
- Venstre.
(2016, 18. november).
Principprogram 2016.
Hentet 09.10.2020, fra https://www.venstre.dk/_Resources/Persistent/2/0/a/e/20ae5a4934215ebc17db9d0af1c06ca6e445d1d7/Principprogram%202016.pdf

- Ward, J. H.
(1963).
Hierarchical Grouping to Optimize an Objective Function.
Journal of the American Statistical Association, 58(301),
236–244.
<https://doi.org/10/fz95kg>
- Wickham, H.
(2014).
Tidy Data.
The American Statistician, 14.
<https://doi.org/10/gdm3p7>
- Aarkrog, V.
(2003).
Rummelighed Og Sammenhæng - Erhvervsuddannelsesreform 2000: Målsætninger, Til-
tag Og Konsekvenser.
I I. M. Bryderup (Red.),
Pædagogisk Sociologi (1. udg., s. 361–374).
Aarhus Universitetsforlag.