

Hur används *de*, *dem* och *dom* i nutida skriftspråk?

En storskalig korpusundersökning
av nyheter och sociala medier

Av EVIE COUSSÉ, YVONNE ADESAM, FATON
REKATHATI & ALEKSANDRS BERDICEVSKIS

Abstract

Coussé, Evie, evie.cousse@gu.se, Associate Professor, Department of Languages and Literatures, University of Gothenburg; *Adesam, Yvonne*, yvonne.adesam@gu.se, Ph.D., Språkbanken Text, Department of Swedish, Multilingualism, Language technology, University of Gothenburg; *Rekathi, Faton*, faton.rekathi@kb.se, M.Sc., KBLab, the National Library of Sweden; *Berdicevskis, Aleksandrs*, aleksandrs.berdicevskis@gu.se, Ph.D., Språkbanken Text, Department of Swedish, Multilingualism, Language technology, University of Gothenburg. “How are *de*, *dem*, and *dom* used in contemporary written Swedish? A large-scale corpus study of news and social media”. *Språk och stil* NF 33, 2023, pp. 39–70.

This study ties in with a longstanding debate on the Swedish spelling variants *de*, *dem* and *dom* for personal pronouns (third person plural) and definite articles (plural). It charts the usage of *de*, *dem* and *dom* in five large corpora with news and social media texts over the past 25 years. The corpora contain more than 1.5 billion tokens, which rules out manual handling of the data. Instead, this study makes use of computational methods (including an AI language model) to automatically identify and classify relevant observations. Analysis of the news corpora shows a relatively stable usage of *de*, *dem* and *dom* over the past 25 years. The forms *de* and *dem* are predominantly used according to the norm: *de* for pronouns in subject position and as a definite article; *dem* for pronouns in object position. The colloquial form *dom* is hardly found in news texts. Analysis of the social media corpora shows more variation and change. The colloquial form *dom* is used in 5–25% of all instances instead of *de* or *dem* and has decreased after an initial rise. The forms *de* and *dem* are sometimes used in a non-standard way: *de* occurs in object position in 4–10% of the observations; *dem* is found in subject position or as a definite article in 1–7% of the cases. Non-standard *dem* is potentially on the rise with younger writers. The corpus analysis also provides details on the usage of *de* and *dem* in relative clauses, and on the users’ ratings of posts containing *de*, *dem* and *dom* on the social media platform Reddit.

Keywords: Swedish, large-scale corpus study, social media, personal pronoun, definite article, norm, spelling.

1 Introduktion

Den långlivade debatten om stavningsvarianterna *de*, *dem* och *dom* blossade upp igen under 2022–2023. Sara Lövestam (2022) startade debatten i Svenska Dagbladet med följande spaning:

Allt fler har tappat distinktionen mellan objektsform och subjektsform för personliga pronomen i tredje person plural. Låt mig därför komma med en spådom: 'de' och 'dem' ersätts på sikt av ett enda 'dom' även i skriftspråket.

Hon säger i samma artikel att vi har kommit till en »tippningspunkt» och hänvisar till 2021 års nyordslista där konceptet definieras som en »punkt där en förändring i ett system blir så stor att ett jämviktsläge oåterkalleligt och abrupt övergår i ett annat läge» (Språktidningen & Språkrådet 2021). Lena Lind Palicki (2022) vidareutvecklar denna observation i Expressen:

Vi står på krönet precis vid tippningspunkten. Snart kommer en ny språklig tid där distinktionen mellan *de* och *dem* börjar försvinna även ur skrift (även om vi kommer att se parallella system under en övergångstid).

Båda debattörerna skissar en bild av att bruket av *de*, *dem* och *dom* är i gungning i nutida svenska. Men vad är det som händer i språket? I denna artikel vill vi undersöka om man kan se en förändring i den faktiska användningen av *de*, *dem* och *dom* i nutida skriftspråk.

Tidigare forskning har kartlagt användningen av *de*, *dem* och *dom* i nyhetstexter från 1965 till 2008 (Ewerth 1978, 1992, 2008, Håkansson & Norrby 2003) och i elevuppsatser från slutet på 1970-talet till 2015 (Hallencreutz 1980, Lind Palicki & Karlsson 2019). Vi kompletterar den befintliga forskningen genom att undersöka bruket av *de*, *dem* och *dom* i flera stora korpusar som tillsammans omfattar de senaste tjugofem åren (1998–2023). Detta tidsfönster låter oss utvärdera om användningen av *de*, *dem* och *dom* har förändrats under de senaste åren. Vi tar ett mikrodiakront perspektiv där vi följer användningen av *de*, *dem* och *dom* mycket detaljerat över tid, närmare bestämt år för år. Detta är möjligt tack vare en stor mängd tillgängliga texter som är publicerade kontinuerligt över tid och som är försedda med information om exakt publiceringstidpunkt. Korpusmaterialet är så stort att vi inte kan hantera det manuellt. I stället använder vi oss av datamaskinella metoder som automatiskt identifierar och klassificerar relevanta träffar i korpusarna.

Några av de korpusar vi använder innehåller nyhetstexter (GP, SVT Nyheter) som låter oss knyta an till tidigare forskning på tidningsmaterial och komplettera den med nyare datapunkter. Vi använder oss även av korpusar med sociala medier såsom bloggar (Bloggmix) och diskussionsforum (Familjeliv, Reddit). Sociala medier är en relativt ny genre, för vilken användningen av *de*, *dem* och *dom* inte har kartlagts än. Texter från sociala medier är i mindre utsträckning redigerade än nyhetstexter, och skrivs inte bara av professionella skribenter. De är därför en intressant källa för att studera normbrytande språkbrytningsförändring. Vissa av våra korpusar med sociala medier är också försedda med metainformation som öppnar för nya forskningsperspektiv.

Vi använder information från Familjeliv om användarnas födelseår för att undersöka om skribenter från olika generationer skiljer sig åt i bruket av *de*, *dem* och *dom*. Här är vi särskilt intresserade av skillnader som pekar mot att yngre generationer i högre utsträckning använder sig av nya varianter än äldre generationer. Detta skulle kunna vara ett tecken på vad man inom sociolingvistik kallar för en förändring i skenbar tid (*apparent time*). På så sätt inkluderar vi språkförändring både i verklig tid och skenbar tid i vår studie. Reddit förser oss med information om i vilken utsträckning användare gillar eller ogillar ett inlägg (uttryckt i så kallade »karmapoäng»). Vi utforskar här om den informationen kan ge oss en fingervisning om hur användare värderar normbrytande användning av *de* och *dem* i sociala medier. På så sätt anknyter vi till forskning om attityder kring *de*, *dem* och *dom*.

Den här studien är först och främst deskriptiv. Vårt huvudmål är att kartlägga vad som sker i nutida svenskt skriftspråk. Vi försöker inte att uttömmande förklara de observerade trenderna (även om vi ibland nämner vissa förklarande hypoteser och kort diskuterar möjliga tolkningar av våra data). Vi försöker inte heller att förutsäga vad som kommer att ske i framtiden; vi utvärderar alltså inte Lövestams och Lind Palickis spådomar. Att förutsäga språkliga förändringar är ett spännande men mycket svårt uppdrag (Berdicevskis m.fl. 2023). Att avgöra huruvida vi står vid en tippningspunkt är kanske ännu svårare, eftersom begreppet *tippningspunkt*, oss veterligen, inte är operationaliserat i språkvetenskaplig forskning. Det är knappast möjligt att identifiera en konkret tröskel (t.ex. 10 % eller 23,783 %) vid vilken systemet oåterkalleligt övergår i ett annat läge: det är ett känt faktum att språkliga förändringar kan »svänga tillbaka» (se t.ex. översikten över »avbrutna» S-kurvor i Blythe & Croft 2012 eller »U-kurvor» i Nevalainen 2015). För att identifiera en möjlig tippningspunkt skulle man behöva analysera ett brett urval av olika språkförändringar, helst med hjälp av en matematisk modell (se t.ex. Phillips 2007),

gärna baserad på en konkret sociolingvistisk teori. Detta ligger långt bortom syftet med denna artikel.

Artikeln är strukturerad enligt följande. I avsnitt 2 presenterar vi våra forskningsfrågor och går igenom tidigare forskning. I avsnitt 3 beskriver vi våra data och metoder. Därefter undersöker vi den talspråksnära stavningen *dom* i avsnitt 4, normbrytande *de* i avsnitt 5, normbrytande *dem* i avsnitt 6, hur normbrytande *de* och *dem* värderas i avsnitt 7, samt utvecklingen av *de som* i avsnitt 8. I avsnitt 9 avslutar vi med att knyta samman resultaten för alla våra delstudier.

2 Bakgrund och tidigare forskning

I dialog med tidigare forskning utarbetar vi här fem forskningsfrågor som fokuserar på olika aspekter av potentiella förändringar i användningen av *de*, *dem* och *dom*.

2.1 Talspråksnära *dom*

De, *dem* och *dom* har sedan 1970-talet stått i centrum för en intensiv språkdebatt (se Lind Palicki & Karlsson 2020 för en översikt). *De* och *dem* är de former som traditionellt används för pronomen i tredje person plural. *De* står då i subjektsposition (1) medan *dem* används i objektsposition (2). *De* är även formen för bestämd artikel i plural (3). Exempelen är tagna från korpusen SVT Nyheter.

- (1) **De** misstänker att försäljningen sker nära skolor.
- (2) Vi ska träffa **dem** för de vill veta vad vi har gjort.
- (3) Här har vi haft **de** godkända sorterna.

Dom har föreslagits under de senaste decennierna som ett talspråksnära alternativ för både *de* och *dem*. Bakgrunden till förslaget är att *de* och *dem* uttalas som [dɔm] av allt fler svenskar (Cassirer 1976). Exempelen nedan visar hur *dom* används som subjekt (4), objekt (5), och som artikel (6) i korpusen Familjeliv.

- (4) **Dom** har raggat på varandra i smyg.
- (5) Tur mina barn har en far som älskar **dom** över allt annat.
- (6) Någon som har erfarenhet från **dom** länderna?

Vår första forskningsfråga undersöker om *dom* har fått genomslag i skriftspråket.

Forskningsfråga 1: Hur ofta används *dom* i stället för *de* eller *dem*? Ser vi en förändring i verklig och/eller skenbar tid?

