

# Analyse

2. juli 2018

## Tonen i debatten

*Af Regitze Wandsøe-Isaksen, Kristine Vasiljeva og Frederik Degn Pedersen*

**”Tonen” i debatten kan afholde nogle borgere fra at deltage i debatten. Den hårde tone i debatten mellem forskellige befolkningsgrupper kan også indikere, at sammenhængskraften er udfordret. I det følgende undersøges vha. af data for diskussioner på Facebook på 200.000 offentligt tilgængelige sider, hvordan kommentartonen er på tværs af kulturel baggrund, køn og politisk tilhørsforhold.**

- Tonen i debatten kan være en indikation af, hvordan sammenhængskraften er. En hård tone indikerer således lav sammenhængskraft. En hård tone kan også føre til, at nogle befolkningsgrupper trækker sig fra den offentlige debat, hvilket kan reducere sammenhængskraften i Danmark fremadrettet.
- Vi måler debattonen ud fra diskussioner på Facebook på 200.000 offentlige sider mellem fremmede diskussionspartnere. Tonen måles vha. en manuel vurdering af knap 5.000 kommentarer. Dette udbredes til hele datasættet bestående af små 10 mio. kommentarer vha. machine learning.
- For etniske danskere (herefter: danskere) er andelen af negative kommentarer rettet mod deres diskussionspartner større når diskussionspartneren er mellemøstlig fremfor når personen er dansker. Tilsvarende ses for personer med mellemøstlig oprindelse, at andelen af negative kommentarer rettet mod danskere er højere end andelen af negative kommentarer rettet mod personer med samme baggrund som dem selv.
- For kvinder er andelen af negative kommentarer rettet mod mænd større end andelen af negative kommentarer rettet mod kvinder. For mænd ligger andelen af negative kommentarer rettet mod kvinder en smule lavere end andelen af negative kommentarer rettet mod mænd.
- På tværs af højre-venstreskalaen ses, at personer, der har tilhørsforhold til et parti på venstrefløjen, er andelen af negative kommentarer rettet mod personer fra højrefløjen – ikke overraskende – større end andelen af negative kommentarer rettet mod personer fra deres egen fløj. Tilsvarende billede ses for højreorienterede.
- Løsningsforslag på den hårde tone i debatten har været diskuteret i medierne. Bl.a. har Institut for Menneskerettigheder foreslået, at medierne skal moderere deres online debatsider mere aktivt. Omvendt har andre fremført argumenter for, at moderering kan føre til polarisering og ekkokammereffekter.

### Kontakt

Senioranalytiker  
Regitze Wandsøe-Isaksen  
Tlf. 31404252  
E-mail [reg@kraka.org](mailto:reg@kraka.org)

Ledende økonom, PhD  
Kristine Vasiljeva  
Tlf. 23939327  
E-mail [kri@kraka.org](mailto:kri@kraka.org)

