

Språkförändring på bar gärning En storskalig korpusstudie av pågående förändringar i stavning, lexikon och grammatik

Yvonne Adesam¹, Aleksandrs Berdicevskis¹, Evie Coussé²

¹Språkbanken Text, Institutionen för svenska, flerspråkighet och språkteknologi

²Institutionen för språk och litteraturer
Göteborgs universitet

1 Introduktion

Språk förändras ständigt och dessa förändringar studeras ofta genom att jämföra äldre historiska texter med varandra eller genom att sätta det historiska språket i relation till nutida språkbruk. Men frågan är om det går att undersöka pågående språkförändringar i nutid.

Med hjälp av ett skarpt språköra kan man observera språkförändring i det dagliga språket omkring sig. Språkspalter brukar gärna innehålla sådana spaningar och vi ger här tre typiska exempel.

(1) ”Allt fler har tappat distinktionen mellan objektsform och subjektsform för personliga pronomen i tredje person plural.” (Sara Lövestam i Svenska Dagbladet 2022-02-22)

(2) ”Nya ord bromsar läshastigheten men motståndet minskar med tiden. Därför sticker vårt könsneutrala pronomen ’hen’ inte längre ut – trots att det bara blivit vanligare.” (Anders Svensson i Dagens Nyheter 2021-01-24)

(3) ”Det är alldeles uppenbart att det blivit mycket vanligare under senare år att infinitivmärket *att* saknas efter *kommer*.” (Lars-Gunnar Andersson i Göteborgs-Posten 2017-01-21)

Även om dessa observationer är outhärliga för att komma språkförändringar på spåret, tenderar de att vara impressionistiska eftersom de är baserade på egen, och därmed begränsad, erfarenhet.

I det nya projektet Cassandra¹ vill vi observera språkförändring i nutida svenska med hjälp av korpuslingvistiska metoder. På så sätt vill vi systematiskt och storskaligt kartlägga olika typer av pågående språkförändringar, för att sedan kunna testa hur bra olika språkförändringsteorier är på att förklara och förutsäga hur språket utvecklas. Vi använder oss av stora korpusar hos Språkbanken Text med texter från flera år i rad. Sammanlagt innehåller dessa korpusar, som består främst av nyhetstexter och texter hämtade från sociala medier, mer än tio miljarder token (löpard, inklusive skiljetecken). Dessa jättelika korpusar låter oss följa hur språkliga innovationer sprider sig i olika textgenrer från år till år.

Vi illustrerar vår forskningsmetod med de tre påstådda fall av förändringar som togs upp i citaten (1) till (3), och utforskar om de går att spåra i vårt material. Först tittar vi på stavning, och undersöker om den uttalsnära stavningen *dom* håller på att ersätta den normativa distinktionen mellan *de* och *dem*, i enlighet med (1). Sedan går vi vidare till svenskans lexikon, och utforskar om det nya könsneutrala pronomenet *hen* ”bara blivit vanligare”, som

¹ <https://spraakbanken.gu.se/projekt/cassandra>.

(2) påstår. Till slut kommer vi till grammatik, och studerar om bortfall av *att* efter *kommer* "blivit mycket vanligare under senare år", såsom (3) ser det. Forskare och kolumnister beskriver inte bara nutiden och det förflutna, utan gör också förutsägelser om framtiden. Vi diskuterar därför i vilken mån sådana förutsägelser är baserade på vetenskaplig grund och om dagens språkvetenskap egentligen kan erbjuda någon grund för sådana frågor.

2 Material och metod

Tabell 1 ger en översikt över vårt textmaterial. Texterna från sociala medier utgör största delen av vårt material med ungefär 10,1 miljarder token, medan nyhetstexterna står för sammanlagt 505 miljoner token. Av tekniska skäl används Twitter inte i de två första fallstudierna.