Tidigt i debatten om *dom* fanns en föreställning om att formen blir allt vanligare. Molde (1971 s. 15) noterar »Det kan inte ha undgått någon språkinsresserad och tidningsläsande svensk att formen *dom* under de senaste åren har blivit allt vanligare i skrift (både som ersättning för *de* och som ersättning för *dem*)». Detta sporrar Ewerth (1978) att undersöka hur vanligt bruket av *dom* egentligen är i nyhetstexter. Han räknar andelen *dom* (jämfört med *de* och *dem*) i åtta dagstidningar och sex veckotidningar publicerade 1977. Det visar sig att andelen *dom* är låg i de undersökta tidningarna: lägst i dagstidningar (2,4 %) och något högre i veckotidningar (10,3 %). Ewerth (1992, 2008) upprepar senare sin undersökning i samma dagstidningar och konstaterar att andelen *dom* där har sjunkit något under åren (från 2,4 % år 1977 till 1,8 % år 1992 och 0,7 % år 2007).

Håkansson & Norrby (2003) kompletterar den diakrona forskningen genom att jämföra den relativa frekvensen av *dom* i fyra nyhetskorporusar från 1965, 1976, 1987 och 1997. De visar att den relativa frekvensen av *dom* först ökar rejält (från 23 fall per miljon löpord år 1965 till 217 fall per miljon år 1976) och sedan faller tillbaka igen (till 116 *dom* per miljon löpord år 1987 och 73 *dom* per miljon år 1997). Sammantaget visar det sig att *dom* är vanligast i tidningstexter från 1970-talet, och sedan dess har tappat mark.

Användningen av *dom* väcker också intresse i skolans värld. Hallencreutz (1980) undersöker hur vanligt *dom* är i uppsatser skrivna av högstadieelever i slutet av 1970-talet, och ser att *dom* är det vanligaste alternativet av *de*, *dem* och *dom* (57,5 %). Lind Palicki & Karlsson (2019) upprepar denna studie för elevuppsatser skrivna under perioden 1996–2015. Gymnasieelever använder *dom* i ganska låg utsträckning, medan högstadieelever oftare använder formen. Vi kan dock konstatera att *dom* inte längre är den vanligaste formen hos högstadieelever. Även Håkansson & Norrby (2003) rapporterar att *dom* inte är den vanligaste formen hos universitetsstudenter i ett lucktest konstruerat som en dialog mellan expedit och kund. Detta är anmärkningsvärt eftersom *dom* länge har varit accepterat för att återge talspråk.

Sammanfattningsvis ser vi att användningen av *dom* skiljer sig mellan olika typer av tidningstext (dagstidning mot veckotidning) och skolåldrar (elever i högstadiet mot gymnasiet). Både i nyhetstexter och elevtexter har *dom* blivit mindre vanligt från mitten av 1970-talet fram till mitten av 2010-talet. Frågan är om vi ser samma nedåtgående trend i vårt material under de senaste tjugo-

fem åren. Här vill vi påpeka att skriftspråksnormen har blivit något mindre restriktiv för användningen av *dom* under perioden. Medan Svenska språknämnden (2005 s. 211) skriver att *dom* »inte vunnit fullt insteg i skriftspråket utan uppfattas just talspråkligt» öppnar Bylin & Melander (2023 s. 174) upp för att »*dom* kan användas i texter med ett generellt vardagsnära stilläge» oberoende av om *dom* är talspråkhärmande eller inte.

2.2 Normbrytande *de* och *dem*

De och *dem* används inte alltid enligt normen i skriftspråk. Ibland förekommer *de* i stället för *dem* som objekt (7), samt *dem* i stället för *de* som subjekt (8) och bestämd artikel plural (9). Vi kallar dessa förväxlingar för normbrytande användning. Exempelen är hämtade från Bloggmix.

- (7) Man känner varandras dåliga och bra vanor och accepterar **de**.
- (8) Mina föräldrar heter Eva och Stefan och **dem** är världens bästa föräldrar.
- (9) Måste ju faktiskt lyfta fram **dem** fina kommentarerna och mailen jag får!

Vår andra och tredje forskningsfråga fokuserar på att undersöka normbrytande användning av *de* och *dem*.

Forskningsfråga 2: Hur ofta används normbrytande *de* i stället för *dem*? Ser vi en förändring i verklig och/eller skenbar tid?

Forskningsfråga 3: Hur ofta används normbrytande *dem* i stället för *de*? Ser vi en förändring i verklig och/eller skenbar tid?

Osäkerheten i valet mellan *de* och *dem* har fått mycket uppmärksamhet i den förnyade *dom*-debatten 2022–2023. Som framgår av Lövestams citat i introduktionen finns det en föreställning bland språkdebattören och en bredare allmänhet om att allt fler har tappat distinktionen mellan *de* och *dem*.

Hallencreutz (1980) och Lind Palicki & Karlsson (2019) har undersökt hur bra skolelever klarar av att välja mellan *de* och *dem* enligt normen när de skriver uppsatser. Hallencreutz (1980) hittar bara några få fall av *de* använt som objekt (5 %). Högstadiееleverna i hennes material »kan – om de tvingas välja mellan *de* och *dem* – träffa ett korrekt val» (Hallencreutz 1980 s. 101). Lind Palicki & Karlsson (2019) rapporterar mer formvariation i sitt uppsatsmaterial. De hittar inte bara några fall av *de* som objekt, utan också några fall där *dem* används som subjekt eller bestämd artikel plural. I deras material dyker

dem först upp som bestämd artikel plural i högstadiuppsatser från 2008 och 2013. Normbrytande *de* och *dem* är relativt ovanligt (0–13 %) och förekommer främst hos högstadieeleverna. Lind Palicki & Karlsson (2019 s. 141) kommer till slutsatsen »Så – kan dagens elever verkligen inte skilja på *de* och *dem*? Jo, många. De flesta faktiskt. Men inte alla, och många gör det inte perfekt».

Håkansson & Norrby (2003) ger oss insikt i hur väl universitetsstudenter upprätthåller distinktionen mellan *de* och *dem*. Majoriteten av studenterna väljer rätt form när de fyller i ett lucktest som beskriver en vardaglig servicesituation. Håkansson & Norrby (2003) lägger dock märke till en förvånande stor andel av normbrytande *dem* i subjektsposition (10–12 %). Närmare analys visar att normbrytande *dem* endast förekommer i positionen efter det finita verbet. Detta gäller både när *dem* står som subjekt (*Vad kostar dem?*) och som bestämd artikel (*Vad tycks om dem här?*). Håkansson & Norrby (2003 s. 14) tolkar sina fynd som att »studenterna uppfattar positionen efter det finita verbet som en klar objektsposition och därför skriver *dem*». Intressant nog förekommer även *dom* enbart i positionen efter det finita verbet i deras material.

Sammanfattningsvis visar den befintliga forskningen att normbrytande användning av *de* och *dem* har ökat något hos skolelever under de senaste decennierna. Bylin & Melander (2023 s. 170) anger dock att »Normbrytande *dem* för subjekt verkar ha ökat sedan 1980-talet, medan normbrytande *de* för objekt verkar ha minskat, så resultaten pekar inte heller i en enda riktning». Frågan är hur normbrytande användning av *de* och *dem* ser ut i det vuxna skriftspråket. Bylin & Melander (2023 s. 170) anger att normbrytande *de* och *dem* »ligger långt under en procent i tidningstext» men ger inte några fler detaljer.

De tre första forskningsfrågorna utgör basen för att diskutera användningen av *de*, *dem* och *dom* i nutida skriftspråk. En möjlig följdfråga är hur man ser på normbrytande *de* och *dem* och om attityden har förändrats över tid. Frågan har spelat en betydande roll i den nuvarande språkdebatten. Lövestam (2022) hänvisar till två enkätstudier som kartlägger vilka språkriktighetsfrågor som upplevs som besvärliga av svensklärare. Grünbaum (1976) lyfter användningen av talspråksnära *dom* i stället för *de* och *dem* som problematisk. Karlsson & Lind Palicki (2017) visar att lärarnas fokus har flyttats från *dom*-frågan till valet mellan *de* och *dem*. Inte mindre än 63 % av alla respondenter uppger att skillnaden mellan *de* och *dem* är besvärlig att hantera i klassrummet.

Forskningsfråga 4: Hur värderas normbrytande *de* och *dem*? Ser vi en förändring av värderingar över tid?

Utan intervjuer eller liknande är det svårt att säga något om attityder. Som ett sidospår utforskar vi detta med hjälp av s.k. karmapoäng på diskussionsforumet Reddit, ett indirekt sätt att mäta synen på normbrott. Även om eventuella skillnader i karmapoäng inte behöver vara kopplade till användningen av *de* och *dem* ger detta en intressant nyans till diskussionen.

2.3 Metodologiska utmaningar gällande normbrytande *de* och *dem*

Frågan om *de* och *dem* används normenligt kompliceras när formerna följs av en relativsats. Traditionellt väljer man formen efter sin funktion i huvudsatsen: *de som* när *de* är subjekt (10) och *dem som* när *dem* är objekt (11). Exempelen är hämtade från GP.

- (10) Men det finns också **de som** är positiva.
- (11) Hon tillhör **dem som** får femmånadersvikariat, som hon uttrycker det själv.

Svenska språknämnden (2005 s. 119–120) påpekar att *de som* har blivit allt vanligare i fall som (12) och (13).

- (12) De tillhör oftast inte **de som** är högavlönade.
- (13) Men Bengt Brölde tillägger att för **de som** har ett arbete är arbetets innehåll viktigare än dess lön.

Enligt traditionen förväntar man sig *dem* eftersom formen är objekt i huvudsatsen. En möjlig förklaring till varför *de* används här är att det är subjekt i relativsatsen i både (12) och (13). En annan förklaring är att *de* egentligen är en bestämd artikel i plural där substantivet utelämnats. I både (12) och (13) kan man tänka sig att *de* följs av det generiska substantivet *människor*. Varianten *de som* väcker stor debatt under 1970-talet (se Lind Palicki & Karlsson 2020 för en översikt). Molde (1980) sätter punkt för debatten genom att visa att det finns grammatiska analyser som stödjer både *dem som* och *de som*. Båda alternativen borde alltså anses vara korrekta. Den rekommendationen gäller än idag (Bylin & Melander 2023 s. 174).

Vår femte forskningsfråga fokuserar på användningen av *de som* gentemot *dem som* efter preposition. Här undersöker vi enbart förändringar i verklig tid.