Indikation på sammenhængskraft	Tonen i debatten kan forventes at være vigtig for sammenhængskraften i et samfund – fx fordi en negativ tone kan få nogle mennesker til ikke at ville deltage i debatten. <sup>1</sup> Hvis nogle herudover udelukkes fra debatten, fordi de benytter en for hård tone, kan det omvendt føre til, at de kun ytrer sig i lukkede fora. Det kan så føre til en højere polarisering og lavere sammenhængskraft <sup>2</sup> – fx fordi man så i ekkokamre etablerer forskellige opfattelser af, hvad der definerer et succesfuldt samfund.
Den anvendte tilgang benyttes i stigende grad	Der bliver i stigende grad foretaget undersøgelser af tonen i kommentarer eller ytringer i samfundsvidenskaben. Undersøgelsen af tonen i ytringer kan bl.a. benyttes til at studere alt fra politisk påvirkning af borgernes følelser (Bail m.fl., 2017) og hadefuld tale ("hate speech") (Burnap og Williams, 2015) til, hvordan subjektive følelser på samfundsplan varierer (Golder og Macy, 2011).
Vores analyse: Omfattende data fra Facebook	I det følgende har vi ud fra en forholdsvis ny metode undersøgt tonen i debatten på Facebook. Denne analyse adskiller sig fra tidligere undersøgelser på danske data, da den omfatter al aktivitet på mere end 200.000 offentligt tilgængelige, danske Facebooksider tilhørende politikere, kendte, nyhedsmedier o.l. i perioden 1. oktober 2015 til 1. oktober 2016. <sup>3</sup> Vi observerer tonen i debatten for alle personer, der har deltaget i diskussioner på offentlige sider i den pågældende periode, hvilket tillader os at undersøge debatten på tværs af grupper med forskellig kulturel baggrund, mellem mænd og kvinder samt mellem højre- og venstreorienterede. <sup>4</sup>
Metoden er baseret på machine learning	Tonen måles som andelen af kommentarer, der er såkaldt "relationelt negative". Relationelt negative kommentarer betyder, at indholdet i kommentaren er en negativ ytring rettet mod en diskussionspartner. Konkret laves først et træningsdatasæt, som består af en manuel vurdering af ca. 5.000 kommentarer – foretaget af syv personer. Ved hjælp af dette træningsdatasæt oplæres en algoritme til at kunne kende forskel på negative kommentarer rettet mod andre og andre slags kommentarer. Algoritmen anvendes så på omkring 10 mio. kommentarer. <sup>5</sup> Dette gør nærværende analyse unik ift. tidligere analyser af tonen i debatten i Danmark.
Nyt forskningsfelt giver øget grad af usikkerhed	Selvom undersøgelsen er baseret på et forholdsvis stort træningsdatasæt og foretaget i samarbejde med forskere fra Københavns Universitet, skal resultaterne imidlertid også læses med forbehold for, at arbejdet er foretaget inden for et nyt forskningsfelt – og at det er kompliceret at oplære en algoritme til fuldstændigt at forstå fx sarkasme.
Tonen mellem kulturelle grupper	Analysen viser, at debatten imellem grupper af personer afhænger af, om de kommunikerer med personer med samme baggrund som dem selv, jf. Figur 1. Konkret gælder det, at personer med dansk baggrund retter en relativt større andel negative kommentarer mod personer med mellemøstlig baggrund, end de gør mod andre personer med dansk baggrund, jf. Figur 1 a). Tilsvarende for personer med mellemøstlig baggrund er andelen af negative kommentarer rettet mod personer med dansk baggrund større end andelen af negative kommentarer rettet mod andre med mellemøstlig baggrund, jf. Figur 1 b). <sup>6</sup>

---

<sup>1</sup> Se Zuleta m.fl. (2017).

<sup>2</sup> Se Mchangama (2017).

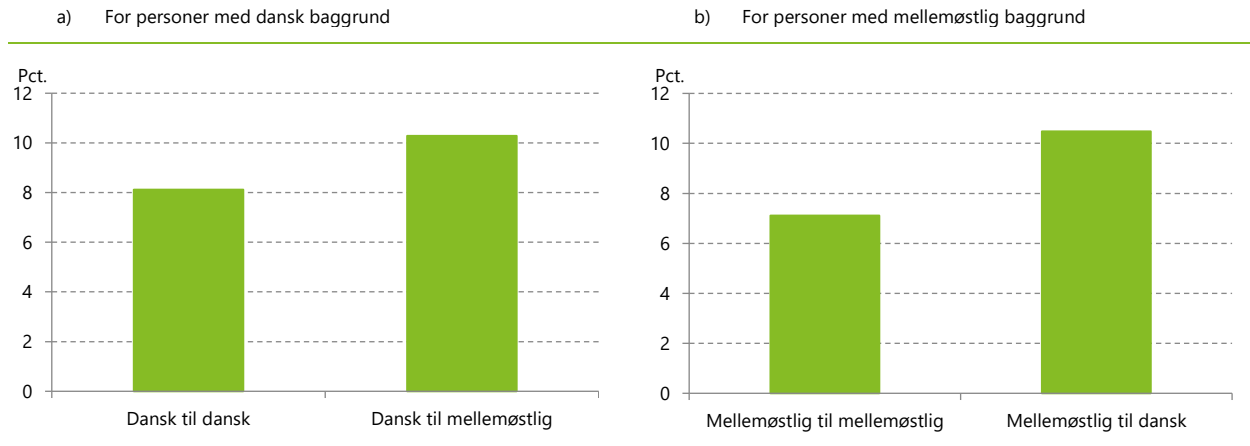
<sup>3</sup> Institut for Menneskerettigheder har tidligere undersøgt tonen i den danske debat på Facebook. Undersøgelsen var imidlertid begrænset til ca. 3.000 kommentarer på DR og TV 2's Facebooksider – og ikke tonen i debatten på Facebook generelt. Se yderligere beskrivelse hos Zuleta m.fl. (2017).

<sup>4</sup> Kulturel baggrund og køn identificeres ud fra den enkelte persons Facebooknavn og politisk tilhørsforhold ud fra, hvad man har liket på Facebook. Se yderligere beskrivelse i Boks 1.

<sup>5</sup> Det skal bemærkes, at tonen kun er målt for diskussioner, der foregår på dansk.

