

NYT FRA SPROGNÆVNET

OKTOBER 2026

Dansk
Sprognævn

INDHOLD

Jagten på den gode og sikre chatbot

Kan man stole på en chatbot hvis man ønsker rådgivning om sproglig korrekthed?

Ord vi hader, og ord vi elsker (rapport fra en sprogfestival)

Farlige ord

Farlige ord er temaet for en række artikler som bringes i *Nyt fra Sprognævnet* i 2026. Artiklen i dette nummer handler om ordet *eskimo*.

Spørgsmål & svar

■ *Perkle rundt*



Jagten på den gode og sikre chatbot¹

Kan man spørge en chatbot om hvor kommaet skal stå i en sætning? Chatbotters evner når det gælder sproglig korrekthed som fx korrekt kommatering, er der tilsyneladende en udbredt tiltro til.



OLIVER RIX JOHANSEN
(f. 2000) er ph.d.-studerende i Dansk Sprognævn.



MARIANNE RATHJE
(f. 1973) er seniorforsker i Dansk Sprognævn.



JØRGEN SCHACK
(f. 1961) er seniorforsker i Dansk Sprognævn.



VOR GODE ER CHATBOTTER TIL SPROGLIG KORREKTHED?

Journalist Søren K. Villemoes skriver det følgende i Weekendavisen om chatbotternes kometkompetencer: ”korrekt tegnsætning er et af chatbotternes mange kendetegn” (Villemoes 2026). Også mediet Copenhagen Review of Communication nævner at et sprogligt træk der afslører at en tekst er skrevet af en chatbot i modsætning til et menneske, er at teksten er ”Helt fri for stavfejl” (Copenhagen Review of Communication 2025). En tekstforfatter udtaler ligeledes i et interview at hun bruger chatbotter dagligt i sit arbejde: ”allermest til rugbrødsarbejdet. For eksempel til første udkast af en oversættelse eller korrekturlæsning” (Obitsø 2026). Endelig kan man på hjemmesider hvor der udbydes AI-hjælp til korrektur, finde det synspunkt at chatbotter gør menneskelig korrekturlæsning overflødig, fordi chatbotterne kan løse opgaven for os: Her markedsføres chatbotter som leveringsdygtige i ”professionel dansk korrektur på sekunder” (<https://retmintekst.dk/>), og som værktøjer der ”fanger stavfejl, kommafejl og grammatik” (<https://skrivsikkert.dk/ai-korrekturlaesning-guide-til-fejlfri-dansk-tekst/>). Men når det kommer til stykket, hvor gode er chatbotterne så egentlig til sproglig korrekthed?

Chatbotter er på kort tid blevet en del af mange danskeres hverdag. I 2025 havde 48

1 Titlen spiller på Philip Diderichsen og Jørgen Schacks artikel *Jagten på den gode og sikre sprogbruger*, Nyt fra Sprognævnet 2015/3.



% af befolkningen mellem 16 og 74 år brugt AI-værktøjer indenfor de foregående tre måneder (Danmarks Statistik 2025). Blandt de yngste, dvs. de 16-24-årige, var tallet hele 77 %. Danmarks Statistik oplyser også at OpenAI's chatbot, ChatGPT, var den mest anvendte chatbot: 9 ud af 10 brugere af AI anvendte ChatGPT.

Det er ikke mindst som skrivehjælp at chatbotterne bruges. En amerikansk undersøgelse af brugerprompter til ChatGPT (Chatterji et al. 2025) viser at "skrivning" ("writing") er et af de hyppigst nævnte formål med at bruge AI, dvs. til bl.a. produktion og retning af tekst, og skrivning er den mest almindelige arbejdsrelaterede brug der rapporteres. Undersøgelsen viste desuden at omkring to tredjedele af alle de skriveopgaver som chatbotten blev brugt til, var at ændre brugerens egen tekst snarere end at producere ny tekst.

Til trods for at chatbotterne i høj grad benyttes som tekstredigeringsværktøjer, er det sparsomt med undersøgelser der har testet kvaliteten af chatbotternes arbejde med sproglig korrekthed. En undtagelse er forskning foretaget af det norske sprognævn, som har testet hvor godt chatbotter skriver norsk (Språkrådet 2025a, 2025b). I den første af to undersøgelser lod Språkrådet ChatGPT-4o skrive tekster på bokmål og nynorsk. Chatbotten begik i gennemsnit 2,6 fejl pr. side på bokmål og 8 fejl pr. side på nynorsk. I bokmålsteksterne var fejltypene altovervejende tegnsætningsfejl og fejl i store og små bogstaver, som tilsammen udgjorde 70 % af alle fejl. I de nynorske tekster var den hyppigste fejltipe forkerte bøjningsformer (21 %). I den anden og større undersøgelse foretaget af

Språkrådet testede man fire chatbotter: ChatGPT, Microsoft Copilot, Le Chat, der nu kaldes Vibe, og den norskudviklede NorMistral. Også her fandt Språkrådet et højt antal fejl i de AI-genererede tekster, især på nynorsk. Således var der mellem 13 (Copilot og NorMistral, begge på bokmål) og 33 fejl (Le Chat (nu Vibe), nynorsk) pr. 1000 ord. Fejlene spændte over fx bøjningsfejl, syntaktiske fejl og kommafejl til uidiomatiske formuleringer. Språkrådet konkluderede på baggrund af undersøgelserne at hvis man skal bruge chatbotter til at hjælpe med sproglig korrekthed, kræver det menneskelig kontrol af chatbottens output.

Den eneste danske undersøgelse af AI og sproglig korrekthed vi kender til, er Blom et al. (2024), der har testet om det er muligt at skelne sprogligt mellem tekster skrevet af danskstuderende og af ChatGPT² (Blom et al. 2024: 80-81). Ud fra løsninger af den samme stillede opgave testede Blom et al. bl.a. afvigelser fra retskrivningsnormen, og her klarede ChatGPT sig væsentligt bedre end de danskstuderende, dvs. de menneskeskabte tekster: De studerende begik 64 sprogfejl pr. 1000 ord, mens chatbotten kun lavede 3 sprogfejl pr. 1000 ord. Forfatterne konkluderer herudfra at "en høj frekvens af sprogfejl afslører det menneskelige fingeraftryk i teksten" (Blom et al. 2024: 94). De forholdsmæssigt få fejl chatbotten begik, var kommafejl, manglende ord, syntaksfejl og idiomatiske fejl.

2 Der er tale om en sammenlignende analyse af to tekst-korpusser; det ene korpus består af opgaveløsninger skrevet af nystartede danskstuderende fra Syddansk Universitet, og det andet korpus er opgaveløsninger (ud fra samme opgave) genereret af ChatGPT-4o.



De undersøgelser af AI og sproglig korrekthed som er nævnt ovenfor, peger altså i hver sin retning: Den danske forskningsgruppe viser at chatbotten er bedre til sproglig korrekthed end mennesket, mens det norske sprognævn konstaterer at chatbotten laver så mange fejl at der må et menneske til at tjekke den.

Denne artikel er vores bidrag til en afklaring af spørgsmålet: Har chatbotterne den sproglige kompetence som kræves af en dygtig korrekturlæser, eller er det for risikabelt at stole på deres sproglige evner? Spørgsmålet besvarer vi ved at teste tre forskellige chatbotter i forhold til et stort antal sproglige spørgsmål af den type som vi i Sprognævnet møder i vores spørgetelefon. I fremtidige undersøgelser vil vi teste chatbotternes evner som korrekturlæsere og som tekstforfattere.

DE STILLEDE SPØRGSMÅL

Til brug for undersøgelsen har vi udarbejdet 1.500 spørgsmål med 2-4 svarmuligheder, som er fordelt på de nedenstående 15 kategorier, de fleste med tilhørende underkategorier, som af pladshensyn er udeladt her.

1. Apostrof
2. Punktum
3. Komma (med startkomma)
4. Komma (uden startkomma)
5. Bindestreg
6. Store og små bogstaver
7. Ét eller flere ord?
8. Stavefejl (fejl i ordets stamme)
9. Morfologi
10. Syntaks
11. Adverbielt -t
12. *Hans/hendes vs. sin/sit/sine, deres vs. sin/sit/sine*
13. Pronominalkasus

14. R-problemer

15. Blandet morfosyntaks og ledstilling

De 15 ovenstående kategorier kan bundtes i 5 overkategorier, nemlig 1) Tegnsætning (1-5), 2) Ortografi (6-8), 3) Morfologi (9), 4) Syntaks (10) og 5) Morfosyntaks (11-15) (jf. afsnittet "Overordnede resultater" nedenfor).

Artiklens tre forfattere har hver især udarbejdet 500 spørgsmål og har efterfølgende læst korrektur på hinandens spørgsmål.

Hver kategori omfatter 100 spørgsmål. Nedenfor vises et par eksempler på de konkrete spørgsmål som vi har stillet til chatbotterne:

Hvilke af følgende muligheder er korrekte?

Svarmulighed a: mur og nagelfast.

Svarmulighed b: mur- og nagelfast.

(B er det korrekte svar).

Hvilke af følgende muligheder er korrekte?

Vær opmærksom på at der sættes startkomma.

Svarmulighed a: Det hold, jeg plejer at komme på, starter kl. 19.

Svarmulighed b: Det hold, jeg plejer at komme på starter kl. 19.