Forskningsfråga 5: Hur ofta används *de som* i stället för *dem som* efter preposition? Ser vi en förändring över tid?¹

Frågan diskuterades främst när *de som* fortfarande ansågs vara normbrytande. Platzack (1976) undersöker med hjälp av en enkätundersökning i vilken utsträckning skribenter med hög språkkompetens lyckas välja rätt mellan *de som* och *dem som*. Han observerar att försökspersoner träffar fel i ungefär en fjärdedel av testmeningarna (26 %), men specificerar inte när *de som* används i stället för *dem som* och vice versa. Hallencreutz (1980) rapporterar att högstadieelever i ett lucktest ofta väljer *de som* efter prepositionen *av* i stället för *dem som*, vilket hon räknar som fel. Det framgår inte av hennes resultat exakt hur stor andelen *de som* är. Ewerth (1978 s. 28) undersöker användningen av *de/dem som* efter preposition eller som objekt i åtta dagstidningar och sex veckotidningar från 1977. Han konstaterar att andelen *de som* gentemot *dem som* bara är 11 % av det totala antalet belägg. Ewerths uppföljningsstudie från 1992 (Ewerth 1992) ger oss en bild av utvecklingen efter att *de som* blev accepterad som alternativ stavning för *dem som*. Hans undersökning av åtta dagstidningar från 1992 visar att andelen *de som* har ökat (från 11 % år 1977 till 32 % år 1992). Språkdebatten och språkforskningen har sedan dess inte ägnat frågan någon större uppmärksamhet. Vi vill dock undersöka hur *de som* förhåller sig till *dem som* i nutida skriftspråk.

Ytterligare en komplikation är att formvarianten *de* inte bara används normbrytande för *dem* utan ibland också för *det*. Exempelen nedan från Familjeliv illustrerar att *de* kan stå för *det* som personligt pronomen singular (14) och för *det* som bestämd artikel singular neutrum (15).

- (14) Får se vad framtiden visar, kollar runt för att se vad de finns för möjligheter och hitta något som passar.
- (15) Har försökt att kolla vem som äger de huset.

Normbrytande *de* i stället för *det* introduceras i den nuvarande språkdebatten av svenskläraren Birkebo (2016), som hos sina elever observerar att det numera är »vanligt att se 'de lilla huset' och 'de är komiskt' i skrift». Stavningen *de* för *det* kan ses som en form av uttalsnära stavning, eftersom *det* uttalas [de:]. Lind

¹ Vi kan inte undersöka förändring i skenbar tid för *de som/dem som* på grund av databrist.

Palicki & Karlsson 2019 är den enda forskningsstudien som undersöker den typen av normbrytande *de*. De hittar flera fall av *de* där det borde stå *det* i uppsatser skrivna av både högstadie- och gymnasieelever under perioden 1996–2015, men dessa fall utgör bara en bråkdel av alla observationer av *det* (0–4 %). Normbrytande användning av *de* i stället för *det* är en separat forskningsfråga som vi inte undersöker här. Vi är dock noga med att utesluta normbrytande *de* för *det* när vi analyserar normbrytande *de* för *dem* (se avsnitt 3.2 för detaljer).

3 Korpus och metod

Vi angriper våra forskningsfrågor med hjälp av korpuslingvistiska metoder. Här presenterar vi vårt korpusmaterial (3.1), de datamaskinella metoder vi har använt för att identifiera relevanta träffar i våra korpusar (3.2) och hur vi visualiserar resultaten i grafer (3.3).

3.1 Korpusmaterial

Vi använder oss av stora korpusar med sociala medier och nyhetstexter från de senaste tjugofem åren för att undersöka användningen av *de*, *dem* och *dom*. Tabell 1 ger en översikt över texttyperna. Nyhetstexterna har sammanlagt 469 miljoner token (löpord, vilket även inkluderar t.ex. skiljetecken). Materialen från sociala medier består av Bloggmix och diskussionsforumet Familjeliv, som tillsammans utgör nästan en miljard token. Diskussionsforumet Reddit består av ett begränsat urval och innehåller enbart inlägg som innehåller *de* eller *dem*. Vi vet inte hur stort materialet är totalt, men urvalet består av 82 miljoner token.

Tabell 1. Korpusar. Reddit-korpusen består enbart av inlägg som innehåller *de/dem* medan övriga korpusar innehåller kompletta texter.

Korpus	Textgenre	Period	Antal miljoner token
Göteborgs-Posten (GP)	Nyheter	2001–2013	250
Sveriges Television (SVT) Nyheter	Nyheter	2004–2021	219
Bloggmix	Blogg	1998–2017	581
Familjeliv-generationer	Diskussionsforum	2003–2021	396
Reddit (r/sweden)	Diskussionsforum	2010–2023	(82)

Tillsammans representerar materialen två olika typer av nutida svenskt skriftspråk. Nyhetstext innehåller ett relativt formellt språk, som i princip följer normerna för det svenska skriftspråket. Genren kan alltså fungera som ett språkligt standardiserat referensmaterial. Nyhetskorpuserna låter oss också jämföra våra resultat med tidigare forskning. Sociala medier utmärker sig genom att de innehåller en mer oredigerad, informell kommunikation. Vi förväntar oss att normbrytande språkförändringar syns tidigare i denna typ av material än i nyhetstexterna, som i betydligt större utsträckning är redigerade och kan antas vara mer språkligt konservativa. Gemensamt för både sociala medier och nyhetstexter är att de har skapats i digital form och alltså inte har digitaliserats i efterhand.

Det svenska diskussionsforumet Familjeliv innehåller diskussioner inom många olika ämnen, men störst fokus ligger på familjerelaterade ämnen, som graviditet och föräldraskap. Vi undersöker enbart inlägg skrivna av de användare som har skrivit minst 10 000 token, som har angett ålder i sin profil, samt som enligt profilen var minst 18 år gamla när de skrev sitt första inlägg. Korpusen heter »Familjeliv-generationer», men kallas framöver »Familjeliv» utom när vi beskriver hur korpusen skapats, då vi behöver skilja mellan de två materialen.

Korpusen Bloggmix innehåller ett urval av bloggar baserat på topplistorna på webbplatsen bloggportalen.se (mest besökta privata bloggar, mest besökta proffsbloggar och de lokala topplistorna för olika regioner).

Det internationella diskussionsforumet Reddit har ett svenskt delforum (subreddit r/sweden) där ett brett spektrum av ämnen diskuteras. På Reddit kan användare rösta upp eller ned andra användares inlägg. Dessa röster ger upphov till »karmapoäng», som indikerar hur populärt ett inlägg är, och som kan påverka inläggets synlighet i tråden genom att användare ges möjlighet att sortera inlägg efter popularitet eller kontroversialitet.

GP, SVT, Bloggmix och Familjeliv underhålls av Språkbanken Text och är tillgängliga via verktyget Korp (Borin m.fl. 2012). Vi skapade Familjeliv-generationer genom att berika Språkbankens version med information om skribenternas ålder. Reddit skiljer sig från de övriga korpuserna på två sätt. Dels finns materialet inte bland Språkbankens korpuser, utan skapades av oss genom att vi laddade ner forumet direkt från webben. Dels skapades korpusen genom en riktad insamling av samtliga inlägg som innehåller *de* och/eller *dem* i delforumet r/sweden. Dessa utgör 82 miljoner token. Vi använder alltså inte alla inlägg i hela r/sweden.

Sammantaget representerar de utvalda korpuserna stora mängder skriftspråk från de senaste tjugofem åren. Storleken på vårt material medför att vi

inte kan hantera det manuellt. Det påverkar inte bara vår sökmetod (se avsnitt 3.2) utan även vår möjlighet att utvärdera och kontrollera sammanställningen på materialet. Här är det viktigt att komma ihåg att våra stora korpusar inte har någon balanserad sammanställning utifrån fördefinierade kriterier (såsom skribentens ålder, kön, osv.) utan ofta innehåller allt material som fanns att tillgå. På så sätt kan vårt material, och därmed våra resultat, vara spretigare än vad som är vanligt i mindre korpusstudier. Samtidigt utgör detta mer »vildvuxna» material den språkliga verkligheten som språkanvändare möter när de läser nyhetstexter eller befinner sig på sociala medier.

3.2 Sökmetod

För att besvara våra forskningsfrågor använder vi oss av datamaskinella metoder som automatiskt identifierar och klassificerar relevanta träffar i korpusarna. Vi använder oss i första hand av sökverktyget Korp. Där söker vi efter relevanta träffar med hjälp av den morfosyntaktiska uppmärkning som har gjorts automatiskt av annoteringsverktyget Sparv (Hammarstedt m.fl. 2022).² Den annotering som är relevant för denna studie består av uppmärkning av ordklasser, vilket också omfattar morfologisk annotering,³ och syntaktisk annotering, i form av dependensstrukturer,⁴ som pekar ut relationer mellan ord (snarare än frasstruktur, som pekar ut relationer mellan grupper av ord). Vi söker inte via webbgränssnittet utan via Korps API (*application programming interface*).⁵ Detta gränssnitt passar bättre för storskaliga studier eftersom det kan anropas från ett program för att utföra fördefinierade sökningar, ladda ner sökträffar vid behov, analysera dem automatiskt och sammanfatta resultaten i en graf eller en tabell. För att söka i Familjeliv-generationer använder vi ett skräddarsytt verktyg som vi specifikt utvecklade för den här studien. Våra sökningar i Korp har en hög träffsäkerhet när det gäller att hitta alla relevanta fall av *de*, *dem* och *dom*, men kan inte tillförlitligt identifiera normbrytande *de* och *dem* (se appendix A för en utvärdering).

Som komplement till Korp använder vi oss därför av en AI-språkmodell som är tränad för att automatiskt identifiera normbrytande *de* och *dem*. Språkmodellen (Rekathati 2023) är en finjusterad version av språkmodellen KB-BERT

² <https://spraakbanken.gu.se/sparv/#/user-manual/available-analyses>

³ <https://spraakbanken.gu.se/korp/markup/msdtags.html>

⁴ <https://svn.spraakdata.gu.se/sb-arkiv/pub/mamba.html>

⁵ <https://ws.spraakbanken.gu.se/docs/korp>

(Malmsten m.fl. 2020), som vi har lärt att klassificera ord utifrån en uppsättning kategorier: *de*, *dem*, *det* och *övrigt* (samt *enda* och *ända*, som inte används för denna artikel).⁶ När vi säger att modellen klassificerar ett ord som exempelvis *de* menas att modellen bedömer att ordet borde vara *de* – baserat på den träningsdata modellen tidigare fått se – även om det är stavat som t.ex. *dem*. Modellen har tränats på textkällor som representerar formellt språkbruk och följer skriftspråkets normer. Vi kan alltså säga att modellens klassificering återspeglar vad normen föreskriver. Vi använder modellens kategorier *de* och *dem* för att avgöra om *de* och *dem* har använts normenligt eller inte. Modellens kategori *det* används för att filtrera bort fall av *de* där det borde ha stått *det*. *Övrigt* är en bakgrundskategori för samtliga ord och tecken som inte har fått sin egen kategori. Modellen visar sig ha en hög träffsäkerhet när den ska bedöma om *de* eller *dem* har använts enligt normen (se utvärderingen i appendix A).