<sup>6</sup> Vores resultat stemmer i høj grad overens med konklusionerne fra Institut for Menneskerettigheders undersøgelse, hvor de finder, at hadefulde ytringer i høj grad er rettet mod personer med mellemøstlig oprindelse.

**Figur 1 Andel negative kommentarer på tværs af kulturel baggrund**



Anm.: Figuren angiver andelen af negative kommentarer rettet mod en diskussionspartner ud af alle kommentarer. Fx viser "Dansker til mellemøstlig" andelen af negative kommentarer fra danskere rettet mod personer med mellemøstlig baggrund ud af alle kommentarer fra danskere til personer med mellemøstlig baggrund.

Kilde: Ralund (2018).

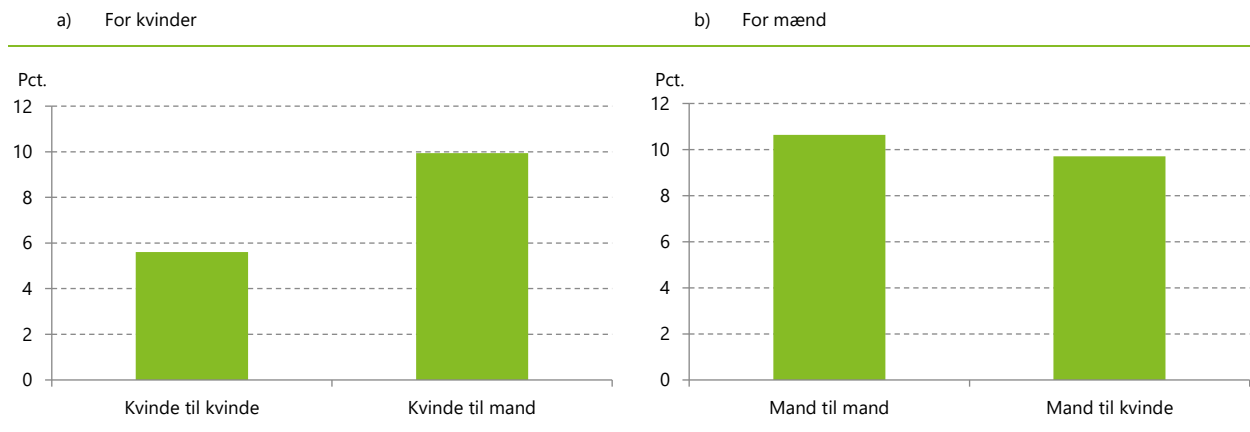
**Sammenligning af resultaterne mellem grupperne**

Man skal være varsom med at sammenligne resultaterne fra Figur 1 a) med resultaterne fra Figur 1 b), da algoritmen bliver oplært til at forstå, hvad negative kommentarer er på et væsentligt mindre antal observationer for minoritetsgrupperne end for majoritetsgrupperne.<sup>7</sup> På samme måde er der væsentligt mindre kommunikation på tværs af de to grupper med forskellig kulturel baggrund, end der er kommunikation inden for grupperne. Derfor kan man forvente, at vi finder et konservativt estimat for den hårdere tone på tværs af grupper. Det betyder imidlertid blot, at personer med dansk baggrund forventes at tale endnu hårdere til personer med mellemøstlig baggrund end til andre personer med dansk baggrund og at personer med mellemøstlig baggrund taler endnu hårdere til personer med dansk baggrund end til personer fra deres egen gruppe, end vi har angivet.

**Tonen mellem kønnene**

Debattonen mellem kønnene afhænger på samme måde af, hvilke grupper der debatterer, jf. Figur 2. Mens kvinder taler væsentligt mere hårdt til mænd, end de gør til andre kvinder, taler mænd nogenlunde lige så hårdt til både kvinder og mænd, som kvinder taler til mænd.

**Figur 2 Andel negative kommentarer på tværs af køn**



Anm.: Se Figur 1.

Kilde: Ralund (2018).

<sup>7</sup> Konkret får vi et konservativt estimat for andelen af negative kommentarer for medlemmer af minoritetsgrupperne, da træningsdatasættet for denne gruppe er relativt småt, jf. Ralund (2018). Forklaringen kan fx være, at algoritmen kan have svært ved at genkende udtryk, som den ikke har set før, og derfor klassificerer algoritmen dem ikke som relationelt negative.