Svarmulighed c: Det hold jeg plejer at komme på, starter kl. 19.

(A er det korrekte svar).

Kategorierne og de tilhørende spørgsmål er primært baseret på en afvigelsestypologi som er udarbejdet af Jørgen Schack i samarbejde med Line Burholt Kristensen og Katrine Falcon Søby (Københavns Universitet) i forbindelse med forskningsprojektet "Broken Grammar & Beyond", som undersøgte hvilke grammatikafvigelser der produceres i tekster, og hvordan disse tekster læses og forstås af andre (Søby & Kristensen 2019).



Afvigelsestypologien bygger hovedsagelig på *Retskrivningsordbogens* regelsæt, men er suppleret med hyppigt forekommende afvigelser som falder udenfor ordbogens regler, først og fremmest morfosyntaktiske afvigelser, dvs. afvigelser som vedrører valget af ordformer i konkrete grammatiske kontekster, fx valget mellem *deres* og *sine* i *De havde ikke hentet deres/sine cykler*, og valget mellem en form uden eller med *-r* i fx *DSB elektrificere/elektrificerer flere banestrækninger*.

Ved udfyldningen af de ovennævnte kategorier, med tilhørende underkategorier, har vi støttet os til en række undersøgelser som beskriver hyppige afvigelser i forskellige tekstarter: fejltyper i diktater ved folkeskolens afsluttende prøver (Jervelund & Schack 2016), fejltyper i avistekster (Rathje et al. 2023), fejltyper i universitetsopgaver (Rathje 2019), og endelig har vi inddraget undersøgelser af hvilke spørgsmål der hyppigst stilles til Sprognævnet (Jarvad 2014). De ovennævnte undersøgelser har bevirket at visse meget hyppigt forekommende morfosyntaktiske spørgsmål, fx adverbialt *-t* og *r*-problemer, har fået selvstændige kategorier og dermed optræder med flere spørgsmål end visse andre kategorier i undersøgelsen.

Til hvert spørgsmål er der som nævnt anført mellem to og fire svarmuligheder som de testede chatbotter skulle vælge imellem. I de tilfælde hvor der er mere end ét rigtigt svar, har vi anført den eller de svarmuligheder som bedst afspejler Sprognævnets svarpraksis, som den eller

de rigtige svarmuligheder. Det medfører at de svarmuligheder der er anført som rigtige, i en del tilfælde afspejler en forholdsvis konservativ norm. Det gælder først og fremmest spørgsmål som vedrører morfosyntaks, fx *De spillede ikke deres (ikke sin) bedste kamp i dag*, *De forstår slet ikke min kone og mig (ikke jeg)*, *Tulipanerne er (ikke har) afblomstret*.

METODE

Vi har testet 6 forskellige sprogmodeller fra 3 forskellige virksomheder – en større og en mindre model fra hhv. amerikanske OpenAI og Anthropic samt franske Mistral, der hhv. udbyder chatbotterne ChatGPT, Claude og Vibe (tidligere kaldet Le Chat). Vi har valgt OpenAI's ChatGPT fordi den er kendt af de fleste og er den mest brugte model, og Anthropics Claude fordi den har ry for at være særligt god til at skrive tekst. Franske Mistrals Vibe har vi valgt fordi den er den vigtigste europæiske konkurrent til de amerikanske chatbotter. Mistrals modeller er desuden open source og indgår i det danske forskningsprojekt "Danish Foundation Models" (Larsen & Blach 2026). Til grund for alle chatbotterne ligger såkaldte sprogmodeller, og det er dem vi har testet i denne undersøgelse. Vi testede modellerne i perioden 4.-11. juni 2026. Bemærk at der siden er kommet nyere versioner af både OpenAI's og Anthropics modeller.

For både OpenAI's ChatGPT og Anthropics

Virksomhed	Chatbot	Modelnavn	Model-ID
OpenAI	ChatGPT	GPT 5.5	gpt-5_5-2026-04-23
OpenAI	ChatGPT	GPT 5.4 mini	gpt-5_4-mini-2026-03-17
Anthropic	Claude	Opus 4.8	claude-opus-4-8
Anthropic	Claude	Sonnet 4.6	claude-sonnet-4-6
Mistral	Vibe	Large	mistral-large-2512
Mistral	Vibe	Medium	mistral-medium-3-5

Tabel 1: Oversigt over de testede modeller



»

Claude gælder det at man som bruger uden betalt abonnement kun har stærkt begrænset adgang til de større og mere avancerede modeller, mens man med et betalt abonnement har udvidet adgang til disse modeller. Derfor har vi testet både den (på testtidspunktet) største model og den nyeste minimodel, omtrent svarende til hvad man har adgang til som hhv. betalende og ikke-betalende bruger.

Når man bruger Mistral's chatbot, Vibe, kan man ikke selv vælge hvilken model der anvendes – det vælges for en, uanset om man er betalende bruger eller ej, men for sammenlignelighedens skyld har vi også her testet både en større og en mindre model.

Hver model blev stillet de samme 1500 spørgsmål 3 gange; vi gentog eksperimentet 3 gange for at se hvor konsistente modellerne er i deres svar. Før selve spørgsmålene blev stillet, blev modellen givet en såkaldt systemprompt, dvs. en generel instruktion om hvilket format modellen skal svare med:

"Returner præcis følgende format:
[SVAR]<bogstaver adskilt af komma>
[/SVAR]
[FORKLARING]<kort begrundelse>
[/FORKLARING]
Ingen anden tekst må forekomme."

Derefter følger en mere specifik promptinstruktion om hvordan modellen skal svare:

"Vælg de korrekte svarmuligheder.
Angiv kun bogstaverne for de korrekte svar i [SVAR]-tagget.
Giv en kort faglig begrundelse i [FORKLARING]-tagget.
<herefter følger spørgsmål og svarmuligheder>."

Vi fortæller altså modellen præcis hvad den skal gøre, og hvordan den skal strukturere sit svar. Det gør det lettere efterfølgende at kunne sammenligne modellens svar med de korrekte svar uden for megen efterbehandling. Modellen skal m.a.o. følge en skabelon hvor hhv. svar og forklaring er klart afgrænsede. I denne undersøgelse beder vi desuden modellerne om både at vælge de rigtige svar og begrunde svarene. Af hensyn til omfanget behandler vi i denne artikel dog kun modellernes svar og altså ikke begrundelserne for svarene.

Vi anvender her såkaldt *zero shot prompting* – en teknik hvor modellen ikke får eksempler på opgaven, men alene en instruktion (Nygaard 2024: 67). Den skal derfor løse opgaven ud fra sin generelle træning og den konkrete formulering af spørgsmålet. Det er til forskel fra fx *one shot* eller *few shot prompting*, hvor modellen først præsenteres for hhv. ét eller få eksempler på spørgsmål og tilhørende svar (Nygaard 2024: 77; Nygaard 2024: 85). Vi har valgt strategien *zero shot prompting* fordi vi antager at det afspejler den måde de fleste brugere anvender en chatbot på, dvs. at stille et spørgsmål uden først at give modellen eksempler på hvordan det skal bevares.

Modellerne blev alle tilgået via API'er, der gør det muligt automatisk at sende spørgsmål til og modtage svar fra modellerne vha. programmer vi har udviklet til formålet. Vi valgte at anvende API'er af flere grunde. Først og fremmest fordi det er meget lettere at sende og modtage 1500 spørgsmål og svar automatisk end manuelt at kopiere og indsætte dem i den webgrænseflade hvor man normalt interagerer med chatbotterne, fx på www.chatgpt.com. For det andet kræver webgrænsefladerne typisk en brugerkonto, og de bevarer som udgangspunkt samtalehistorikken, så tidligere spørgsmål kan påvirke efterfølgende svar. I vores API-baserede opsætning blev spørgsmålene sendt uden tidligere samtalehisto-



rik, så alle spørgsmål blev behandlet uafhængigt af hinanden. Det giver et mere kontrolleret og reproducerbart datagrundlag.

Når man tilgår en sprogmodel via et API, er det muligt at definere en række parametre som påvirker modellens adfærd. Det er dog ikke alle modeller der understøtter de samme parametre. GPT 5.4 mini, Sonnet 4.6, Mistral Medium og Mistral Large understøtter fx en såkaldt temperaturindstilling, der regulerer graden af tilfældighed i modellens svar, dvs. hvor meget modellens svar varierer. Denne indstilling blev sat til 0 (den laveste indstilling) for at minimere variationen. GPT 5.5 og Opus 4.8 har ikke en sådan indstilling, og disse modellers svaradfærd er derfor fastlagt internt af modellen og kan ikke justeres via API'et. Bl.a. for at tage højde for dette har vi foretaget eksperimentet tre gange. Modellerne valgte nøjagtig de samme svar i 89-97 % af tilfældene på tværs af de tre kørsler, hvilket indikerer at modellerne generelt var meget stabile i deres svar.

Modellen GPT 5.5 er en såkaldt *reasoning model*. Det betyder at den bruger ressourcer på at 'tænke' inden den svarer, til forskel fra traditionelle sprogmodeller der genererer svar mere direkte ud fra brugerens input ([OpenAI 2026a](#)). For denne model er parameteren *reasoning* derfor relevant; den specificerer hvor grundigt modellen skal 'tænke' inden den svarer. Vi satte værdien til "medium", der er "the recommended balanced starting point for quality, reliability, latency, and cost" ([OpenAI 2026b](#)). En lavere indstilling er mindre omkostningsfuld, men der er større risiko for fejl, mens en højere indstilling kan reducere antallet af fejl på bekostning af længere svartid og højere omkostninger.