Tabell 2 sammanfattar vilken metod och vilka korpusar som använts för att besvara våra fem forskningsfrågor.

Tabell 2. Metod och korpusar per forskningsfråga.

Fråga	Fokus på	Metod	Korpusar
1	dom	Korp	GP, SVT, Bloggmix, Familjeliv
2	de (i stället för dem)	Språkmodell	GP, SVT, Bloggmix, Familjeliv, Reddit
3	dem (i stället för de)	Språkmodell	GP, SVT, Bloggmix, Familjeliv, Reddit
4	värdering normbrott	Språkmodell	Reddit
5	de som	Korp + Språkmodell	GP, SVT, Bloggmix, Familjeliv

För **forskningsfråga 1** söker vi i Korp efter alla ord stavade som *de*, *dem* eller *dom* där ordklass är pronomen eller determinerare. För att utesluta fall där *dom* egentligen är ett substantiv, men felaktigt uppmärkt som pronomen eller determinerare, lägger vi till ett villkor. Om ordet stavas som *dom*, får det föregående ordet inte ha dependensrelation »framförställt attribut» eller »determinerare, bestämningsord».⁷ Det utesluter fall som *gammal dom* eller *tingsrättens dom* på ett mycket effektivt sätt.

⁶ Modellen är fritt tillgänglig via <https://huggingface.co/Lauler/deformer>. Den intresserade läsaren kan själv testa modellen under rubriken »Hosted Inference API» genom att skriva in en text i rutan och trycka på »Compute».

⁷ I CQP (*Corpus Query Processor*, det formella språket som används för att söka i Korp) ser sökningen ut så här: [deprel!= 'DT' & deprel!= 'AT'] [word = 'dom' %c & (pos = 'PN' | pos = 'DT')].

För **forskningsfrågorna 2–4** bedömer vi om *de* och *dem* har använts enligt normen eller inte med hjälp av språkmodellen. Språkmodellen får som input alla meningar som innehåller *de* eller *dem* (Reddit) eller alla meningar som innehåller *de* eller *dem* uppmärkta som pronomen eller determinerare (alla andra korpusar). Vi sällar bort alla meningar där *de* eller *dem* följs av *som*, eftersom normen här ofta tillåter båda formerna.

För **forskningsfråga 2** använder vi alla fall av *de* och *dem* där modellen bedömer att det borde stå *dem* enligt normen. Med andra ord använder vi oss av alla fall där *de* och *dem* använts som objekt.

För **forskningsfråga 3** tar vi samtliga fall av *de* och *dem* som enligt modellen borde vara *de*. Det handlar alltså om alla förekomster av *de* och *dem* som subjekt eller bestämd artikel.

För **forskningsfråga 4** används enbart Reddit, eftersom bara den korpusen har den relevanta informationen (karmapoäng). Vi använder språkmodellen för att identifiera normbrytande *de* (i stället för *dem*), normbrytande *dem* (i stället för *de*) och samtliga fall av *de* och *dem* som används enligt normen. Sedan beräknar vi för dessa tre grupper inläggens genomsnittliga karmapoäng.

För **forskningsfråga 5** använder vi både Korp och språkmodellen. Först söker vi i Korp efter alla sekvenser som motsvarar följande mall: preposition + *de* eller *dem* uppmärkt som pronomen eller determinerare + *som*. Sedan använder vi språkmodellen för att filtrera bort alla observationer där *de som* egentligen står för *det som*, eftersom de inte är relevanta för den här forskningsfrågan.

3.3 Visualisering

Vi presenterar våra resultat som förändringar i relativa frekvenser för konkurrerande varianter över tid (per år). Med *relativ frekvens* menar vi frekvensen av den nya varianten (eller varianten i fokus) i förhållande till frekvensen av alla varianterna (den kan variera mellan 0 och 1). Observera att skalan på Y-axlarna varierar mellan de olika graferna. Eftersom de relativa frekvenserna vanligen är låga justerar vi skalan för att synliggöra trenden. Visualiseringar som även innehåller information om skribentens ålder är mer komplexa. Vi delar in skribenterna i Familjeliv i fyra generationer, se tabell 3. Vi representerar de relativa frekvenserna inom varje generation som en separat linje. Beräkning av relativa frekvenser görs för varje generation och årtal.

I samtliga grafer utesluter vi de år där den totala frekvensen av variabeln (alla varianter sammantaget) i den givna korpusen eller generationen är mindre

än eller lika med 500. I generationsgraferna tillämpar vi ytterligare en tröskel och utesluter även de år där antalet skribenter i den givna korpusen eller generationen är mindre än eller lika med 20.

De resultat som presenteras i graferna kan i olika grad påverkas av hur korpusarna är sammanställda, hur vi identifierar relevanta observationer för undersökningen, samt hur väl språkmodellen klassificerar dessa observationer. Vi försöker redogöra för dessa faktorer dels genom att utvärdera språkmodellen (appendix A), dels genom att inkludera en alternativ version av graferna som ger läsaren en bild av antalet observationer för varje korpus och år (appendix B).

Vi väljer att inte visualisera statistiska osäkerhetsmått i form av konfidensintervall eftersom vi för flera korpusar har samlat in hela materialet för de aktuella åren. Konfidensintervall kvantifierar osäkerhet som beror på att ett mindre slumpmässigt urval dragits från en större population. För våra data existerar ingen sådan osäkerhet i de fall där urvalet är komplett. Förekomsten av en språklig variant kan trots detta ibland variera kraftigt från år till år. Sådan variation reflekterar – i de fall urvalet är komplett – vad som faktiskt skrivits i korpusen det aktuella året. Variationen mellan år kan då bero på att färre inlägg skrivits ett visst år, eller att olika skribenter varit aktiva. Vi försöker minimera effekten av detta genom de trösklar vi har satt för antalet inlägg och skribenter.

Vår visualiseringsmetod låter oss visa potentiella förändringar både i verklig tid och skenbar tid. Skenbar tid kan utforskas genom att jämföra olika generationslinjer. Om till exempel en yngre generation A är mer innovativ än en äldre generation B under en viss period, ligger linjen för A högre än linjen för B under den perioden (oavsett linjens trend). Även om det inte finns någon synlig ökning av en innovativ variant i verklig tid kan det alltså fortfarande finnas en ökande trend i skenbar tid, vilket tyder på att en förändring är på gång.

Tabell 3. Generationer i Familjeliv.

Generation	Födelseår	Antal skribenter	Antal miljoner token
1	före 1970	258	25
2	1970–1979	1650	136
3	1980–1989	3582	217
4	från 1990	478	18
Totalt		5968	396

Graferna som är relevanta för forskningsfråga 4 beskrivs mer detaljerat i avsnitt 7.

Vi betonar att våra visualiseringar visar relativa frekvenser på korpusnivå snarare än på skribentnivå. De bör alltså tolkas i bemärkelsen »hur ofta förekommer en språklig variant i respektive korpus», snarare än »hur ofta används en språklig variant av den genomsnittliga skribenten».

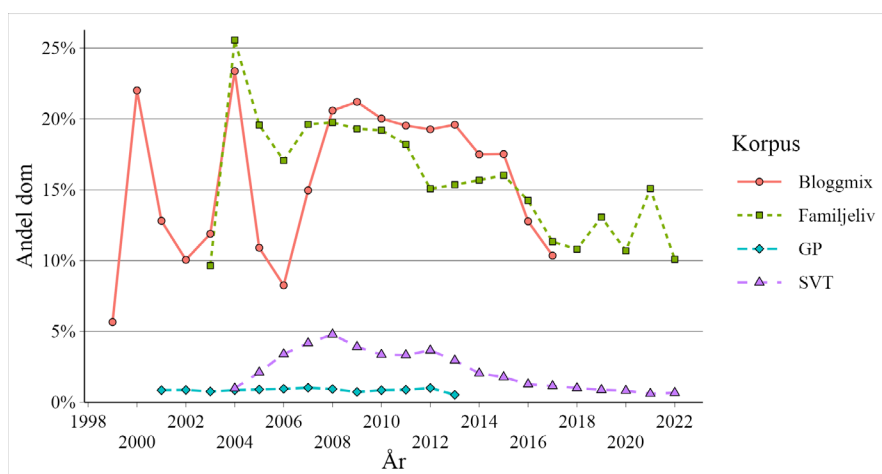
4 Användning av talspråksnära *dom*

Vi inleder med att undersöka **forskningsfråga 1**: Hur ofta används *dom* i stället för *de* eller *dem*? Ser vi en förändring i verklig och/eller skenbar tid? Litteraturöversikten visade att andelen *dom* har sjunkit sedan 1970-talet och skiljer sig något mellan de undersökta genrerna samt åldrarna. I vårt material förväntar vi oss att *dom* används oftare i sociala medier än i tidningstexter. Sociala medier närmar sig i vissa avseenden talat språk (Wiktorsson 2018) vilket kan gynna *dom* som talspråksnära variant. Sociala medier kan dessutom betraktas som »texter med ett generellt vardagsnära stilläge» där *dom* kan användas enligt Bylins & Melanders (2023 s. 174) nya rekommendation. Generellt förväntar vi oss att den nedåtgående trenden för *dom* som framkom i nyhetstexter och elevtexter också syns i vårt material.

4.1 Talspråksnära *dom* i olika korpusar

Figur 1 visar andelen *dom* för fyra av våra fem korpusar. Som förväntat är *dom* vanligare i sociala medier (Bloggmix, Familjeliv) än i nyhetstexter (GP, SVT). Andelen *dom* ligger på bara några procentenheter i våra korpusar med nyhetstexter, i linje med Ewerths (1978, 1992, 2008) tidigare resultat. På sociala medier ligger andelen *dom* mellan 5 och 25 %. De diakrona trendlinjerna i grafen varierar mellan korpusarna och är svåra att relatera till den nedåtgående tendensen som har rapporterats i litteraturen. Här måste man dock komma ihåg att den befintliga forskningen bara jämför några enstaka tidpunkter som ligger långt ifrån varandra i tiden. Vårt mikrodia krona perspektiv fångar förändringar över tid mer detaljerat så att även fluktueringar från år till år syns. Just fluktueringar är något som präglar de första åren för Bloggmix (vilket kan bero på att antalet observationer är mindre, se appendix B), därefter ser vi en nedgång. Familjeliv har en mer stadig utveckling med initialt en ganska kraftig

ökning av andelen *dom*, följt av en långsam tillbakagång. SVT visar samma mönster med en liten ökning i början och sedan en tillbakagång. Gemensamt för de tre korpusarna är en nedgång av andelen *dom* under de senaste 10–15 åren. Det verkar som om *dom* får ett uppsving i början av 2000-talet – samtidigt som internet och sociala medier får sitt breda genombrott – men att varianten därefter sakta överges av skribenterna igen. GP tar inte del i den trenden eftersom *dom* där ligger på samma låga nivå under hela perioden.