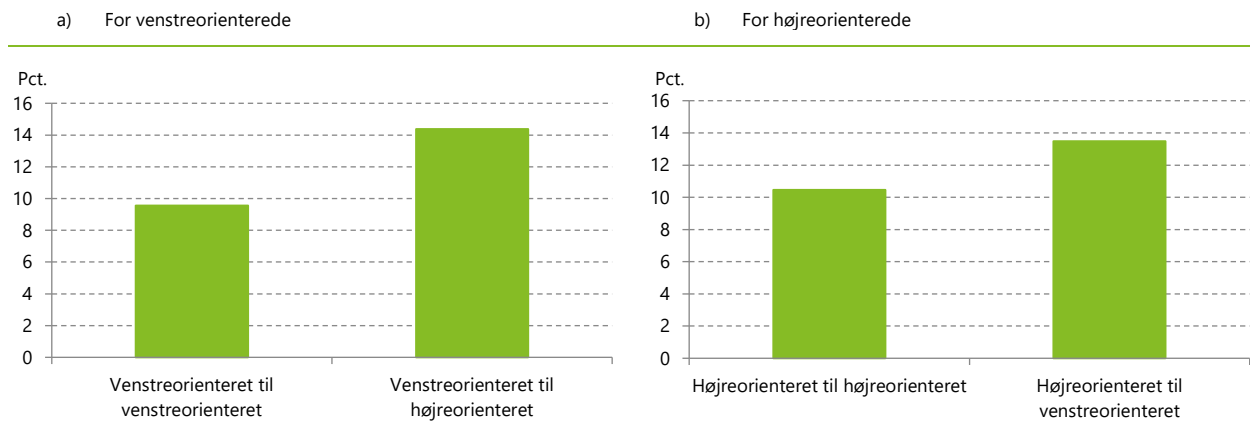
Flere kvinder undgår at deltage i den offentlige debat

På trods af, at tonen i kommentarer rettet mod kvinder ikke er mere negativ end tonen rettet mod mænd, er der flere kvinder, der trækker sig eller undgår at deltage i debatten på sociale medier pga. tonen i debatten. Konkret finder Danmarks Statistik (2017), at 58 pct. af kvinderne i Danmark mener, at tonen i debatten på sociale medier afholder dem fra at udtrykke deres mening og deltage i debatten – mod kun 43 pct. af mændene.<sup>8</sup>

Tonen på tværs af politiske fløje

På tværs af højre-venstreskalaen gælder det, at personer med tilhørsforhold til et parti på én side af skalaen – ikke overraskende – har en større andel negative kommentarer rettet mod personer fra den anden side af skalaen end deres egen, jf. Figur 3.

**Figur 3 Andel negative kommentarer på tværs af politisk tilhørsforhold**



Anm.: Se Figur 1.

Kilde: Ralund (2018).

Diskussion om løsninger

Løsningsforslag på den hårde tone i debatten har været diskuteret i medierne. Bl.a. har Institut for Menneskerettigheder foreslået, at medierne skal moderere deres online debatsider mere aktivt, jf. Zuleta m.fl. (2017). Omvendt har Mchangama (2017) fremført argumenter for, at moderering kan føre til polarisering og ekkokammereffekter.

Centralt med spilleregler for online færden

Fra barnsben oplæres danskere i regler og normer ift. hvordan man kommunikerer ansigt til ansigt. Børn lærer fx, at man ikke må slå andre børn, og at man skal tale ordentligt til hinanden. Det giver et sæt fælles spilleregler, som gør det nemmere at leve sammen. En udfordring på de sociale medier er, at folk ikke nødvendigvis følger de eksisterende regler for god opførsel, fx fordi man ikke står ansigt til ansigt med sin diskussionspartner. Det kommer muligvis med tiden. Indtil da kan man definere grundlæggende spilleregler og undervise børn og unge – såvel som den ældre generation – i disse. Ligeledes kan man oplyse om konsekvenserne af en hård tone imod diskussionspartnere.

<sup>8</sup> 58 pct. og 43 pct. dækker over alle personer med det pågældende køn, der har svaret "Meget enig" eller "Enig" til udsagnet "Tonen i debatter på sociale medier afholder mig fra at udtrykke min mening og deltage i debatten".

**Boks 1 Beskrivelse af datagrundlag**

Krakas beregninger er foretaget på anonymiserede og aggregerede data – og kan således ikke tilbageføres til enkeltpersoner.

*Population:* Undersøgelsen er baseret på offentlig aktivitet på mere end 200.000 offentligt tilgængelige danske Facebooksider tilhørende politikere, kendte, nyhedsmedier osv. i perioden 1. oktober 2015 til 1. oktober 2016. Det betyder, at undersøgelsen er lavet med udgangspunkt i omkring 10 mio. kommentarer.