RESULTATER

I denne artikel fremlægger vi som nævnt af pladshensyn kun modellernes svar på spørgs-

målene – modellernes forklaringer på svarene udskydes til en senere artikel. Resultaterne for hver af de seks modeller fremgår nedenfor. For hver af de 15 sproglige kategorier, som er beskrevet i afsnittet "De stillede spørgsmål" (fx store og små bogstaver), angiver vi fire mål for modellens præstation.

Ved evaluering af modeller der skal klassificere eller udvælge svar, er det almindeligt at anvende målene **præcision** (*precision*) og **fuldstændighed** (*recall*) samt **F1**, der kombinerer de to. Herudover anvender vi et supplerende mål, **eksakt**, som måler hvor ofte modellen vælger netop de rigtige svar – hverken flere eller færre. De fire mål besvarer følgende spørgsmål:

- **Eksakt:** Hvor ofte vælger modellen alle de korrekte og kun de korrekte svar?
- **Præcision** (*precision*): Hvor stor en andel af de valgte svar er korrekte?
- **Fuldstændighed** (*recall*): Hvor stor en andel af de korrekte svar vælges?
- **F1:** Hvor godt balancerer præcision og fuldstændighed?

Det strengeste mål er det **eksakte**: Hvis de rigtige svar på et spørgsmål fx er {a,b,c}, men modellen kun vælger fx {a,b}, får den scoren 0. Kun hvis modellen vælger det samlede sæt af korrekte svar, {a,b,c}, dvs. kun rigtige og alle rigtige svar, får modellen scoren 1. Det er altså alt eller intet. **Præcision** angiver hvor stor en andel af modellens svar der er korrekte. Hvis de rigtige svar på et spørgsmål fx er {a,b,c}, og modellen svarer {a,b}, er præcisionen 100 % – alle svar er rigtige, selvom modellen overser det rigtige svar {c}. **Fuldstændighed** angiver hvor stor en andel af de korrekte svar der vælges. Hvis de rigtige svar er {a,b,c}, og modellen giver svarene {a,b}, er fuldstændigheden = $2/3 \approx 66,6\%$. Fuldstændig-



»

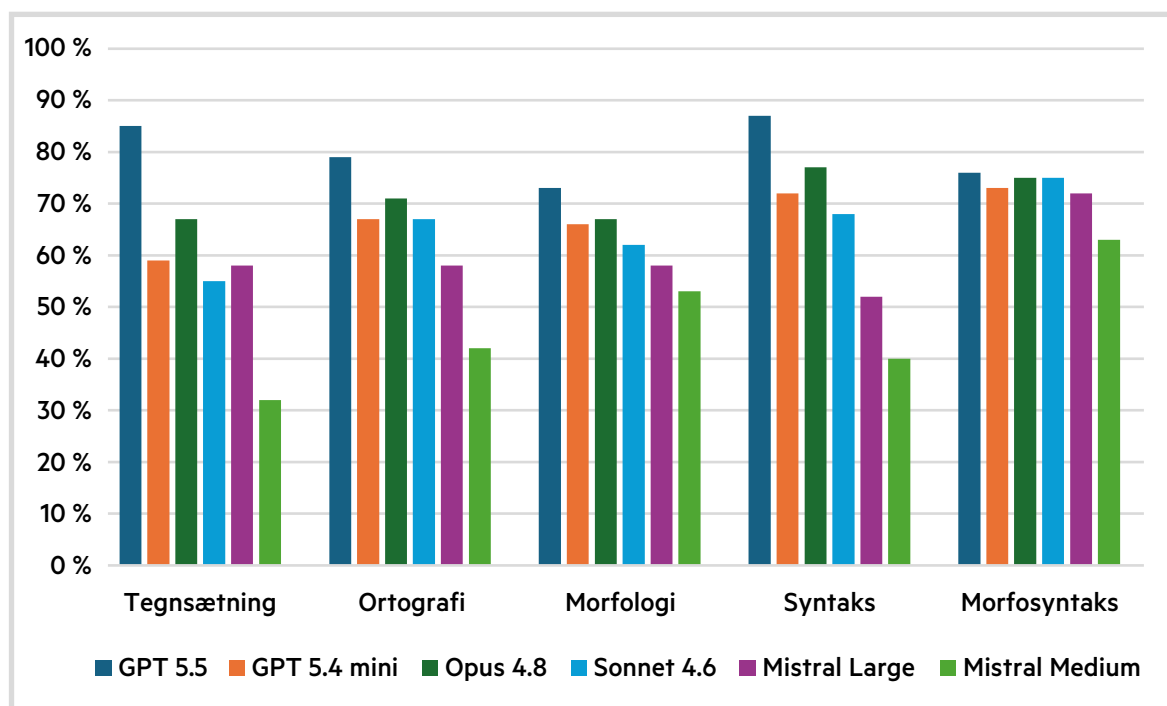
hedsmålet straffer altså oversete, korrekte svar, mens præcisionsmålet straffer forkerte gæt. **F1** er et mål der forener præcision og fuldstændighed, nemlig det harmoniske gennemsnit af præcision og fuldstændighed. At gennemsnittet er harmonisk, betyder at F1 kun er høj hvis både præcision og fuldstændighed er høje. Er den ene lav, falder F1 markant. F1 er altså et samlet mål der vægter præcision og fuldstændighed lige højt og straffer ubalancer mellem mange falske positive (lav præcision) og mange falske negative (lav fuldstændighed).

I modsætning til det eksakte mål gør præcision, fuldstændighed og F1 det muligt at skelne mellem forskellige typer af fejl: Mange falske positive (lav præcision) betyder at modellen gætter

for meget, og mange falske negative (lav fuldstændighed) betyder at modellen er for forsigtig og systematisk overser korrekte svar.

OVERORDNEDE RESULTATER

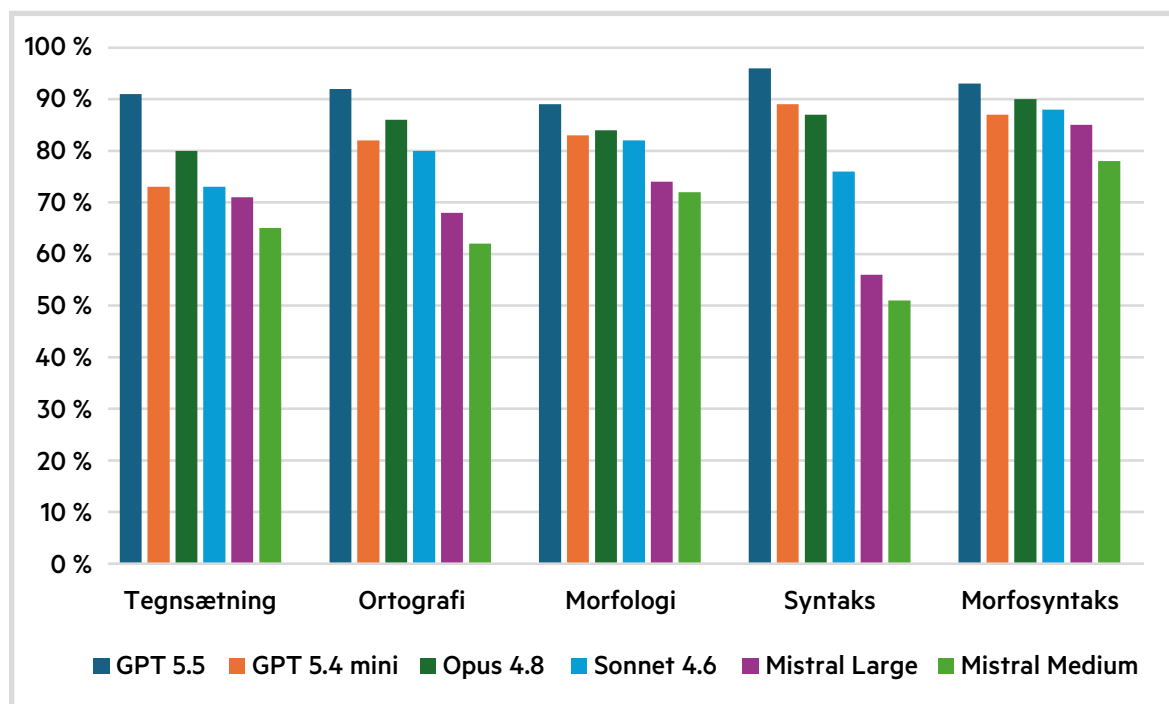
I det følgende giver vi først et overordnet billede af de seks modellers resultater ved at gruppere de 15 sproglige kategorier beskrevet i afsnittet "De stillede spørgsmål" i de fem overkategorier tegnsætning, ortografi, morfologi, syntaks og morfosyntaks. Figurene nedenfor viser at modellerne på trods af forskelle i deres samlede præstation overordnet er stærke og svage indenfor de samme sproglige kategorier: De sproglige kategorier der er vanskelige for én model, er det typisk også for de øvrige modeller.