Tabell 1. Korpusar

| Korpus | Textgenre | Period | Antal miljoner token |
|---|------------------|------------------|----------------------|
| Göteborgs-Posten (GP) | Nyheter | 2001-2013 | 21 |
| Sveriges Television (SVT) Nyheter | Nyheter | 2004-2021 | 214 |
| Webbnyheter (nyheter från Svenska Dagbladets och Dagens Nyheters webbplatser) | Nyheter | 2001-2013 | 270 |
| Twitter | Mikroblogg | 2006-2019 | 2 000 |
| Flashback | Diskussionsforum | 2001-2021 | 3 700 |
| Familjeliv | Diskussionsforum | 2003-2021 | 4 400 |
| Totalt | | 2001-2021 | 10 605 |

Vi använder texterna från sociala medier av främst två anledningar. Den första är att dessa korpusar utgör en mycket stor mängd text, som dessutom kan undersökas diakront över de senaste två decennierna. Materialet kan alltså sägas utgöra ett urval av nutida svenska. Den andra anledningen är att texterna utmärker sig genom att vara en mer oredigerad, talspråksnära kommunikation. Vi gör antagandet att detta innebär att vissa språkförändringar syns tidigare i denna typ av material än i nyhetstexterna, som i betydligt större utsträckning är redigerade och därmed kan antas vara mer språkligt konservativa. Nyhetstexterna används alltså i viss mån som ett språkligt standardiserat referensmaterial. Även om nyhetstexterna inte utgör ett lika stort material som texterna från sociala medier så är det fortfarande ett stort material med mycket text över flera år.

Diskussionsforumet Familjeliv innehåller diskussioner inom många olika ämnen, men störst fokus ligger på familjerelaterade ämnen, som graviditet och föräldraskap. Flashback har också diskussioner kring en stor mängd ämnen – från privatekonomi, politik och sport till droganvändning, kriminalitet och psykisk ohälsa – och finns med bland de mest använda sociala medierna i Sverige². På Familjeliv dominerar kvinnliga användare, medan de manliga dominerar på Flashback³. Twitter skiljer sig textuellt främst från diskussionsforumen genom att varje inlägg har en begränsad längd. Vad gäller nyhetsmaterialet har SVT och

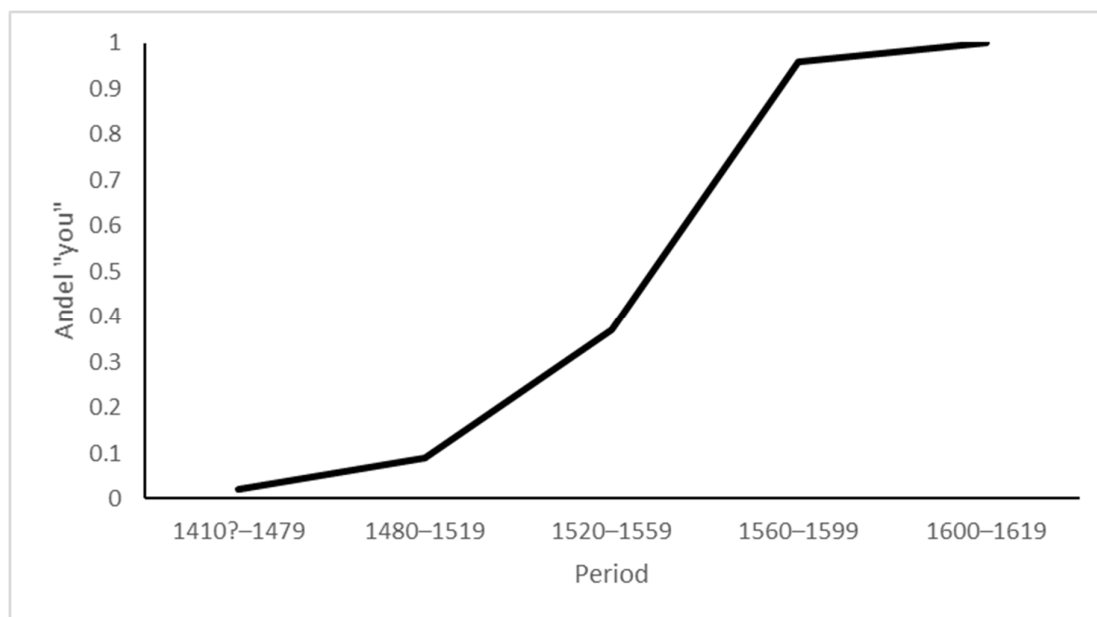
² <https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2022/sociala-medier/>.