*Inddeling i kategorier:* Opdeling på køn foretages ved en sammenligning af navne med en navneliste godkendt af regeringen. Opdeling på partitilhørsforhold er baseret på personers likes på Facebook af 1.500 politikere forbundet til et parti. Flere end 1,9 mio. danskere har offentligt tilkendegivet deres opbakning til en politiker. Opdeling på kulturel baggrund er foretaget vha. en algoritme, som på baggrund af 57 mio. kontaktlister kan klassificere personens baggrund ud fra deres navn.

*Mål for diskussionspartner:* Det vigtige er at være sikker på, hvem der svarer hvem, og om de er venner og kender hinanden, eller om de diskuterer. En diskussionspartner defineres på baggrund af to forhold. For det første skal der udtrykkes manglende venskabelige tegn – dvs. man deltager i en diskussion, som man ikke er tagget i. For det andet skal man have trykket "svar" til en kommentar, således at der opstår et subkommentarspor.

*Mål for om man har mellemøstlig baggrund eller ej:* Her benyttes samme tilgang som i Ambekar m.fl. (2009) til at opdele personer efter mellemøstlig baggrund og dansk baggrund. Mellemøstlig og dansk baggrund underestimeres i forhold til andre kategorier. Da der ses på relationer imellem grupper, er der imidlertid ingen grund til at tro, at dette påvirker konklusionerne i vores analyse.

*Fordele og ulemper ved en machine learning-tilgang til at bestemme tonen i debatten:* Fordelen ved at benytte en machine learning-tilgang til bestemmelse af tonen i debatten er, at det muliggør, at man kan se på tonen i et væsentligt større antal kommentarer og debatter, end hvis man skal benytte manuel vurdering af hver kommentar. Omvendt kræver tilgangen, at man har et omfangsrigt træningsdatasæt således, at træningssættet repræsenterer så meget information som muligt (Hutto og Gilbert, 2014). Derfor er der til analysen oprettet datasæt igennem 7 personers manuelle gennemgang af knap 5.000 tilfældigt udtrukne kommentarer. Der kan stadigvæk opstå subjektiv fejlagtig opfattelse af tonen i de enkelte kommentarer, derfor blev der udført krydsvalidering på 10 pct. af træningsdatasættet, hvor mere end en person har evalueret tonen i kommentaren. Her var der i knap 90 pct. af tilfældene enighed om tonen af kommentarerne, som er et højt tal i denne type undersøgelse. Når resultaterne blev krydsvalideret ift. modellen, så var der en enighed på knap 80 pct. med computerens vurdering.

## 1. Litteratur

Ambekar, A., Ward, C., Mohammed, S. og Skiena, S. (2009). "Name-ethnicity classification from open sources." *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. s. 49-58.

Bail, C., Brown, T. og Mann, M. (2017). "Channeling Hearts and Minds: Advocacy Organizations, Cognitive-Emotional Currents, and Public Conversation." *American Sociological Review*, 82(6), s. 1188-1213.

Burnap, P. og Williams, M. (2015). "Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making." *Policy and Internet*. Vol. 7(2), s. 223-242.

Danmarks Statistik (2017). *Resultater fra undersøgelsen onsdag den 14. juni 2017*. [https://www.dst.dk/ext/8808853785/0/formid/Svar-fra-befolkningen-\(uddybende\)--pdf](https://www.dst.dk/ext/8808853785/0/formid/Svar-fra-befolkningen-(uddybende)--pdf).

Golder, S. og Macy, M. (2011). "Diurnal and seasonal mod vary with work, sleep, and daylength across divers cultures." *Science*, 33(6051), s. 1878-1881.

Hutto, C. og Gilbert, E. (2014). "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text." *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*.

Mchangama, J. (2017). *Nej, SÅDAN skal vi bekæmpe den hadefulde debat på nettet: Sæt den fri*. <https://www.zetland.dk/historie/seEIEPxr-aevmJ36a-18bd6>.

Ralund, S. (2018). *Dokumentation af databehandling – fremsendt ved datalevering*.

Zuleta, L., Burkal, R. og Thomsen, T. (2017). *Hadefulde ytringer i den offentlige online debat*. [https://menneskeret.dk/sites/menneskeret.dk/files/media/dokumenter/udgivelser/ligebehandling\\_2017/old\\_rapport\\_om\\_hadefulde\\_ytringer\\_2017.pdf](https://menneskeret.dk/sites/menneskeret.dk/files/media/dokumenter/udgivelser/ligebehandling_2017/old_rapport_om_hadefulde_ytringer_2017.pdf).