Figur 1: Gennemsnitlig eksakt-score i overkategorierne

På trods af forskelle i det generelle præstationsniveau følger modellerne stort set det samme mønster ift. sproglige kategorier: Morfosyntaks er den stærkeste overkategori for næsten alle

modeller, ortografi og morfologi ligger i midterfeltet, og tegnsætning og syntaks volder flere af modellerne de største udfordringer. Det er dog i overkategorien syntaks at modellerne adskiller



Figur 2: Gennemsnitlig F1-score i overkategorierne

sig mest fra hinanden. Omvendt er morfosyntaks den overkategori hvor modellerne ligger tættest.

Figurerne viser også tydeligt at modellen GPT 5.5 opnår de bedste resultater i alle fem overkategorier. Især i tegnsætning, der ellers volder flere af de andre modeller de største problemer, skiller GPT 5.5 sig ud ved at klare sig omtrent lige så godt som i de andre kategorier. I den anden ende af skalaen ligger Mistral Medium konsekvent lavt. Opus 4.8, Sonnet 4.6 og GPT 5.4 mini præsterer mere jævnt på tværs af kategorierne uden markante styrker eller svagheder.

I det store og hele adskiller modellerne sig dog fra hinanden ved hvor godt de klarer opgaverne generelt, ikke ved hvilke typer sproglige spørgsmål de har svært ved.

DETALJEREDE RESULTATER

I det følgende gives resultaterne for hver model i tabelform. For hver af de 15 kategorier angives de fire mål beskrevet ovenfor. Øverst i hver tabel

angives desuden modellens gennemsnitlige score på tværs af alle kategorier. De fem overkategorier er adskilt i tabellerne vha. vandrette streger. Procentsatserne er farvekodet efter en fælles farveskala, hvor rød angiver lave værdier, gul mellemværdier og grøn høje værdier. Den samme skala er anvendt i alle tabeller for at lette sammenligningen mellem modellerne. Farvekodningen tjener alene som en visuel hjælp til forståelse af resultaterne.



»

GPT 5.4 mini		Eksakt	Præcision	Fuldstændighed	F1
Gennemsnit		67 %	83 %	81 %	80 %
Apostrof	1	53 %	71 %	74 %	70 %
Punktum	2	65 %	77 %	88 %	80 %
Komma (med startkomma)	3	72 %	82 %	82 %	81 %
Komma (uden startkomma)	4	50 %	55 %	60 %	57 %
Bindestreg	5	58 %	78 %	87 %	79 %
Store og små bogstaver	6	47 %	74 %	72 %	70 %
Sam- og særskrivning	7	74 %	92 %	86 %	87 %
Stavefejl	8	83 %	88 %	91 %	89 %
Morfologi	9	66 %	90 %	80 %	83 %
Syntaks	10	72 %	76 %	77 %	76 %
Adverbielt -t	11	61 %	92 %	77 %	82 %
Hans/hendes eller sin	12	64 %	97 %	83 %	87 %
Pronominalkasus	13	65 %	85 %	77 %	79 %
R-problemer	14	98 %	99 %	98 %	98 %
Blandet morfosyntaks	15	80 %	90 %	86 %	87 %

Tabel 2: Resultater for OpenAI's GPT 5.4 mini

OpenAI's GPT 5.4 mini klarer sig bedst i kategorierne *r*-problemer og stavefejl; næsten alle spørgsmål (98 %) i kategorien *r*-problemer besvares fuldstændig korrekt. Omvendt har den sværest ved komma (uden startkomma) og store og små bogstaver, hvor kun omtrent halvdelen af spørgsmålene besvares korrekt. I kategorierne 1-5 (tegnsætning) har modellen en tendens til at vælge for mange svar. Det ses ved at fuldstændigheden er højere end præcisionen, mens det forholder sig omvendt i kategorierne 10-15 (morfosyntaks). Her er modellen mere forsigtig: Den vælger sjældent forkerte svar, men overser korrekte svar. Samlet set er forskellen mellem modellens gennemsnitlige præcision og fuldstændighed dog kun 2 procentpoint, dvs. at modellen generelt ikke har en særlig udpræget tendens til at være forsigtig. Det giver en F1-score på 80 %, mens modellen i 67 % af tilfældene vælger udelukkende korrekte og alle korrekte svar.

OpenAI's GPT 5.5 er den bedst præsterende model når man sammenligner modellerne i undersøgelsen, med en gennemsnitlig F1-score på 92 % og en eksakt-score på 82 %. Modellen klarer sig bedst i kategorierne *r*-problemer, stavefejl, komma (uden startkomma), punktum og syntaks. Til gengæld er de morfosyntaktiske kategorier adverbielt -t, *hans/hendes* eller *sin* og pronominalkasus blandt de vanskeligste for modellen, når man ser på de eksakte scorer. I disse kategorier er præcisionen højere end fuldstændigheden, hvilket viser at modellen har en tendens til at være forsigtig: Den vælger sjældent forkerte svar, men overser af og til korrekte svar. Som for den mindre model GPT 5.4 mini er forskellen mellem den gennemsnitlige præcision og fuldstændighed dog kun 2 procentpoint, dvs. at modellen generelt ikke har en særlig udpræget tendens til at være forsigtig.



GPT 5.5			Eksakt	Præcision	Fuldstændighed	F1
Gennemsnit			82 %	94 %	92 %	92 %
Apostrof	1		75 %	89 %	83 %	84 %
Punktum	2		93 %	97 %	98 %	97 %
Komma (med startkomma)	3		85 %	94 %	91 %	92 %
Komma (uden startkomma)	4		94 %	95 %	96 %	95 %
Bindestreg	5		80 %	89 %	90 %	88 %
Store og små bogstaver	6		74 %	95 %	87 %	89 %
Sam- og særskrivning	7		80 %	93 %	92 %	91 %
Stavefejl	8		90 %	94 %	97 %	95 %
Morfologi	9		73 %	94 %	88 %	89 %
Syntaks	10		88 %	95 %	100 %	96 %
Adverbielt -t	11		72 %	93 %	93 %	91 %
Hans/hendes eller sin	12		68 %	99 %	85 %	89 %
Pronominalkasus	13		73 %	95 %	89 %	90 %
R-problemer	14		99 %	100 %	100 %	100 %
Blandet morfosyntaks	15		86 %	96 %	96 %	95 %

Tabel 3: Resultater for OpenAI's GPT 5.5

Sonnet 4.6			Eksakt	Præcision	Fuldstændighed	F1
Gennemsnit			66 %	81 %	82 %	80 %
Apostrof	1		36 %	61 %	69 %	62 %
Punktum	2		59 %	74 %	88 %	78 %
Komma (med startkomma)	3		71 %	85 %	88 %	85 %
Komma (uden startkomma)	4		51 %	58 %	64 %	60 %
Bindestreg	5		57 %	73 %	88 %	78 %
Store og små bogstaver	6		56 %	78 %	70 %	72 %
Sam- og særskrivning	7		65 %	85 %	81 %	81 %
Stavefejl	8		81 %	85 %	88 %	86 %
Morfologi	9		62 %	86 %	82 %	82 %
Syntaks	10		69 %	75 %	80 %	76 %
Adverbielt -t	11		59 %	88 %	77 %	80 %
Hans/hendes eller sin	12		76 %	98 %	90 %	92 %
Pronominalkasus	13		76 %	94 %	87 %	89 %
R-problemer	14		90 %	92 %	92 %	91 %
Blandet morfosyntaks	15		78 %	89 %	92 %	89 %

Tabel 4: Resultater for Anthrops Sonnet 4.6

Sonnet 4.6 opnår en gennemsnitlig F1-score på 80 % og en eksakt-score på 66 %, hvilket placerer den på niveau med GPT 5.4 mini. Det er

kategorierne *r*-problemer og stavefejl (målt på eksakt-score) modellen klarer sig bedst i, mens tegnsætningskategorierne apostrof og kommate-



»

ring uden startkomma volder de største vanskeligheder for denne model. I modsætning til GPT 5.5 er den gennemsnitlige fuldstændighed en anelse højere end præcisionen (82 % mod 81 %),

hvilket tyder på at modellen har en svag tendens til at vælge lidt for mange svar. Forskellen er dog kun på et enkelt procentpoint, så overordnet er modellen velafbalanceret i sine svar.

Opus 4.8

		Eksakt	Præcision	Fuldstændighed	F1
Gennemsnit		72 %	87 %	87 %	85 %
Apostrof	1	60 %	76 %	78 %	75 %
Punktum	2	64 %	79 %	95 %	84 %
Komma (med startkomma)	3	84 %	90 %	94 %	91 %
Komma (uden startkomma)	4	60 %	67 %	76 %	70 %
Bindestreg	5	69 %	81 %	86 %	82 %
Store og små bogstaver	6	60 %	86 %	86 %	83 %
Sam- og særskrivning	7	72 %	89 %	84 %	85 %
Stavefejl	8	84 %	89 %	92 %	90 %
Morfologi	9	67 %	88 %	85 %	84 %
Syntaks	10	78 %	85 %	91 %	87 %
Adverbielt -t	11	70 %	96 %	85 %	88 %
Hans/hendes eller sin	12	67 %	100 %	84 %	89 %
Pronominalkasus	13	76 %	95 %	86 %	89 %
R-problemer	14	93 %	94 %	94 %	94 %
Blandet morfosyntaks	15	78 %	90 %	91 %	89 %

Tabel 5: Resultater for Anthropic's Opus 4.8

Opus 4.8 er med en gennemsnitlig F1-score på 85 % og en eksakt-score på 72 % den næstbedst præsterende model i undersøgelsen. Det er kategorierne *r*-problemer, stavefejl, kommatering med startkomma og syntaks som modellen klarer sig bedst i, mens apostrof og kommatering uden startkomma er de vanskeligste kategorier for modellen. Til forskel fra både GPT 5.4 mini, GPT 5.5 og Sonnet 4.6 er modellens gennemsnitlige præcision og fuldstændighed identiske (87 %), hvilket indikerer at den hverken gætter for meget eller er for forsigtig.