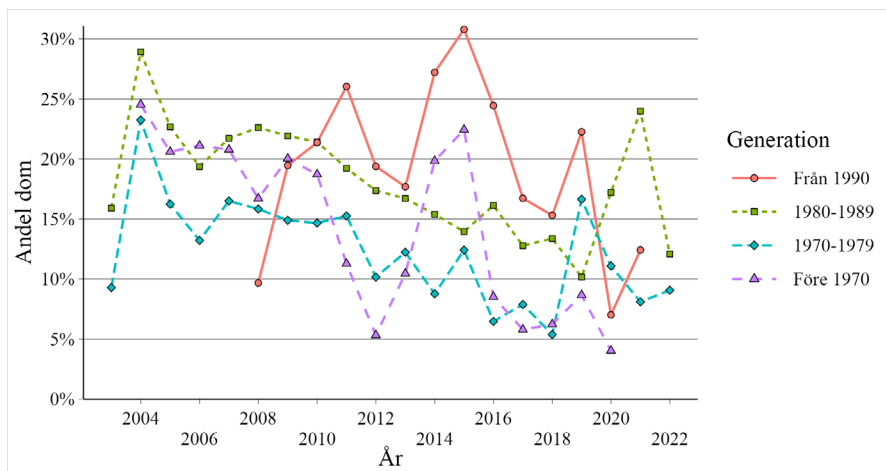


Figur 1. Andel *dom*, av samtliga *de/dem/dom* (per korpus).

4.2 Talspråksnära *dom* i olika generationer

Figur 2 visar hur ofta *dom* används i olika generationer. Skribenter födda på 1970- och 1980-talet följer den allmänna trendlinjen för Familjeliv (som är avbildad i figur 1) ganska väl. Det som skiljer de båda generationerna åt är att de yngre, 1980-talisterna, systematiskt har en högre andel *dom* än 1970-talisterna. Trendlinjerna för skribenterna födda före 1970 och skribenter födda från 1990 visar en mycket mer vacklande användning av *dom* över tid. Detta hänger ihop med att vi har mindre data för dessa generationer (jfr tabell 3 och appendix B). Trots det ser vi att skribenterna födda 1990 och senare överlag använder *dom* mest av alla generationer. Sammanfattningsvis visar alltså yngre skribenter en högre preferens för *dom* än äldre skribenter. Denna trend kan peka på en pågående språkförändring i skenbar tid – *dom* är på väg att bli vanligare bland

de yngre skribenterna – som inte syns i verklig tid. Tendensen kan dock också förklaras på annat sätt, nämligen som åldersgradering: när man är yngre är man mer innovativ, men man blir mer konservativ med åldern. För att ta reda på om det är språkförändring eller åldersgradering kan vi återkomma till verklig tid och titta på vad som sker med varje generation. Här ser vi en nedåtgående trend för alla generationer, vilket är ett tecken på åldersgradering.



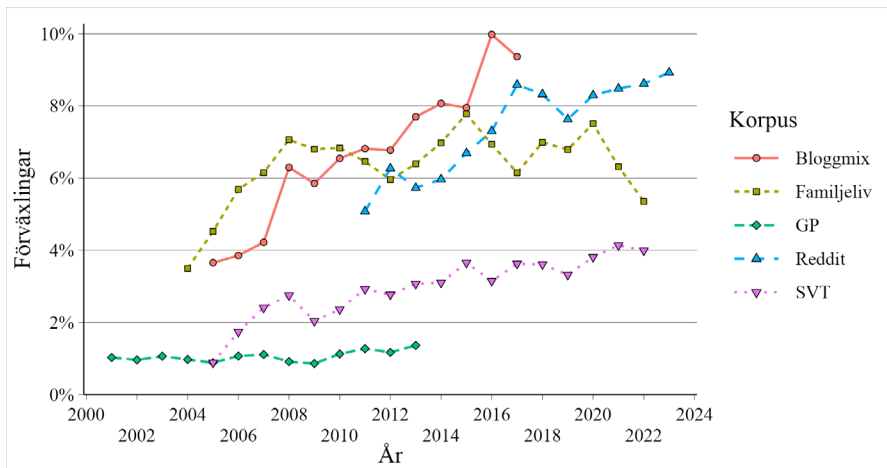
Figur 2. Andel dom, av samtliga de/dem/dom (per generation).

5 Användning av normbrytande *de* (i stället för *dem*)

Här tittar vi närmare på **forskningsfråga 2**: Hur ofta används normbrytande *de* i stället för *dem*? Ser vi en förändring i verklig och/eller skenbar tid? Normbrytande *de* används som objekt där det enligt normen borde stå *dem*. Litteraturoversikten visar att andelen normbrytande *de* är relativt låg i både skoluppsatser och tidningstexter. Vi förväntar oss att andelen normbrytande *de* ligger högre i sociala medier än i tidningstexter. Den bakomliggande tanken är här att inlägg på sociala medier, framför allt på diskussionsforum som Familjeliv eller Reddit, skrivs snabbt och utan någon egentlig redigering. Detta kan leda till att normbrytande *de* står kvar på sociala medier medan dessa former redigeras bort från nyhetstexter. Dessutom är inte alla användare på sociala medier professionella skribenter utan privatpersoner med varierande skrivvana.

5.1 Normbrytande *de* i olika korpusar

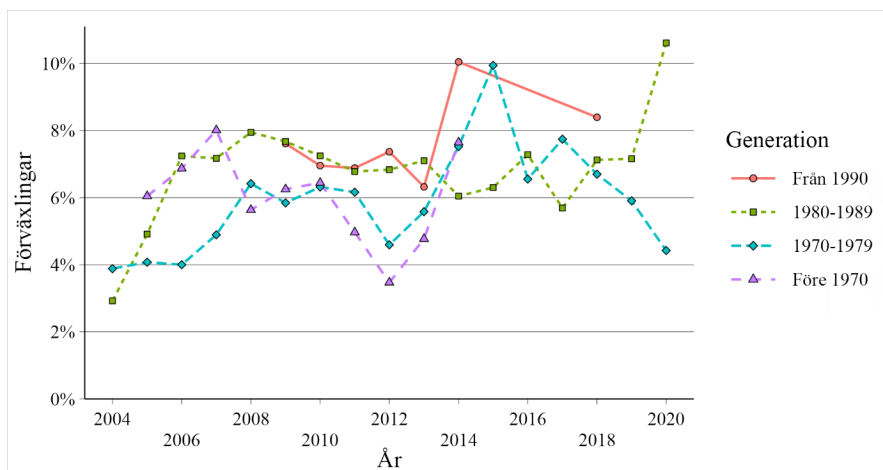
Figur 3 visar att andelen normbrytande *de* är lägre i nyhetstexter än sociala medier. GP ligger stabilt på 1 % vilket är i linje med Bylin & Melander (2023 s. 170) som anger att normbrytande *de* (och *dem*) »ligger långt under en procent i tidningstext». SVT ligger först på samma låga nivå som GP men sedan ökar andelen normbrytande *de* till omkring 4 %. Sociala medier visar på en större osäkerhet i valet mellan *de* och *dem*. Där hittar vi upp till 10 % *de* där det borde ha stått *dem* enligt normen. Över tid ser vi att andelen normbrytande *de* ökar med några procentenheter i flera av våra korpusar. Förändringarna över tid är dock små och ganska ostadiga.



Figur 3. Andel *de*, av samtliga *de/dem* som är objekt (per korpus).

5.2 Normbrytande *de* i olika generationer

Generationslinjerna i figur 4 följer den allmänna trenden för Familjeliv (som är avbildad i Figur 3) mycket nära. Trendlinjerna korsar varandra ofta, vilket betyder att användningen av normbrytande *de* knappt skiljer sig mellan de olika generationerna. Vi ser att skribenter födda på 1980-talet och senare visar ett något större bruk av normbrytande *de*. Det handlar dock om mycket små skillnader som inte är bestående över tid. Vi kan alltså inte dra några slutsatser om förändringar i skenbar tid.



Figur 4. Andel *de*, av samtliga *de/dem* som är objekt (per generation).

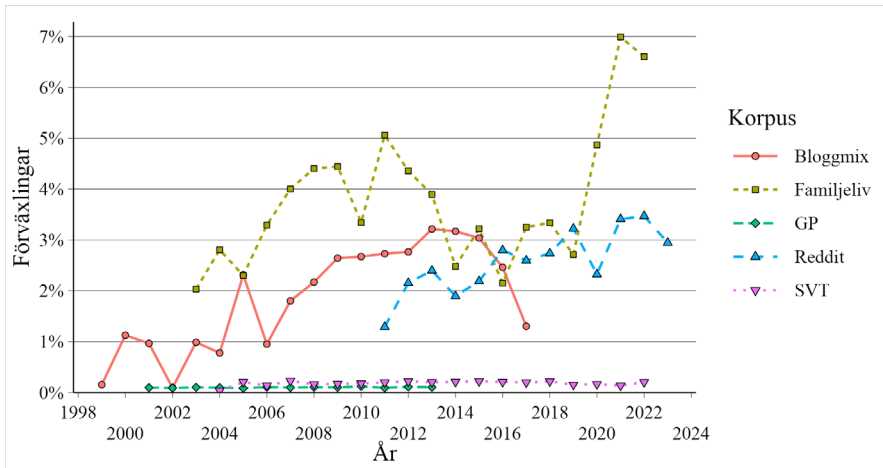
6 Användning av normbrytande *dem* (i stället för *de*)

Nu går vi vidare till **forskningsfråga 3**: Hur ofta används normbrytande *dem* i stället för *de*? Ser vi en förändring i verklig och/eller skenbar tid? Normbrytande *dem* används för subjekt eller bestämd artikel plural där normen föreskriver *de*. Tidigare forskning visar att normbrytande *dem* förekommer relativt sällan i både elevuppsatser och nyhetstexter. Normbrytande *dem* verkar vara yngre än normbrytande *de* i elevuppsatser (Hallencreutz 1980, Lind Palicki & Karlsson 2019): stavningen dyker först upp som subjekt i det undersökta uppsatsmaterialet på 1990-talet och sedan som bestämd artikel i plural på 2000-talet. Man kan tänka sig att normbrytande *dem* haft mindre tid att sprida sig än normbrytande *de*. Vi förväntar oss därför att andelen normbrytande *dem* är lägre än normbrytande *de* i vårt material.