³ <https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2021/sociala-medier/man-dominerar-pa-twitch-reddit-och-twitter-och-kvinnor-pa-pinterest-och-familjeliv/>.

Webbnyheter hämtats från webbsidor, men även GP finns på webben och får antas ha texter som skrivits för webbpublicering. Gemensamt för alla materialen är att de har skapats i digital form och alltså inte har digitaliserats i efterhand.

Vi söker i alla korpusar med hjälp av Korps (Borin et al. 2012) API⁴, vilket också tillåter oss att genomföra sökningar för avgränsade tidsintervaller, i detta fall ett år i taget. Vi söker i första hand på ordform. Det beror på att den språkliga uppmärkningen i korpusarna, annoteringen, är automatiskt gjord och inte alltid är pålitlig. Detta är framför allt tydligt i sociala medier som inte alltid följer traditionella skrivkonventioner (inga fullständiga meningar, avvikande användning av skiljetecken, emojis) och därför innebär en utmaning för annoteringsverket Sparv (Hammarstedt et al. 2022). Annoteringen behövs dock ofta eftersom enbart ordformen inte är tillräcklig för att avgränsa en fallstudie. Vi väljer att använda oss sparsamt av Sparvs automatiska annoteringar, fullt medvetna om att det kan medföra en viss risk för missar. Fler detaljer om våra sökningar ges i beskrivningen av varje fallstudie nedan.

Våra resultat presenteras som trendlinjer i tre grafer (en för varje fallstudie). Varje trendlinje avbildar den relativa frekvensen av den nya språkliga varianten (*dom*, *hen* och *att*-bortfall) gentemot en eller flera befintliga varianter som den nya varianten konkurrerar med. Vi tar bort år där det totala antalet observationer (av både den nya och den befintliga varianten) är mindre än 100. Vissa typer av språkförändring tycks ofta genomgå flera faser där ökningen av en nya variant först sker långsamt, sedan tar fart, och till slut mattas av igen, vilket ger en graf i form av en så kallad S-kurva (Labov 1972, Kroch 1989, Nevalainen 2015). Figur 1 visar andelen *you* (i relation till *ye*) som subjekt i engelska (Nevalainen & Raumolin-Brunberg 2016:251) och kan sägas uppvisa en klassisk S-kurva.



Figur 1. Andel *you* (frekvens av *you* mot frekvens av *you* och *ye* sammantaget). Efter Nevalainen & Raumolin-Brunberg (2016:251).

⁴ <https://ws.spraakbanken.gu.se/docs/korp>.

Vi utforskar för våra tre fallstudier om trendlinjerna visar tecken på en snabb frekvensökning följt av en eventuell avmattning av ökningen, med andra ord huruvida de uppvisar en S-kurva eller ett fragment därav.

3 Stavning

Vår första fallstudie behandlar en långlivad debatt gällande stavning. Lövestam (2022) hävdar att ”Allt fler har tappat distinktionen mellan objektsform och subjektsform för personliga pronomen i tredje person plural”. Enligt normen stavar man *de* i subjektsposition och *dem* i objektsposition (Språkriktighetsboken 2005:203). Nedanstående exempel⁵ (4) och (5) från Flashback illustrerar den normerade distinktionen mellan *de* och *dem*.

(4) Sega gubbar på bygge, vad orkar *de* med??

(5) Meningen är inte att katten ska äta *dem*.

Distinktionen mellan *de* och *dem* motsvarar inte längre någon skillnad i uttalet hos många talare i Sverige, utan uttalas [dɔm] (Cassirer 1976). Språkvårdare har därför sedan sjuttioalet förordat att ersätta *de* och *dem* med den mer uttalsnära stavningen *dom* i både subjekts- och objektsposition (se Lind Palicki & Karlsson 2020 för en översikt). Exemplet (6) och (7) nedan visar *dom* i respektive subjekts- och objektsposition.