Mistral Medium er den svagest præsterende model i undersøgelsen med en gennemsnitlig F1-score på 68 % og en eksakt-score på 46 %. Modellen klarer sig som de øvrige modeller bedst i kategorien *r*-problemer. Generelt klarer modellen

sig bedst i de morfosyntaktiske kategorier med undtagelse af adverbielt -t, mens tegnsætning volder de største problemer, navnlig apostrof, punktum og kommatering med startkomma. Til forskel fra de andre modeller har Mistral Medium en tydelig tendens til at vælge for mange svar: Den gennemsnitlige fuldstændighed er 74 %, mens præcisionen kun er 68 %. Modellen overser altså relativt få korrekte svar, men vælger oftere forkerte svar.



Mistral Medium			Eksakt	Præcision	Fuldstændighed	F1
Gennemsnit			46 %	68 %	74 %	68 %
Apostrof	1		19 %	50 %	70 %	56 %
Punktum	2		34 %	61 %	93 %	70 %
Komma (med startkomma)	3		25 %	59 %	90 %	69 %
Komma (uden startkomma)	4		42 %	56 %	71 %	61 %
Bindestreg	5		39 %	64 %	85 %	71 %
Store og små bogstaver	6		26 %	59 %	63 %	57 %
Sam- og særskrivning	7		47 %	66 %	62 %	62 %
Stavefejl	8		57 %	64 %	69 %	66 %
Morfologi	9		53 %	77 %	72 %	72 %
Syntaks	10		40 %	49 %	55 %	51 %
Adverbielt -t	11		39 %	71 %	58 %	62 %
Hans/hendes eller sin	12		62 %	93 %	80 %	84 %
Pronominalkasus	13		67 %	86 %	78 %	80 %
R-problemer	14		80 %	84 %	87 %	85 %
Blandet morfosyntaks	15		69 %	81 %	79 %	79 %

Tabel 6: Resultater for Mistral Medium

Mistral Large			Eksakt	Præcision	Fuldstændighed	F1
Gennemsnit			62 %	77 %	74 %	74 %
Apostrof	1		40 %	60 %	64 %	60 %
Punktum	2		68 %	75 %	79 %	76 %
Komma (med startkomma)	3		73 %	84 %	86 %	84 %
Komma (uden startkomma)	4		42 %	51 %	59 %	53 %
Bindestreg	5		65 %	79 %	84 %	80 %
Store og små bogstaver	6		45 %	67 %	56 %	60 %
Sam- og særskrivning	7		55 %	72 %	64 %	67 %
Stavefejl	8		74 %	77 %	77 %	77 %
Morfologi	9		58 %	82 %	71 %	74 %
Syntaks	10		52 %	56 %	58 %	56 %
Adverbielt -t	11		64 %	93 %	79 %	83 %
Hans/hendes eller sin	12		69 %	97 %	83 %	88 %
Pronominalkasus	13		70 %	88 %	79 %	82 %
R-problemer	14		85 %	87 %	87 %	87 %
Blandet morfosyntaks	15		77 %	87 %	83 %	84 %

Tabel 7: Resultater for Mistral Large

Mistral Large opnår en gennemsnitlig F1-score på 74 % og en eksakt-score på 62 %, hvilket placerer den i den nedre halvdel af feltet, men klart over

Mistral Medium. Modellen klarer sig som de andre bedst i kategorien *r*-problemer. Målt på F1-score har modellen lettest ved de morfosyntaktiske



» spørgsmål (kategori 11-15), mens kommatering uden startkomma, apostrof, store og små bogstaver samt sær- og samskrivning volder de største problemer. Den gennemsnitlige præcision er en smule højere end fuldstændigheden (77 % mod 74 %), hvilket tyder på at modellen har en svag tendens til at være forsigtig: Den vælger sjældnere

forkerte svar, men overser til gengæld af og til korrekte svar. Mest udtalt er denne tendens i kategorierne adverbialt *-t* og *hans/hendes* eller *sin*.

RANGORDNING AF MODELLERNE

Tabellen nedenfor viser en rangordning af de seks modeller for hvert af de fire evalueringsmål.

Rang	Eksakt	Præcision	Fuldstændighed	F1
1.	GPT 5.5 (82 %)	GPT 5.5 (94 %)	GPT 5.5 (92 %)	GPT 5.5 (92 %)
2.	Opus 4.8 (72 %)	Opus 4.8 (87 %)	Opus 4.8 (87 %)	Opus 4.8 (85 %)
3.	GPT 5.4 mini (67 %)	GPT 5.4 mini (83 %)	Sonnet 4.6 (82 %)	GPT 5.4 mini (80 %)
4.	Sonnet 4.6 (66 %)	Sonnet 4.6 (81 %)	GPT 5.4 mini (81 %)	Sonnet 4.6 (80 %)
5.	Mistral Large (62 %)	Mistral Large (77 %)	Mistral Medium (74 %)	Mistral Large (74 %)
6.	Mistral Medium (46 %)	Mistral Medium (68 %)	Mistral Large (74 %)	Mistral Medium (68 %)

Tabel 8: Rangordning af sprogmodellerne

Tabellen viser tydeligt at OpenAI’s model GPT 5.5 er den bedst præsterende model på alle fire evalueringsmål. Anthropic’s Opus 4.8 indtager andenpladsen. GPT 5.4 mini og Sonnet 4.6 ligger meget tæt og udgør et midterfelt. GPT 5.4 mini scorer en anelse højere på eksakt, præcision og F1, mens Sonnet 4.6 har en marginalt højere fuldstændighed. Mistral’s to modeller placerer sig nederst i rangordningen; især Mistral Medium adskiller sig ved at opnå markant lavere scorer end de øvrige modeller. På tværs af alle fire evalueringsmål er rangordningen bemærkelsesværdigt stabil. Det viser at modellerne præsterer konsistent uanset hvilket evalueringsmål der anvendes. Som det fremgår, er det hver virksomheds betalingsmodel der klarer sig bedst.

SAMMENFATNING OG DISKUSSION

Vores undersøgelse viser først og fremmest at chatbotter kan være ganske gode til sproglig korrekthed, men også at der er store forskelle mellem de enkelte modeller. Den bedst præsterende model, GPT 5.5, opnåede en eksakt-score på 82 %³, mens den svageste model, Mistral Medium, havde en eksakt-score på 46 %. Undersø-

3 Til sammenligning viser den os bekendt nyeste større undersøgelse af 9.-klasseelevers stavefærdighed at eleverne ved besvarelsen af diktatdelen af afgangsprøven i retskrivning havde en gennemsnitlig rigtighedsprocent på 78 (Jervelund & Schack 2016: 13), der bedst kan sammenlignes med præcisionsscoren i denne undersøgelse. I diktaten testes elevernes færdigheder i ortografi samt visse morfosyntaktisk betingede staveformer, dvs. færdigheder som i vid udstrækning svarer til de spørgsmål som vi har stillet til chatbotterne. Elevernes og chatbotternes rigtighedsprocenter er ikke direkte sammenlignelige men giver dog et fingerpeg om forholdet mellem skoleelevers og chatbotters sproglige færdigheder.



gelsen viser desuden at det er de modeller man betaler for, der overordnet set klarer sig bedst. Det er altså ikke ligegyldigt hvilken chatbot man bruger, hvis man vil have hjælp til sproglig korrekthed.

På trods af forskellene mellem modellerne er der også et fælles mønster i deres styrker og svagheder når det gælder sproglige kategorier:

Nemmest: *r*-problemer og stavefejl

Sværest: tegnsætning, især apostrof, og derudover store/små bogstaver samt enkelte morfosyntaktiske fænomener, især adverbialt *-t*.

At spørgsmål om *r*-problemer gennemgående er de letteste for chatbotterne, er interessant når man sammenholder dette med tidligere undersøgelser, der viser at både elever i folkeskolen og studerende på universitetet har store udfordringer med netop *r*-fejl (Jervelund & Schack 2016, Rathje 2019). At chatbotterne også har styr på spørgsmålene om stavefejl (fejl i ordets stamme), er en god nyhed, set i lyset af at dette ifølge undersøgelser er en af de hyppigste fejltyper i både universitetsopgaver og danske aviser (Rathje 2019, Rathje et al. 2023). *R*-fejl og stavefejl kan man altså generelt godt bruge chatbotterne som hjælp til.

Især tegnsætning volder derimod modellerne problemer, og tegnsætning, især Kommatering, er samtidig den sproglige kategori Sprognævnet hyppigst får spørgsmål til (jf. Dansk Sprognævn 2025: 20). Så forestillingen om at chatbotterne har styr på ”korrekt tegnsætning”, som det blev antaget i artiklens indledning, kan altså med vores undersøgelse afkræftes. Modellen GPT 5.5 skiller sig dog ud ved også at klare de fleste tegnsætningsspørgsmål godt. Vores undersøgelse viser også at har man problemer med apostrof,

store og små bogstaver og adverbialt *-t*, skal man heller ikke forlade sig ukritisk på sproglig hjælp fra sin chatbot.