6.1 Normbrytande *dem* i olika korpusar

Figur 5 bekräftar vår hypotes om att normbrytande *dem* är mindre vanligt än normbrytande *de*. Normbrytande *dem* förekommer väldigt sällan i våra korpusar med nyhetstexter. Här ligger användningen i storleksordningen »långt

under en procent» som Bylin & Melander (2023 s. 170) anger för förväxlingar av både *de* och *dem* i nyhetstexter. Vårt material visar att det främst är normbrytande *dem* som används så lite i nyhetstexter. Sociala medier visar en något större andel normbrytande *dem* än nyhetstexter – för upp till 7 % av alla subjekt och bestämda artiklar står det *dem* i stället för *de*. Andelen *dem* i stället för *de* verkar bli lite vanligare över tid i sociala medier, men det handlar om små skillnader.

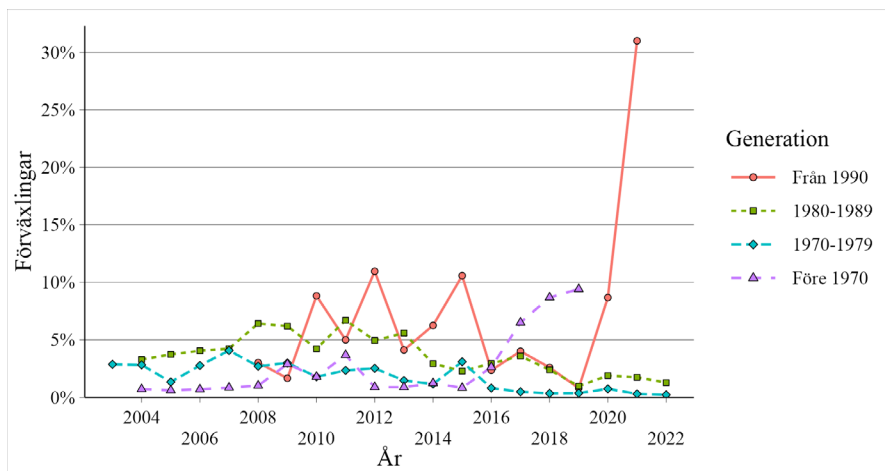


Figur 5. Andel *dem*, av samtliga *de/dem* som är subjekt/artikel (per korpus).

6.2 Normbrytande *dem* i olika generationer

Figur 6 visar att det finns små men systematiska skillnader mellan generationerna. Vi ser närmare bestämt att normbrytande *dem* är vanligare hos yngre skribenter än äldre. Att yngre skribenter har en lite högre preferens för normbrytande *dem* än äldre skribenter kan man tolka som ett tecken på en pågående språkförändring bland de yngre skribenterna. Här ser vi att förändringen i skenbar tid matchar den observerade förändringen i verklig tid. Vi kan dock inte utesluta att åldersgradering även här spelar in (alltså att skribenter är innovativa när de är unga men blir mer konservativa när de blir äldre). När vi tittar på förändringarna i verklig tid ser vi en svagt nedåtgående trend för de två »stora» generationerna (1970- och 1980-talisterna), vilket också kan peka mot åldersgradering. Trenden går dock upp för den äldsta generationen (där vi har

mindre data, se appendix B). Vi vill inte dra några slutsatser om den ovanligt kraftiga ökningen av normbrytande *dem* bland de yngsta talarna 2021 utan att veta vad som skett efter det.



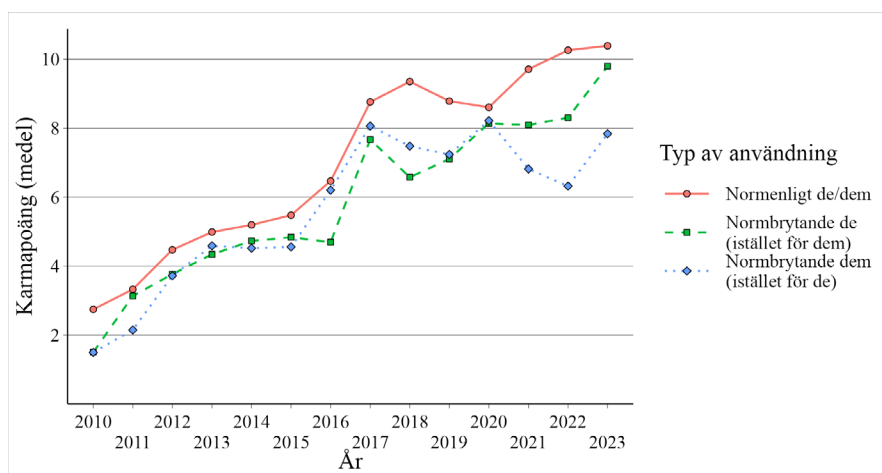
Figur 6. Andel *dem*, av samtliga *de/dem* som är subjekt/artikel (per generation).

7 Värdering av normbrytande *de* och *dem*

För att undersöka olika vinklar byter vi fokus från användning till värdering av normbrytande *de* och *dem*. Valet mellan *de* och *dem* har pekats ut som en besvärlig språkriktighetsfråga av svensklärarna. Utöver en ökning av normbrotten kan detta också visa på en förändrad syn på dessa normbrott. Här undersöker vi värdering av normbrytande *de* (i stället för *dem*) och *dem* (i stället för *de*) med hjälp av karmapoäng på Reddit, för att försöka besvara **forskningsfråga 4**: Hur värderas normbrytande *de* och *dem*? Ser vi en förändring över tid? Karmapoäng utgör givetvis ett indirekt sätt att undersöka värdering av språkliga fenomen och det är svårt att exakt bedöma vilka faktorer som påverkar inläggets karmapoäng. Tidigare forskning har dock visat att språkliga faktorer spelar en roll utöver mer strukturella faktorer, som t.ex. att inlägg i början av en diskussionstråd får fler s.k. uppröster (se Medvedev m.fl. 2019 för en forskningsöversikt). Vår hypotes är att inlägg med normbrytande *de* eller

dem får lägre karmapoäng och därmed värderas lägre än inlägg där *de* och *dem* används enligt normen.

Figur 7 visar att alla inlägg som innehåller *de* eller *dem* får högre karmapoäng över tid. Vi vet inte om ökningen gäller för alla inlägg överlag eller för just vårt urval av inlägg med *de* eller *dem*. Grafen visar att inlägg med normbrytande *de* eller *dem* genomgående får lägre karmapoäng än inlägg där *de* och *dem* används enligt normen. Detta skulle kunna vara en indikation på att normbrytande *de* eller *dem* värderas lägre än normenlig användning. Vi kan dock inte säga något säkert om anledningen och entydigt koppla värderingen till detta specifika normbrott. Det skulle till exempel kunna vara så att normbrytande *de/dem* samförekommer med fler normbrytande språkföreteelser. Det är svårt att se en systematisk skillnad i värdering mellan normbrytande *de* och *dem*. Från 2020 får dock inlägg med normbrytande *dem* i snitt lite lägre karmapoäng än resten. Det kan vara en fingervisning om att normbrytande *dem* har värderats lägre under de senaste åren.



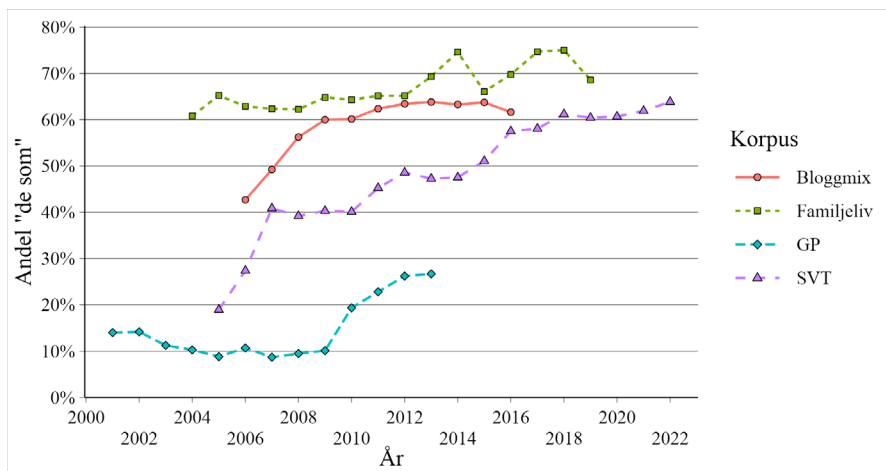
Figur 7. Karmapoäng (Reddit).

8 Användning av *de som* (i stället för *dem som*)

Nu fokuserar vi på **forskningsfråga 5**: Hur ofta används *de som* i stället för *dem som* efter preposition? Ser vi en förändring över tid? Den lilla forskning

som finns visar på en ökad användning av *de som* i nyhetstexter under slutet av 1900-talet.

Figur 8 visar att andelen *de som* även ökar i vårt material under de senaste tjugofem åren. GP har lägst andel *de som*. Vi ser först en lång period av stabil användning strax över 15 % och sedan en ökning till knappt 30 % – ett procenttal inte så långt från andelen på 32 % som Ewerth (1992) redovisar för åtta dagstidningar publicerade 1992. SVT har en större preferens för *de som* och ökar sin användning från relativt blygsamma 20 % till strax över 60 %. Detta innebär att *de som* har gått förbi *dem som* i frekvens under de senaste åren och nu är den vanligaste varianten. Samma utveckling syns på Bloggmix där *de som* gradvis knappar in på *dem som* och snart blir den dominanta varianten. Familjeliv, slutligen, visar en preferens för *de som* från start. Vi ser att andelen *de som* ligger ganska stabilt mellan 60 och 75 % över åren. Sammanfattningsvis ser vi att *de som*, många år efter att varianten har accepterats som alternativ för *dem som*, är eller har blivit den vanligaste varianten i tre av våra fyra korpusar. I korpusar där *de som* har varit dominant under flera år ser vi att dess andel har stabiliserats kring 60–75 %. Detta innebär att stavningen *dem som* inte helt överges av språkanvändare.



Figur 8. Andel *de som*, av samtliga *de som/dem som* (per korpus).

9 Avslutande diskussion

Med hjälp av en mikrodiakron korpusstudie har vi nyanserat och kompletterat kunskapen om användningen av *de*, *dem* och *dom* under de senaste tjugofem åren. Vi har utforskat fem forskningsfrågor som berör olika aspekter av dessa tre ord.