(6) Lol sitter *dom* under ett skynke eller nått då typ?

(7) Det är bara det att jag vill undvika *dom*.

Här fokuserar vi på frågan om hur vanlig den nya stavningsvarianten *dom* har blivit i skriftspråket. Redan på sjuttioalet observerade Molde (1971:15) att ”formen *dom* under de senaste åren har blivit allt vanligare i skrift (både som ersättning för *de* och som ersättning av *dem*)”. Detta sporrar Ewerth (1978, 1992, 2008) att undersöka hur vanligt bruket av *dom* egentligen är i pressen. Hans undersökning begränsar sig inte bara till stavningen av det personliga pronomenet i tredje person plural utan behandlar även den bestämda artikeln i plural. Artikeln ska enligt normen stavas *de*, men uttalas [dɔm], vilket ibland återspeglas i den uttalsnära stavningen *dom*. Exempel (8) och (9) illustrerar artikelns stavning som respektive *de* och *dom* i Flashback.

(8) Vad säger *de* andra grannarna egentligen??

(9) Åk med *dom* vanliga bussarna istället.

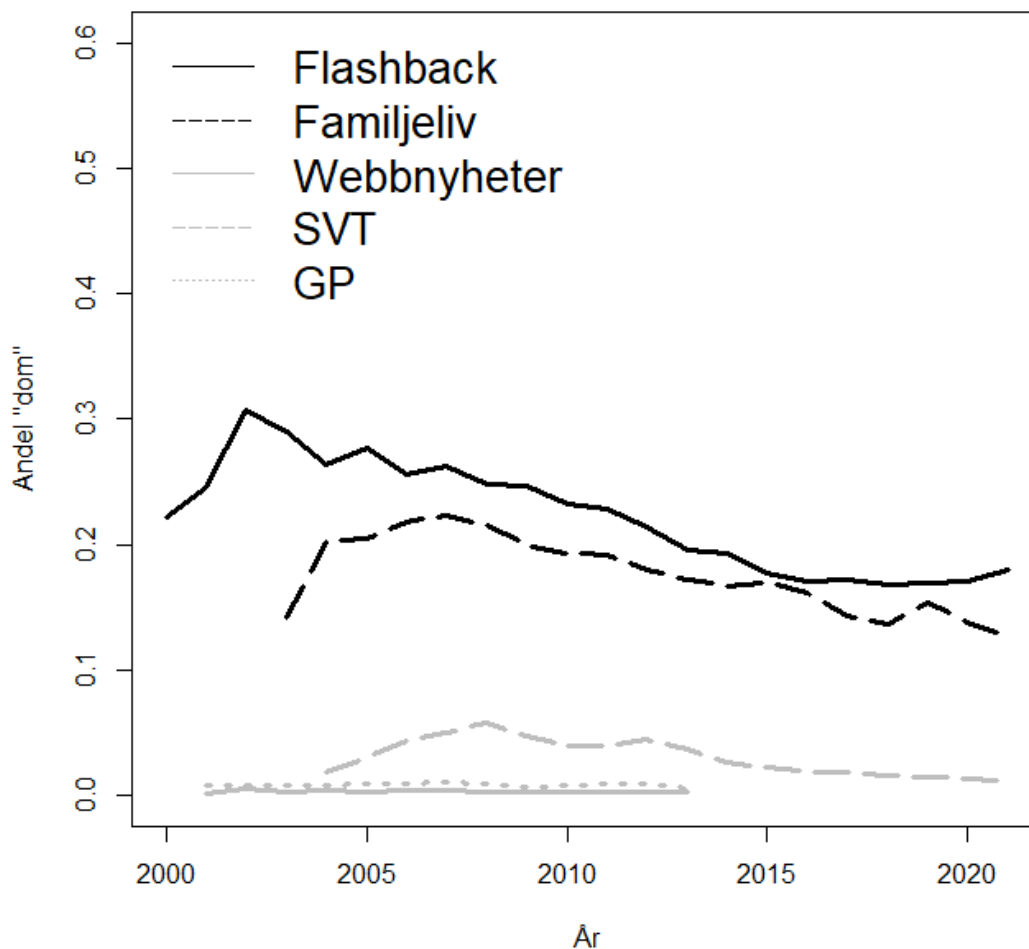
Ewerth finner att andelen *dom* på *de/dem/dom* är 2,4% i åtta dagstidningar från år 1977, vilket minskar till 1,8% i motsvarande tidningar år 1992 och till 0,7% år 2007. Stavningen *dom* är alltså ganska ovanlig i nyhetstexter, och verkar till och med bli allt ovanligare under perioden 1977–2008.