Resultaterne nuancerer den forestilling om chatbotters næsten fejlfrie sprog som vi tog udgangspunkt i i indledningen. Vores resultater viser at der er noget om chatbotternes ry for at være sprogligt kompetente, men de viser samtidig at *god* ikke er det samme som *fejlfri* eller *sikker*. I selv den bedste models svar var der mindst én fejl eller forglemmelse i ca. hvert femte spørgsmål.

Dermed kan vores resultater også være med til at forklare de tilsyneladende modsatrettede konklusioner i tidligere undersøgelser. Blom et al. (2024) fandt betydeligt færre sproglige fejl i tekster skrevet af ChatGPT end i tekster skrevet af universitetsstuderende, mens det norske Språkrådet (2025a, 2025b) på baggrund af sine undersøgelser konkluderede at chatbotter stadig begår så mange fejl at deres output bør kontrolleres af et menneske. De to konklusioner behøver imidlertid ikke at være modstridende. En chatbot kan godt være bedre end mange mennesker til sproglig korrekthed (fx universitetsstuderende) og samtidig være for usikker til at fungere som en autoritativ korrekturlæser.

For den almindelige danske bruger er konklusionen derfor at en chatbot godt kan bruges som sproglig hjælper, men at den ikke bør betragtes som en sproglig facitliste: Brugeren bør være opmærksom på at chatbotten også kan give forkerte svar og ikke sjældent overser korrekte svar.

Vores næste skridt bliver at undersøge om modellerne kan identificere og rette fejl i autentiske tekster. Indtil da peger denne undersøgelse på at chatbotterne kan være gode sproglige assistenter, men at jagten på den gode og sikre chatbot endnu ikke er afsluttet.



»

REFERENCE:

Blom, Jonas Nygaard, Alexandra Holsting & Jesper Tinggaard Svendsen (2024): På sporet af chatbottens sproglige fingeraftryk: En sproglig ophavsanalyse af tekster skrevet af danskstuderende og ChatGPT. I: *NyS - Nydanske Sprogstudier*, 1(65), 79–110. <https://doi.org/10.7146/nys.v1i65.142936>

Chatterji, Aaron, Thomas Cunningham, David J. Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan & Kevin Wadman (2025): How People Use ChatGPT. *NBER Working Paper* 34255. Cambridge, Massachusetts: National Bureau of Economic Research. <https://doi.org/10.3386/w34255>

Copenhagen Review of Communication (2025): Snif snif det lugter af AI: Hvordan sniffer du egentlig om en tekst er skrevet af AI?, *Copenhagen Review of Communication*, 9. oktober 2025. <https://www.copenhagen-review-of-communication.com/snif-snif-det-lugter-af-ai-hvordan-sniffer-du-egentlig-om-en-tekst-er-skrevet-af-ai/>, lokaliseret 16. juni 2026.

Danmarks Statistik (2025): <https://www.dst.dk/da/Statistik/udgivelser/NytHtml?cid=51815>, lokaliseret 12. juni 2026.

Dansk Sprognævn (2025): *Årsberetning 2025*. <https://dsn.dk/wp-content/uploads/2026/05/Aarsberetning-2025.pdf>

Jarvad, Pia (2014): *Spørg om sprog – 40 år i Sprognævnets tjeneste*. Dansk Sprognævns skrifter 44. København: Dansk Sprognævn.

Jervelund, Anita Ågerup & Jørgen Schack (2016): *Den der skriver d i gjort. Staveproblemer i folkeskolens ældste klasser*. Dansk Sprognævns skrifter 48. København: Dansk Sprognævn.

Larsen, Rasmus & Torben Blach (2026): Munin 1.0-udgivelsesnote. <https://www.foundationmodels.dk/da/news/2026/06/11/munin-10-udgivelsesnote.html>, lokaliseret 6. august 2026.

Nygaard, Claus (2024): *Prompt engineering. En grundbog*. København: Samfundslitteratur.

Obitsø, Ole (2026): Tekstforfatter om AI: Det er lynhurtigt blevet en del af mit arbejde. *Dansk Markedsføring*, 15. juni 2026.

OpenAI (2026a): Reasoning models. <https://developers.openai.com/api/docs/guides/reasoning>, lokaliseret 6. august 2026.

OpenAI (2026b): Using GPT-5.5. <https://developers.openai.com/api/docs/guides/latest-model?model=gpt-5.5>, lokaliseret 6. august 2026.

Rathje, Marianne (2019): Afvigelser fra retskrivningsnormen i universitetsopgaver. I: Yonatan Goldshtein, Inger Schoonderbeek Hansen & Tina Thode Hougaard (red.). *17. Møde om Udforskningen af Dansk Sprog*. Institut for Kommunikation og Kultur, Aarhus Universitet, Århus, 415–437.

Rathje, Marianne, Jørgen Schack, Jonas Blom & Eckhard Bick (2023): Stavefejl i aviserne i 1999 og 2019. I: *Nyt fra Sprognævnet* 2023/2, 6–16.

Språkrådet (2025a): *KI-språkets fallgruver*, publiceret 14. januar 2025 på https://sprakradet.no/aktuelt/ki-sprakets-fallgruver/?utm_source=chatgpt.com

Språkrådet (2025b): *Språket i KI-genererte tekster. Rapport fra Språkrådet om språklig kvalitet i tekster produsert av prateroboter basert på fire språkmodeller*, oktober 2025. <https://sprakradet.no/wp-content/uploads/Rapport-fra-test-av-sprakroboter-2025.pdf>

Søby, Katrine Falcon & Line Burholt Kristensen (2019): Hjælp! Jeg har mistede min yndlings rød taske: Et studie af grammatikafvigelser. I: *Ny forskning i grammatik* 26, 89–104.

Villemoes, Søren K. (2026): Slopificeringen, *Weekendavisen*, 29. maj 2026.



Ord vi hader, og ord vi elsker (rapport fra en sprogfestival)

Ved festivalen Sprogense, tilbage i maj måned, fik jeg mulighed for at spørge en hel sal fuld af sproginteresserede om hvilke ord i det danske sprog de holder mest af, og hvilke de ikke kan fordrage. Resultatet var ikke som forventet, faktisk langt fra, og det vækker en mistanke. Måske mener danskerne noget andet om vores sprog end det officielle Danmark går og tror?



**PETER JUEL
HENRICHSEN**
(f. 1962) er senior-
forsker i Dansk
Sprognævn.



ESTIVALLLEN SPROGENSE

Sprogense er den årlige sprogfestival i Bogense. Hvert år mødes mere end 5000 festivalgæster

i købstaden ved Fyns nordkyst og fejrer det danske sprog i al sin pragt. Alle er velkomne, og den næste Sprogense er d. 30/4-1/5 2027 (se sprogense.dk). I år havde festivalarrangørerne fået lov til at lægge en foredragsserie i rådhusets fine saloner, og det var tiltrængt, for byens øvrige sale kan ikke længere rumme alt det der foregår på festivalen. Mit foredrag var dagens sidste i rådhusalen, og som allerede nævnt sluttede jeg med et spørgsmål til publikum: Hvilke danske ord kan I bedst lide? Og hvilke mindst? Denne lille artikel giver et overblik over publikums svar. Selv om ordmaterialet ikke er stort, sætter det alligevel skub i nogle tanker om det danske sprogs udvikling (og især hvor skoen trykker).

ORD DER VÆKKER FØLELSER

En almindelig dansk sprogbruger kender mere end 50.000 ord. Hvordan kan det være at nogle få af dem stikker ud og vækker særligt stærke følelser?

Det kan jo simpelthen skyldes deres betydning. Nogle ord er knyttet til gode oplevelser og positive tanker, såsom *ferie*, *glad* og *ven-skab*. Dem vil vi betegne som **plusord** (med en velkendt term fra 70'erne). Deres modsætning, **minusord**, er dem der giver ubehagelige associationer, for eksempel *slaveri*, *sorg* og *kræftsyg*. Plusord og minusord er både almindelige og gammelkendte, og de er slet ikke mulige at undvære i ordforrådet.



»

Anderledes er det med de nye ord der pludselig dukker op og gatecrasher sproget. Nye ord er fortræffelige til at vække følelser, ikke mindst ærgrelse, men også munterhed og begejstring. Særlig omtalt er de **engelske nyord** (fx *goat*, *floor*, *at onboarde*), men også **danske nyord** kan få blodtrykket op hos mange (*bedekvinde*, *jernmark*, *at vækste*).

Og endelig kan man som sprogbruger have sine helt private og personlige grunde til at føle stærkt for et ord, måske fordi det vækker gode eller dårlige minder. Blandt publikums hadeord fik vi for eksempel både *inderlighed* og *gudfrygtig*. Man kan ikke lade være at tænke på hvordan disse to ord – der engang var klare plusord – er endt som vederstyggeligheder for nogle enkelte. Ord med en særlig personlig valør vil vi i mangel af bedre kalde **privatord**.

Man kunne let foreslå flere kategorier, men til denne lille undersøgelse er det nok med de fem herover (mere om Sprognævnets metoder sidst i artiklen).