Forskningsfråga 1 undersöker användningen av *dom* som talspråksnära alternativ för *de* och *dem*. Vi ser att *dom* sällan används i nyhetstexter, där formen utgör några enstaka procent av alla *de*, *dem* och *dom*. Däremot är *dom* vanligt i sociala medier, där vi ser en användning på mellan 5 och 25 %. Där ökar varianten i början av 2000-talet, samtidigt som internet och sociala medier får sitt stora genombrott, men minskar därefter igen. Den nedåtgående trenden syns för alla generationer av skribenter, men yngre skribenter använder *dom* i större utsträckning än äldre skribenter.

Forskningsfråga 2 tittar på användningen av normbrytande *de* i stället för *dem* i objektsposition. Normbrytande *de* förekommer relativt sällan i våra korpusar, även om vi kan se att andelen ökar något. Överlag är normbrytande *de* vanligare i sociala medier – normalt mellan 4 och 10 % av alla *de/dem* som är objekt – än i nyhetstext – mellan 1 och 4 % av alla *de/dem* i objektsposition. Vi ser inga större skillnader mellan de olika generationernas användning av normbrytande *de*.

Forskningsfråga 3 fokuserar på normbrytande *dem* som används i stället för *de* i subjektposition eller som bestämd artikel. Vi ser att normbrytande *dem* är mindre vanligt än normbrytande *de* men att det finns en antydning till ökning över tid i sociala medier, från runt 1 % till uppemot 3–7 % av *dem* i subjektposition. Varianten förekommer ytterst sällan i våra korpusar med nyhetstexter. Normbrytande *dem* är något vanligare hos yngre skribenter än äldre.

Forskningsfråga 4 undersöker hur normbrytande *de* eller *dem* värderas på sociala medier. Vi ser att inlägg med normbrytande *de* eller *dem* i genomsnitt får lägre karmapoäng på Reddit än inlägg där *de* och *dem* används enligt normen. De senaste åren har inlägg med normbrytande *dem* lägre karmapoäng.

Forskningsfråga 5 tittar på användningen av *de som* i relation till *dem som* efter preposition. Vi kan se att *de som* har något högre frekvens i sociala medier. Andelen *de som* ökar i alla våra material och mot slutet av den undersökta tidsperioden är det den vanligare varianten på över 60 % av alla *de som/dem som* i de flesta korpusarna. *Dem som* tycks dock inte helt överges av språkanvändare.

Med våra samlade resultat kan vi nu besvara frågan som vi ställde i början av artikeln: kan vi se en förändring i den faktiska användningen av *de*, *dem* och *dom* i nutida skriftspråk?

Våra korpusar med nyhetstexter visar en bild av relativ stabilitet. Talspråksnära *dom* samt normbrytande *de* och *dem* förekommer ganska sällan i vårt nyhetsmaterial, och deras frekvens ligger kvar på samma låga nivåer de senaste tjugofem åren. Dessa resultat bekräftar bilden från tidigare forskning om att *de*, *dem* och *dom* överlag används enligt normen i nyhetstexter. Andelen *de som* ökar däremot stort i vårt nyhetsmaterial – en utveckling som inte har lyfts i den befintliga forskningen. Att forskningen inte har intresserat sig så mycket för *de som* under de senaste åren har givetvis att göra med att *de som* inte längre är en normbrytande variant. Det är dock intressant att se att varianten fortsätter att öka i frekvens många år efter att den har accepterats som norm.

Vår undersökning av sociala medier avslöjar ett mer vacklande bruk av *de*, *dem* och *dom*. Här pekar utvecklingen i verklig tid åt olika håll. Talspråksnära *dom* syns ganska ofta på sociala medier och blir initialt också vanligare men tappar sedan mark igen. Normbrytande *de* och *dem* blir däremot lite vanligare över tid, om än på låga nivåer. Dessa skilda trender visar att utvecklingen av talspråksnära *dom* samt normbrytande *de* och *dem* inte nödvändigtvis är ihopkopplade utan har gått sina egna vägar under de senaste åren. Detta kan vara av intresse för den aktuella språkdebatten, där en ökad osäkerhet i valet mellan *de* och *dem* har lyfts som argument för att genomföra en *dom*-reform.

Tack vare information om användarnas ålder på sociala medier kan vi komplettera vår kunskap om utvecklingen i verklig tid med utveckling i skenbar tid. Där ser vi att yngre skribenter använder sig mer av talspråksnära *dom* och normbrytande *dem* än äldre skribenter. Man kan tolka den trenden som ett tecken på språkförändring i skenbar tid. Om det är så kan vi förvänta oss fler talspråksnära *dom* och normbrytande *dem* i framtiden för hela populationen. Tendensen kan dock också förklaras som åldersgradering: som ung är man språkligt innovativ, men med åldern blir man mer konservativ. Om det stämmer förväntar vi oss att 1990-talisterna om tio år skriver ungefär som 1980-talisterna gör nu, 1980-talisterna som 1970-talisterna osv. och att det inte blir någon språkförändring på gruppnivå.

Så, kan vi säga att bruket av *de*, *dem* och *dom* är i gungning i nutida skriftspråk? Ja, vi har sett att andelen talspråksnära *dom* samt normbrytande *de* och *dem* har förändrats under de senaste tjugofem åren, framför allt på sociala medier. Här ska man dock komma ihåg att det ofta handlar om små förändringar och att förändringarna inte alltid går åt samma håll. Möjligen har norm-

brytande *de* och *dem* ökat något, framför allt bland yngre skribenter på sociala medier. Däremot verkar inte *dom* ha framträtt som ett tydligt alternativ för att undvika osäkerhet.

Efterord

Forskningen beskriven i denna artikel har gjorts inom ramen för projektet Cassandra som finansieras av Marcus och Amalia Wallenbergs Minnesfond (MAW 2020.0060) 2021–2024. Den har möjliggjorts genom den nationella forskningsinfrastrukturen Nationella språkbanken, finansierad av Vetenskapsrådet (2018–2024, kontrakt 2017-00626) och de 10 deltagande partnerinstitutionerna. Den har också möjliggjorts av språkmodeller tränade på KB-labb, den nationella forskningsinfrastrukturen på Kungliga biblioteket (Börjeson m.fl. 2023). Författarna har alla bidragit till artikelns koncept och text. EC har även stått för litteraturstudien och en stor del av skrivandet. YA har också tagit en större del av textskrivandet. FR har extraherat data från Reddit, utvecklat och applicerat språkmodellen och visualiserat data. AB har sammanställt korpusen Familjeliv-generationer, samt extraherat och analyserat data från Språkbankens korpusar. Vi tackar Olle Josephson för hans kommentarer på en tidigare version av denna artikel och två anonyma granskare för deras noggranna läsning och konstruktiva kommentarer.

Litteratur

- Berdicevskis, Aleksandrs, Coussé, Evie, Koplenig, Alexander, & Adesam, Yvonne, under utgivning: To drop or not to drop? Predicting the omission of the infinitival marker in a Swedish future construction. I: *Corpus Linguistics and Linguistic Theory*.
- Birkebo, Henrik, 2016: Nu är det dags att slopa de och dem. I: *Svenska Dagbladet*. 2016-10-22.
- Blythe, Richard & Croft, William, 2012: S-curves and the mechanisms of propagation in language change. I: *Language* 88(2). S. 269–304.
- Borin, Lars, Forsberg, Markus & Roxendal, Johan, 2012: Korp – the corpus infrastructure of Språkbanken. I: *Proceedings of LREC 2012*. Istanbul: ELRA. S. 474–478.
- Bylin, Maria & Melander, Björn, 2023: Språkrådet rekommenderar. Perspektiv, metoder och avvägningar i språkriktighetsfrågor. Stockholm: Språkrådet.

- Börjeson, Love, Haffenden, Chris, Malmsten, Martin, Klingwall, Fredrik, Rende, Emma, Kurtz, Robin, Rekathati, Faton, Hägglöf, Hillevi & Sikora, Justyna, 2023: Transfiguring the Library as Digital Research Infrastructure: Making KBLab at the National Library of Sweden. SocArXiv preprint: doi:10.31235/osf.io/w48rf.
- Cassirer, Peter, 1976: Formerna av pronomenet *de* i södra Sverige. I: Nysvenska studier 55–56. S. 251–290.
- Ewerth, Sten, 1978: Ska dom eller skall de? I: Språkvård 1978(2). S. 25–30.
- Ewerth, Sten, 1992: Tidningstexter, om dom igen. I: Språkvård 1992(3). S. 17–19.
- Ewerth, Sten, 2008: Dom skriver fortfarande dem. I: Språktidningen 2008(6). S. 66–67.
- Grünbaum, Catharina, 1976: Språknämndens enkät till svensklärare. I: Språkvård 1976 (4). S. 3–23.
- Hallencreutz, Katharina, 1980: De-dem-dom i svenskt elevspråk. I: Kent Larsson (red.): Elevsvenska. (Ord och stil. Språkvårdssamfundets skrifter 12.) Lund: Studentlitteratur. S. 84–103.
- Hammarstedt, Martin, Schumacher, Anne, Borin, Lars & Forsberg, Markus, 2022: Sparv 5 User Manual. Teknisk rapport.
- Håkansson, Gisela & Norrby, Catrin, 2003: Vad kostar dem? I: Språkvård 2003(3). S. 11–16.
- Karlsson, Susanna & Lind Palicki, Lena, 2017: Svensklärares upplevelser av besvärliga språkriktighetsfrågor. I: Emma Sköldberg, Maia Andréasson, Henrietta Adamsson Eryd, Filippa Lindahl, Sven Lindström, Julia Prentice & Malin Sandberg (red.): Svenskans beskrivning 35. Göteborg: Institutionen för svenska språket vid Göteborgs universitet. S. 93–106.
- Lind Palicki, Lena & Karlsson, Susanna, 2019: Kan dagens elever verkligen inte skilja på *de* och *dem*? En granskning av elevers formvariation. I: Marco Bianchi, David Håkansson, Björn Melander, Linda Pfister, Maria Westman & Carin Östman (red.): Svenskans beskrivning 36. Uppsala: Institutionen för nordiska språk vid Uppsala universitet. S. 131–142.
- Lind Palicki, Lena & Karlsson, Susanna, 2020: Till doms över dom. Och över dom andra. En språkideologisk analys av en språkvårdsdebatt. I: HumaNetten 44. S. 175–206.
- Lind Palicki, Lena, 2022: Dags att vi skriver 'dom gamla husen'. I: Svenska Dagbladet. 2022-04-05.
- Lövestam, Sara, 2022: 'Dom' slår snart ut 'de' och 'dem'. I: Svenska Dagbladet. 2022-02-22.
- Malmsten, Martin, Börjeson, Love & Haffenden, Chris, 2020: Playing with Words at the National Library of Sweden – Making a Swedish BERT. arXiv preprint arXiv:2007.01658. <https://arxiv.org/abs/2007.01658>.
- Medvedev, Alexey N., Lambiotte, Renaud & Delvenne, Jean-Charles, 2019: The anatomy of Reddit: An overview of academic research. I: F. Ghanbarnejad, R. Saha Roy, F. Karimi, J.-C. Delvenne & B. Mitra (red.): Dynamics On and Of Complex Networks III. DOOCN 2017. Springer Proceedings in Complexity. S. 183–204.
- Molde, Bertil, 1971: Dom och andra talspråksformer. I: Språkvård 1971(2). S. 14–16.
- Molde, Bertil, 1980: Ingen av de som? I: Språkvård 1980(4). S. 12–20.