Vi upprepar Ewerths undersökning i vårt korpusmaterial genom att söka efter ordformerna *dom*, *de* och *dem* (skiftlägesoberoende). För att undvika substantivet *dom* och ordformen *de* i

⁵ Vi framhäver de nya och befintliga språkvarianterna i alla exemplen med kursiv stil.

utländska egennamn (som i *Rio de Janeiro* eller *Tour de France*), begränsar vi urvalet till ordklasserna 'pronomen' och 'determinator'.

Figur 2 avbildar andelen *dom* på den totala frekvensen av *de/dem/dom* per år i alla våra korpusar utom Twitter. Där ser vi att andelen *dom* generellt är låg i nyhetskorpusarna (de grå trendlinjerna). Andelen *dom* är nästan obefintlig i Webbnyheterna och GP (mellan 0,1% och 1%) medan nyhetstexterna på SVT uppvisar ett något högre bruk av *dom* (mellan 1% och 6%). Vårt material bekräftar alltså Ewerths observation om att *dom* är mycket ovanligt i tidningstexter. Andelen *dom* i nyhetstexter ligger ganska stabilt över tid. Webbnyheterna och GP ligger kvar på samma låga nivå under hela perioden 2001–2021, medan SVT uppvisar en liten ökning under perioden 2004–2008 och sedan en tillbakagång. Nyhetstexterna visar alltså under de senaste åren inga tecken på att *dom* håller på öka i frekvens.



Figur 2. Andelen *dom* (frekvens av *dom* mot frekvens av *de* och *dem* sammantaget). Observera skalan på y-axeln: den går till 0,6 och inte 1,0 för bättre synlighet. Det totala antalet observationer är ca 78 miljoner, varav ca 14 miljoner *dom*.

Andelen *dom* ligger betydligt högre i våra korpusar med sociala medier (svarta trendlinjer). På diskussionsforumen Flashback och Familjeliv ligger andelen *dom* närmare bestämt mellan

13% och 29%. Båda korpusarna visar initialt en kort men ganska kraftig ökning av andelen *dom*, följt av en långsam tillbakagång. Trendlinjerna uppvisar därmed ingen klassisk S-kurva. Blythe & Croft (2012:208) introducerar begreppet ”avbruten S-kurva” (en språklig innovation som efter en snabb uppgång tappar mark igen), vilket möjligtvis kan användas om den här trendlinjen. Så vitt vi vet finns det dock inga tydliga kriterier för att känna igen en avbruten S-kurva. Det verkar som om uppkomsten av sociala medier under tidigt 2000-tal leder till ett kortvarigt uppsving i bruket av *dom* men att onlineskribenterna sedan gradvis överger varianten igen. Även i sociala medier kan vi alltså inte se att *dom* är på väg att ta över som stavningsvariant för *de* och *dem*. Tendensen verkar snarare vara tvärtom.

Våra resultat för både nyhetstexter och sociala medier ger sammanfattningsvis idag inte något stöd för Lövestams observation i (1). Men Lövestam erbjuder inte bara en observation om det förflutna, utan också en förutsägelse om framtiden.