INDSAMLINGEN FRA PUBLIKUM

Ekperimentet i rådhusalen gav en smuk samling af elske- og hadeord (omhyggeligt nedskrevet af publikum på små sedler med en grøn og en rød side). Indsamlingen gav over hundrede ord, som efterfølgende blev ordnet i kategorierne herover. Jeg må indrømme at jeg havde nogle ret stereotype forventninger til publikums smag. Da nu de engelske nyord fylder så meget i sprogdebatten, forventede jeg at ord som *stealth*, *cringe*, *CEO* og *business school* ville fylde godt op i den røde liste, mens den grønne var rig på de tilsvarende fordanskninger (*kondomsnyd*, *pinligt*, *direktør*, *handelshøjskole*, osv.). Der fik jeg en lang næse.

Men før afsløringen af dens længde er jeg nødt til at fortælle om en begivenhed der indtraf

i rådhusalen lige inden indsamlingen og truede med helt at umuliggøre den.

INTERMEZZO I RÅDHUSALEN

Mit foredrag var begyndt med en gennemgang af det danske sprogs udvikling fra 4000 f.Kr. til nutiden. For ikke at tabe tråden undervejs havde jeg brugt den skønne plakat som Mål & Mæle udgav i år 2000 som illustration. Plakaten viser sproghistorien som en flod i stadig bevægelse, fra sit udspring i proto-indoeuropæisk til det dansk vi taler i dag, med bifloder der dels forlader hovedfloden – som da det nordiske ordforråd flød ind i det engelske sprog i vikingetiden – og dels løber til hovedfloden fra de dominerende europæiske sprog, fra nedertysk i middelalderen, fra fransk i oplysningstiden og i nyere tid især fra engelsk. Ned ad den brede flod sejler en ung kvinde på et surfbræt med en laptop foran sig, særdeles nutidigt udseende. Hende kaldte jeg frøken Arild; og det vågne og snilde publikum forstod ordspillet med det samme. Hun var nu foredragets turguide gennem sprogstadierne, og takket være hende nåede jeg tørskoet helt frem til nudansk. Her havde jeg tænkt mig at slå et slag for at det danske sprog slet ikke er så truet af nye **ord** – engelske, arabiske, slang, gadesprog, lingo – men langt mere truet af dets nye **sprogbrugere**. Vores elektroniske hjælpemidler begynder nemlig også at blande sig i samtalen. I dag opfinder chatbots helt nye ord som vi mennesker begynder at bruge uden at ane at de er kunstige. Og det er bare begyndelsen.

DEN LUKKEDE DØR

På dette sted i foredraget havde jeg forberedt en dommedagsprofeti, og meningen var at frøken Arild på mit stikord skulle brase ind i lokalet og afbryde seancen med ordene "Det er helt okay at du står og fortæller folk om sproget i gamle



dage. Og det er også okay at du skræmmer livet af dem med din robotsnak. Men jeg er faktisk mere interesseret i de ord vi går rundt og siger lige nu!”. Jeg havde derfor allieret mig med min datter, en fortrinlig amatørskuespiller. I kostume og med paryk, solbriller og surfbræt lignede hun frøken Arild til forveksling (se billedet herunder). Uheldigvis havde vi ikke fået holdt generalprøve (mit foredrag var det sidste af en lang, ubrudt række i rådhusalen), og derfor kom det bag på frøken Arild at den lille gang som hun gemte sig i, og som førte lige op til scenen, var for smal til at tillade passage af surfbrættet. Hun sad fast. Imens fortsatte jeg min brandtale om sprogets forestående nederlag, men den forberedte del af min tekst slap hurtigt op eftersom den jo kun skulle tjene til stikord, og jeg måtte improvisere videre ud ad en flod der nu begyndte at minde

mere om en dæmning med hul i. Talende, gestikulerende og efterhånden perspirerende kantede jeg mig over mod den lukkede dør og forsøgte at banke diskret på den, med hundrede blikke klistret til min besynderlige opførsel. Krise. Var frøken Arild faldet i staver?

FRØKEN ARILDS ENTRE

Pludselig sprang døren op. Surfbrættet kom ind først, ikke som på plakaten, men snarere som bovsprødet på en fregat fra pirattiden med frøken Arild som galionsfigur – heldigvis slagfærdig som altid og med fuld kontrol over replikkerne dérfra. Publikum lod venligt som om de ikke havde bemærket noget usædvanligt, og endelig kunne hun indtræde i den planlagte rolle som seddel-uddeler og -indsamler. Og dermed lykkedes det at indsamle de to ordlister som denne artikel



» egentlig handler om. Hvilken indflydelse episoden har haft på det indsamlede materiale, tør jeg ikke gisne om.

ORSEDLERNE OPTALT OG SORTERET

Over halvtreds personer afleverede en udfyldt ordseddel, og mange afgav endda tre eller flere ordforslag. I alt høstede vi 80 forskellige ord, nye som gamle (se detaljerne i tabellen).

Hvad angår de grønne sedler, elskeordene, er der ingen tvivl om deres fordeling. Fire ud af fem er gode og veltjente danske gloser som har stået i *Retskrivningsordbogen* og de andre ordbøger, hvis ikke fra arilds tid, så i alt fald i generationer. Det er plusord som *ferie*, *forår*, *venskab*, *livsglad* og privatord som *purløg*, *krabask*, *snobrød*, *nul-lermænd*. Der er kun ganske få egentlige nyord¹ blandt de grønne ord, heriblandt *dogmestrik*, *haveglæde* og *ketchupere* (sidstnævnte er done-ret af en 59-årig mand og forklaret med citatet: ”Mennesket skal selv kunne folde sin faldskærm, så man ikke bliver *ketchuperet* når man lander.”). Og der er minsandten også en enkelt anglicisme: *floor*.

Den røde afdeling er meget mere sammensat. Over halvdelen er privatord, en broget samling af personlige irritationer som *samvær*, *medarbejdersamtale*, *øjebæ*, *at maksimere*, *lodret uenig*, *ha’ en god dag*, *sindssygt* (som forstærkningsord), *spurgt* (slang for ’jeg er ligeglad’); *fæo* (ordet ’færge’ i nyere udtale), *pizzaria* (stavefejl). Der er også adskillige minusord i den røde liste (fx *sorg*, *corona*, *komplot*); men interessant nok endnu flere plusord (*gudfrygtig*, *inderlig*, *agil*, *fremadrettet*, *fantastisk*, *at vækste*). Plusord er jo ord der oprindeligt er ment som positive,

men hvis de bruges for tit, kan de hurtigt få den omvendte virkning.

Og hvad så med de engelske nyord? Der var kun fire: *at shine*, *at onboard*, *goat* og *sale*.

OPSAMLING

Det mest overraskende i denne lille undersøgelse er altså det vi **ikke** ser. Der er ingen samlet front mod de engelske indlån, intet der tyder på et ønske om sprogrensning, ingen store tendenser, faktisk bare en bred interesse for sproget og en udtrykt glæde ved vores ordforråd. Og så nogle sten i skoen.

Karakteristisk nok er *sale* det eneste nyord i hele materialet der fik mere end to stemmer. Netop dét ord lyder for mig mere som et ekko af mediedebatten end som en personlig iagttagelse. De kommentatorer vi lytter til, har som bekendt stor indflydelse på vores holdninger. En livlig sprogdebat er både god og nødvendig; men vi må ikke glemme at indsatsen for at bevare og styrke det danske sprog skal have et demokratisk udgangspunkt og ikke styres af smagsdommere og æqqehoveder (som undertegnede).



Frøken Arild. Udsnit af plakat (*Mål & Mæle*, 2000). Se hele plakaten på <https://webshophum-dk.ku.dk/shop/maal-og-maeles-406p.html>

¹ Ord der har været i dansk i mere end 30 år, fx *internet*, *okay*, *tackle*, *Udkantsdanmark*, regnes ikke for nyord i denne artikel (se belæg i Sprognævnets ordbøger, *Den Danske Ordbog* og *Ordbog over det Danske Sprog*).



GIV DIT BESYV MED

Ved årets Sprogense-festival skød Sprognævnet en stor kampagne i gang under navnet ”Kan vi på dansk?” for at blive klogere på **hele** befolkningens synspunkter og følelser (med særligt henblik på indflydelsen fra engelsk). Da artiklen her blev skrevet, havde mere end 2300 danskere responderet, fordelt over alle aldre, køn og landsdele.² Læs mere om hele kampagnen på Sprognævnets hjemmeside (dsn.dk).

Kategori	Hadeord	Elskeord
plusord	6	17
minusord	4	0
danske nyord	2	7
engelske nyord	4	1
privatord	22	17

Indsamlingens 80 forskellige ord, fordelt på kategorier (ulæselige sedler blev frasorteret).



2 Blandt de over 9000 ordforslag vi på daværende tidspunkt havde modtaget, var *sale* langt det hyppigste. Det havde mere end dobbelt så mange stemmer som nummer to.

Eskimo

I sensommeren 2016 udtalte en tidligere kollega i Sprognævnet til Ekstra Bladet at nogle danske sprogbrugere opfatter ordet *eskimo* som racistisk, og at man derfor står sig bedre ved at bruge ordet *inuit*, som mange grønlændere bruger om sig selv.



JØRGEN SCHACK
(f. 1961) er seniorforsker i Dansk Sprognævn.



MARIANNE RATHJE
(f. 1973) er seniorforsker i Dansk Sprognævn.