- Nevalainen, Terttu, 2015: Descriptive adequacy of the S-curve model in diachronic studies of language change. I: Christina Sanchez-Stockhammer (red.): *Can We Predict Linguistic Change?* (Studies in Variation, Contacts and Change in English 16.) Helsinki: VARIENG.
- Phillips, Fred. 2007. On S-curves and tipping points. I: *Technological Forecasting and Social Change* 74(6). S. 715–730.
- Platzack, Christer, 1976: Hur dom som blir de/dem som. I: Lars Svensson, Anne-Marie Wieselgren & Åke Hansson (red.): *Festskrift tillägnad Gösta Holm på 60-årsdagen den 8 juli 1976*. Lund: Studentlitteratur. S. 353–361.
- Rekathati, Faton, 2023: Deformer (Revision 83a43f2). <https://huggingface.co/Lauler/deformer>.
- Språktidningen & Språkrådet, 2021: Nyordslistan 2021. <https://www.isof.se/lar-dig-mer/kunskapsbanker/lar-dig-mer-om-nyord/nyordslistor/nyordslistan-2021>.
- Svenska språknämnden, 2005: *Språkriktighetsboken*. Stockholm: Norstedts förlag.
- Wiktorsson, Maria, 2018: How hybrid is blog data? A comparison between speech, writing and blog data in Swedish. I: *Nordic Journal of Linguistics* 41(3). S. 367–377.

Appendix A

Här utvärderar vi sökmetodens kvalitet (se avsnitt 3.2). För att testa hur pålitliga våra sökresultat är har vi undersökt stickprov manuellt. Vi har dragit 126 slumpmässiga observationer av respektive *de*, *dem* och *dom* från korpusar som innehåller sociala medier (inklusive Flashback och Twitter, som vi planerade att använda men sedan exkluderade från studien av tekniska skäl).

Korp

Vi har kontrollerat alla observationer i stickprovet manuellt och noterat om uppmärkningen innehåller kritiska fel. För **forskningsfråga 1** är kritiska fel av typen fel ordklass eller fel grundform (t.ex. *de* som står för [de:] istället för [dɔm]). Alla *dem* och *dom* i stickprovet har korrekt uppmärkning av grundform, men för *de* är enbart 89 % av observationerna korrekta (resterande 11 % borde normenligt ha skrivits *det* istället för *de*). Annoteringen av ordklass är korrekt för 98 % av observationerna för både *dem* och *de*, samt för 94 % av *dom*. Observera dock att alla ordklassfel gäller pronomen uppmärkta som determinerare eller vice versa. Eftersom vi inte ställer frågor som kräver att vi kan skilja mellan de här två ordklasserna, är träffsäkerheten egentligen 100 % för våra behov. Vi bedömer alltså att annoteringen är tillräckligt pålitlig.

Observera att våra stickprov inte visar hur många relevanta fall vi missar. Om t.ex. ett *dom* i en korpus felaktigt är uppmärkt som substantiv, när det ska vara ett pronomen, kommer vi inte att hitta det. Det bör dock poängteras att även om det finns *dom* som felaktigt annoterats som substantiv så är sådana fall få i förhållande till mängden *dom* som annoterats korrekt som pronomen. Vår sökning missar alltså några relevanta fall, men de är relativt få i sammanhanget.

Stickproven visar dock att det inte går att använda annoteringen för att identifiera normbrytande användning av *de* och *dem*. Det finns många fall av felaktig annotering på syntaktisk nivå. Tillsammans med ordklassfel (här är det viktigt att kunna skilja mellan pronomen och determinerare) och grundformsfel (*de* som står för [de:] istället för [døm]) innebär det att Korp inte kan användas för våra andra forskningsfrågor.

Annoteringsfelen gör det också svårt att utforska andra aspekter av *de*, *dem* och *dom* som har tagits upp i litteraturen. Vi kan till exempel inte undersöka om normbrytande *dem* används som subjekt eller artikel (jfr Hallencreutz 1980, Lind Palicki & Karlsson 2019) eftersom det automatiska annoteringsverktyget t. ex. aldrig märker upp *dem* som subjekt. Det går inte heller att automatiskt avgöra om *de*, *dem* eller *dom* kommer före eller efter det finita verbet (jfr Håkansson & Norrby 2003), eftersom annoteringsverktyget inte pålitligt kan identifiera rätt finit verb.

Språkmodellen

Med hjälp av vårt manuella kvantitativa stickprov utvärderar vi även språkmodellen som ska identifiera normbrytande användning av *de* och *dem*.

Täckningsgraden (*recall*) för modellens klassificering av samtliga fall där en skribent använt *dem* är 100 %. Av dessa är täckningen:

- 100 % för fall där *dem* används på ett normenligt sätt i texten (modellen klassificerar alltid ett *dem* använt som objekt till kategorin *dem*).
- 100 % för fall där *dem* används på ett normbrytande sätt i texten (modellen klassificerar alla *dem* som borde varit subjekt/artikel till kategorin *de*).

Täckningsgraden för modellens klassificering av samtliga fall där en skribent använt *de* är 96 %. Av dessa är täckningen:

- 96 % för fall där *de* används på ett normenligt sätt i texten.

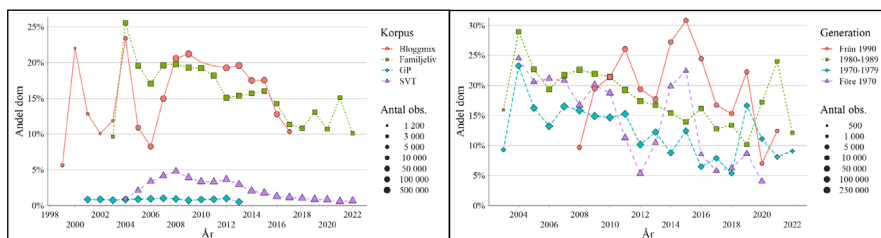
- 93 % för fall där *de* används normbrytande men där objektsformen *dem* egentligen var korrekt. Eftersom stickprovet inte innehöll tillräckligt med tydliga fall där *de* användes istället för *dem* (flera fall var tvetydiga, där skribenten med stor sannolikhet i sin användning av *de* avsåg betydelsen *det* snarare än *dem*), berikades stickprovet med 15 manuellt utvalda fall där *de* uppenbart användes istället för *dem*.
- 100 % för fall där *de* används normbrytande men där singularformen *det* egentligen avsågs (modellen fångar upp alla *de* som borde varit singularformen *det* och klassificerar dessa till kategorin *det*).

När skribenten använt *de* och modellen förutsäger *det* har modellen gissat rätt i 78 % av alla fall. Modellen fångar alltså upp 100 % av de fall där *de* används normbrytande när singularformen *det* egentligen avsågs, men påstår ibland felaktigt att normenligt *de* står för *det*.

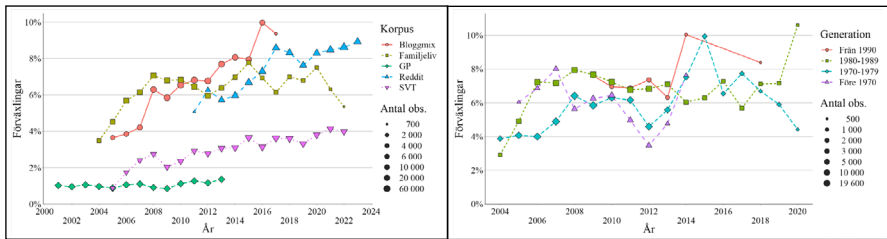
Modellen presterar alltså bra och kan, till skillnad från Korp, användas för att studera normbrytande användning av *de* och *dem*. Man bör dock tänka på att en del observationer felaktigt filtreras bort, eftersom modellen överreagerar på *de* som kan vara *det*.

Appendix B

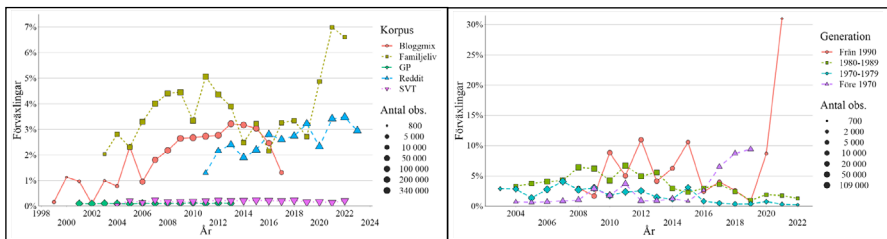
För att visa hur många observationer varje datapunkt representerar, visualiserar vi våra resultat i figur B1–B4 där vi skalar punktstorlek efter antalet observationer under ett år för en given korpus eller generation.



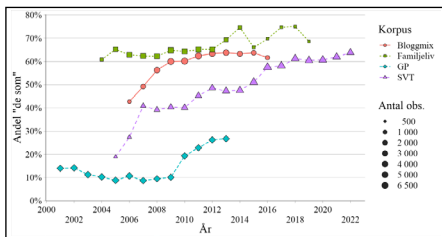
Figur B1. Andel *dom*, av samtliga *de/dem/dom*. Korpus (vänster bild). Generationer (höger bild).



Figur B2. Andel *de*, av samtliga *de/dem* som är objekt. Korpus (vänster bild). Generation (höger bild).



Figur B3. Andel *dem*, av samtliga *de/dem* som är subjekt/artikel. Korpus (vänster bild). Generation (höger bild).



Figur B4. Andel *de som*, av samtliga *de som/dem som*.

© 2024 Författaren/författarna

Denna artikel publiceras med öppen tillgång under villkoren i
Creative Commons Erkännande-licensen ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))

Språk och stil: Tidskrift för svensk språkforskning, NF 33, 2023. ISSN: 1101-1165 E-ISSN: 2002-4010