(10) ”Låt mig därför komma med en spådom: ’de’ och ’dem’ ersätts på sikt av ett enda ’dom’ även i skriftspråket.” (Sara Lövestam i Svenska Dagbladet 2022-02-22)

Det är betydligt svårare att avgöra om våra data stödjer denna spådom. En enkel (linjär) extrapolering av den nuvarande trenden gör inte det – trenden går neråt. Men det är inte givet att trenden måste vara linjär eller monoton (enbart ökande eller enbart avtagande), som figur 2 visar. Hur vet vi då om vi befinner oss vid en vändpunkt just nu? Och går det egentligen att veta om det kommer en vändpunkt? Vi återkommer till den här frågan i kommande avsnitt.

4 Lexikon

Nu riktar vi vår uppmärksamhet mot ett nytt tillskott i svenskans lexikon: det könsneutrala pronomenet *hen*. Språkriktighetsboken (2005:78) påpekar att ”det är svårt att lansera pronomen från ingenting i ett så stabilt språkssystem som svenskan” och bedömer därför att utsikten för att ett nytt köns neutralt pronomen ska etableras är minimal. Ändå var det vad som hände när användningen av det könsneutrala *hen* tog fart 2012, efter att ha diskuterats i debattartiklar (bl a Milles et al. 2012), även om det först föreslogs mycket tidigare.

Det nya pronomenet *hen* används främst i subjektposition, men även i objektposition (Milles 2013:136). Nedanstående exempel (11) och (12) illustrerar bruket av *hen* som respektive subjekt och objekt i Familjeliv.

(11) TS får tro på vad *hen* vill.

(12) Försök få kontakt med en kurator eller psykolog och be *hen* om hjälp.

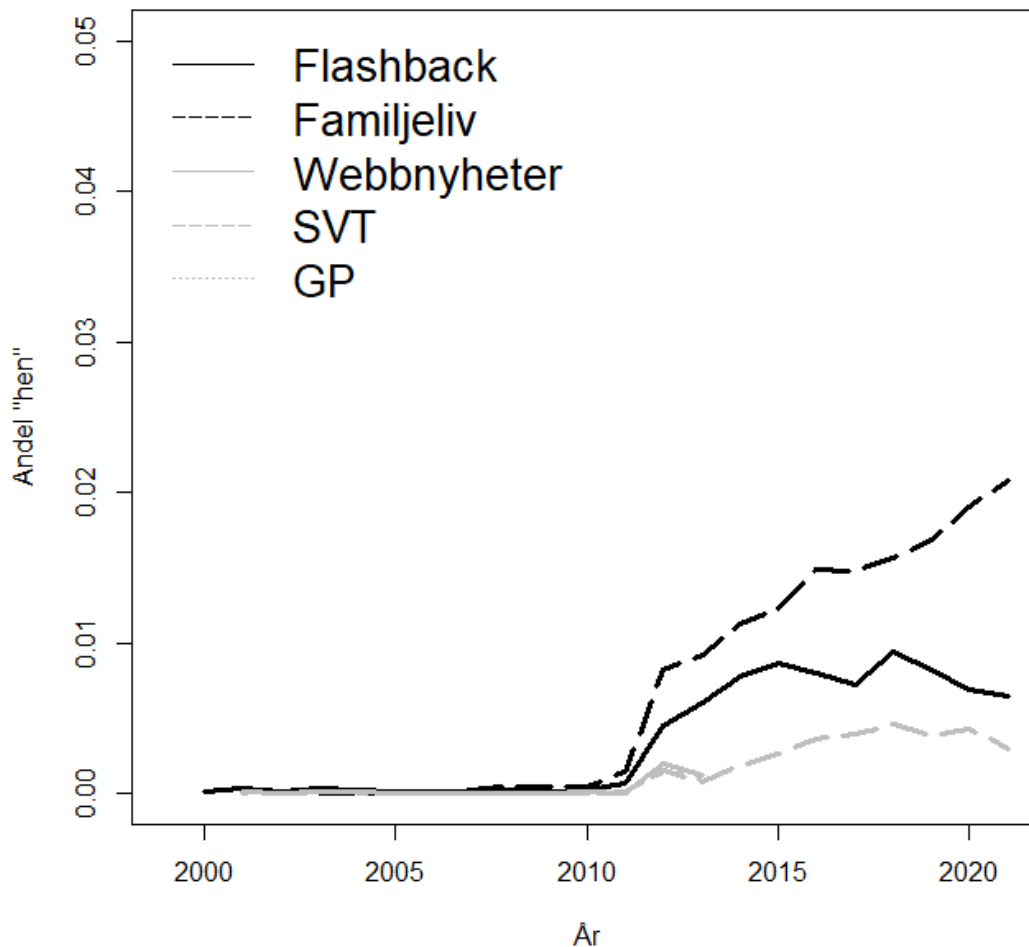
Pronomenet kan ha definit referens, för att syfta på en specifik individ som man inte vet eller vill avslöja könet på, som TS (som står för *trådstartaren*) i (11), eller icke-definit referens, som i (12) (Ledin & Lyngfelt 2013). Sedan kan man också urskilja meta-*hen*, som i (13), där man benämner uttrycket.