FAKTA – FARLIGE ORD

Overskriften *Farlige ord* er temaet for en række artikler som bliver bragt i *Nyt fra Sprognævnet* i 2026. Med "farlige ord" menes i denne sammenhæng ord som tilhører det identitetspolitiske område, og som i en eller flere henseender kan opfattes som kontroversielle, dvs. ord som fx *brun*, *bøsse*, *eskimo*, *etnisk*, *hudfarvet*, *indianer*, *slave(gjort)*. I dette nummer bringer vi en artikel med titlen "Eskimo".



NLEDNINGEN TIL AVISENS HENVENDELSE

var en begyndende debat om isnavnet *Kæmpe Eskimo*. Kollegaens udtalelse til

Ekstra Bladet blev stærkt fordrejet i flere aviser, bl.a. i metroxpress 11.9.2016: "Is-gigant anklaget for racisme." Det gik så vidt at Sprognævnets daværende direktør, Sabine Kirchmeier, følte sig nødsaget til at skrive et dementi, som afsluttes med ordene:

Sprognævnets velmente råd og vejledning om at man skal vælge sine ord med omhu, bliver fremstillet som anklager om racisme og krav om at produkter skal trækkes tilbage. Vi oplever desværre ofte at vi bliver påduttet en rolle som vi ikke har, men denne her tager prisen. (dsn.dk 12.9.2016)

Debatten om *Kæmpe Eskimo* og andre ispinde hvis navne indeholder ordet *eskimo*, vendte tilbage med fornyet styrke i sommeren 2020. Nogle isproducenter ændrede meget hurtigt navnene på deres ispinde. Den vel nok bedst kendte ispind af "eskimotypen", Friskos *Kæmpe Eskimo*, ændrede navn til *Kæmpe Solbær* i 2021. Det er sandsynligvis de omtalte isdebatter der har givet ordet *eskimo* det sidste skub ud i kulden, i hvert fald i almensproget; blandt fagfolk har man længe opfattet ordet som utidssvarende.

I knap 100 år kunne man uddanne sig til eskimolog på Københavns Universitet. Man læste faget eskimologi, og når man var færdiguddannet, kunne man kalde sig *eskimolog*. Faget eksisterer stadig, men det har siden 2019 heddet



grønlandske og arktiske studier. Navneændringen har dog ikke noget med debatten om isen (*Kæmpe*) *Eskimo* at gøre. I en artikel om emnet i Forskerforum 27.8.2020 oplyser institutlederen på Institut for Tværkulturelle og Regionale Studier, Københavns Universitet, at det nye navn, *grønlandske og arktiske studier*, ikke var resultatet af den krænkellesdebat der udspandt sig i forbindelse med navnet på den ulyksalige ispind: "Det havde været et ønske fra fagmiljøet i årevis, simpelthen fordi navnet var utidssvarende og begyndte at lyde mærkeligt og gammeldags – lidt som når man siger *cikle* i stedet for *cykle*." En lektor ved faget grønlandske og arktiske studier giver i samme artikel udtryk for en tilsvarende opfattelse: Navneændringen var "ikke et spørgsmål om at imødekomme krænkede grønlandere. Dem var han nemlig ikke stødt på." Der havde ifølge lektoren været en diskussion i sommeren 2020, men ikke tidligere: "Det har nok mere været i Canada, hvor ordet *eskimo* ikke har været helt politisk korrekt. I Alaska er man gået mindre op i om man kalder det *inuit* eller *eskimo*, måske fordi de kalder sig *yupik*." *Eskimo* gav ifølge lektoren forkerte associationer i retning af noget forhistorisk, men faget har i virkeligheden et stærkt nutidigt indhold: "Det nye navn signalerer, at vi har fokus på Grønland, men også arktisk historie og kultur."

Sådan forholder sagen sig altså ifølge eksperterne på området: Ordet *eskimo* og fagbetegnelsen *eskimologi* er utidssvarende snarere end krænkende. Det ændrer dog ikke ved at *eskimo* i dag af mange danskere opfattes som krænkende. Ifølge **Den Danske Ordbog** kan ordet "virke

stødende, idet betegnelsen *inuit* nu foretrækkes af disse folkeslag frem for *eskimo*." At *eskimo* af nogle grønlandere betragtes som et ord der bør undgås, fremgår af at vejnavnet *Eskimovej* i Århus Kommune pr. 2. marts 2026 er ændret til *Silavej*. Ændringen skyldes en henvendelse i april 2025 fra den grønlandske studenterforening, Avalak (Århus), og Det Grønlandske Hus "med ønske om et mere neutralt vejnavn".¹

Før isdebatterne optrådte *eskimo* ret sjældent i dagspressen. Vi kender alle sammen ordet, men har sjældent haft anledning til at bruge det. I Danmark taler vi af gode grunde ofte om *grønlandere*; de beslægtede befolkningsgrupper i den øvrige del af Arktis taler vi ikke ret meget om, men hvis man har behov for at betone etnicitet snarere end nationalitet, bør man i dag bruge ordet *inuit*, i hvert fald hvad angår grønlandske forhold; *eskimo* opfattes ikke længere som passende.



¹ <https://aarhus.dk/nyt/teknik-og-miljoe/2026/februar-2026/eskimovej-skifter-navn>



Perkle rundt



Jeg er ved at læse romanen "Det jeg elskede" af den amerikanske forfatter Siri Hustvedt, og der er jeg stødt på et udtryk jeg ikke kender, nemlig *perkle rundt*. I den engelske version af bogen står der "Bill filmed children crawling at high speed in the park", og det er oversat til "Bill filmede børn der perkede rundt i parken på alle fire ...". Jeg kan jo godt nogenlunde regne ud hvad *perkle* betyder, men jeg kan ikke slå det op. Er det et ord I kender?



Vi kender ikke umiddelbart udtrykket. Det findes ikke i nogen af de almensproglige ordbøger vi har adgang til, og vi har heller ikke registreret det i vores ordsamling. Vi har dog fundet eksempler på at udtrykket er blevet brugt, bl.a. i dette citat fra Politiken 9.7.2008: "Koncerterne afvikles i et lokale, der må være enhver lydmands værste mareridt. Stort, forskellige loftshøjder og et hav af krinkelkroge, hvor lyden kan perkle rundt, helt ude af kontrol". Et andet eksempel er fra marineforeningen.dk, hvor der står "Nu kan bestyrelsen jo ikke bruge al sin tid på at perkle rundt og spørge gud og hvermand om vedkommende har været i marinen ...". (nr. 3 1990).

Udtrykket findes altså, men det er tilsyneladende ikke særligt udbredt. Måske er der en sammenhæng med forskellige dialektale udtryk som fx *pikle af*, der dog er et jysk udtryk for at skrubbe af, hvilket ikke rigtig giver mening i forhold til citatet om "børn der perkede rundt i parken på alle fire". Andre dialektale udtryk er *perre (rundt)*, der vist primært er fynsk, og *pøle (rundt)*, der er fra nordjysk. Disse udtryk bruges om at rende lidt forvirret rundt uden mål og



med, hvilket bedre giver mening i forhold til såvel Hustvedts roman som citaterne fra Politiken og Danmarks Marineforening. Dialektforskerne fra Jysk Ordbog kender dog ikke udtrykket, så det er vanskeligt at pege på hvor det kommer fra.

Vores bedste bud er at *perkle rundt* er et dialektalt udtryk for at rende lidt forvirret rundt, måske i samme boldgade som *perre rundt* og *pøle rundt* – uden at vi kan komme oprindelsen og den præcise betydning nærmere.

MHA

Dansk Sprognævn

- Ansvarshavende redaktør: Thomas Hestbæk Andersen
- Redaktion: Margrethe Heidemann Andersen, Kirsten Lindø Dolberg-Møller og Johanne Niclasen Mortensen

- ISSN: 2446-3124
- Forsidefoto: © Kirsten Lindø Dolberg-Møller
- Layout: Pernille Kleinert
- DTP: René Malchow

Usignerede artikler og artikler med initialer giver udtryk for Sprognævnets mening. Artikler med navn giver ikke nødvendigvis i enhver henseende udtryk for Sprognævnets mening.

ASH	Anna Sofie Hartling
ESJ	Eva Skafte Jensen
JNM	Johanne Niclasen Mortensen
JS	Jørgen Schack
KLHK	Kasper Lind-Hansen Kristjansson
KLD	Kirsten Lindø Dolberg-Møller
MHA	Margrethe Heidemann Andersen
MR	Marianne Rathje
ORJ	Oliver Rix Johannsen
PJH	Peter Juel Henriksen
THA	Thomas Hestbæk Andersen
TW	Thomas Widmann
WIC	Winnie Collin

Nyt fra Sprognævnet udgives af Dansk Sprognævn. Tidsskriftet udkommer i udgangspunktet 3 gange om året og er gratis. Du tilmelder dig *Nyt fra Sprognævnet* ved at indtaste din e-mail i formularen på dsn.dk/nyt-fra-sproгнаevnet. Du kan altid afmelde dig igen via et individuelt link som sendes med hvert nyhedsbrev.

Følg også Dansk Sprognævn på vores hjemmeside, på LinkedIn og på Facebook:

- Hjemmesiden: dsn.dk
- LinkedIn: [Dansk Sprognævn](#)
- Facebook: [Dansk Sprognævn](#)

Dansk
Sprognævn

Adelgade 119 B
5400 Bogen