(13) *Hen* är ju bara ett ord, jag förstår inte varför folk blir så upprörda över ordet *hen*?

Här fokuserar vi på spridningen av *hen* i skriftspråket under det senaste tjugo åren. Ledin & Lyngfelt (2013) visar att *hen* förekommer sparsamt i bloggtexter och nyhetstexter från mitten av 2000-talet, men att spridning av det nya pronomenet tar fart från år 2012. Milles (2013)

hittar samma trend i tidningstexter från 2011 till och med 2013, där *hen* först börjar dyka upp i texter från 2012. Spridningen av *hen* verkar hålla i sig under senare år, enligt Anders Svenssons spaning i (2).

För att undersöka hur ofta det könsneutrala *hen* används istället för *hon/henne* eller *han/honom*, söker vi efter ordformen *hen* kontra *hon*, *henne*, *han*, *honom* (skiftlägesoberoende) i våra korpusar. Lagg märke till att vår automatiska sökning inte skiljer mellan det faktiska bruket av *hen* och meta-*hen*⁶.



Figur 3. Andelen *hen* (frekvens av *hen* vs. frekvens av *hon*, *han*, *henne*, *honom* sammanlagt). Observera skalen på y-axeln: andelen är alltid mycket liten. Trendlinjen för GP syns nästan inte, eftersom den är täckt av snarlika linjer för SVT och Webbnyheterna. Det totala antalet observationer är ca 81 miljoner, varav ca 285 000 *hen*.

Figur 3 visar andelen *hen* på den totala frekvensen av *hen*, *hon*, *han*, *henne* och *honom* i vårt korpusmaterial. Där ser vi att *hen* används i väldigt liten utsträckning under hela perioden 2000–2011. Nedanstående exempel (14) på meta-*hen* från 2005 visar dock att det

⁶ Detta gäller givetvis även för våra andra sökningar.

könsneutrala *hen* redan cirkulerade i vissa kretsar (här i tråden ”Feministmuffins och huligankaffe del 3!” på Familjeliv) innan det stora genomslaget år 2012.

(14) För övrigt vill jag för 2005 slå ett slag för orden ”*hen*” och ”*henom*” som komplement till *hon/han* och *henne/honom*.

År 2012 pekar sedan alla trendlinjer uppåt. Diskussionsforumet Familjeliv använder och diskuterar *hen* mest av alla våra korpusar under 2012. Webbnyheterna, GP och SVT däremot verkar vara mer återhållsamma i bruket av *hen*. Flashback ligger mitt emellan. Ordet fortsätter sedan att öka i frekvens efter 2012. För Flashback och SVT ser vi en avmattning eller till och med en nedgång från 2018. Trendlinjerna i dessa två korpusar har alltså formen av en S-kurva. Efter en ökningen av *hen* strax efter 2012 tycks ordet under de senaste åren ha legat på en plåtå i flera korpusar.

Sammantaget visar vår undersökning alltså att det könsneutrala *hen* fick sitt genombrott år 2012 och att det därefter har blivit något vanligare (märk dock att den relativa frekvensen av *hen* jämfört med de traditionella pronomina fortfarande är ytterst liten). Några av våra korpusar visar tecken på att den uppåtgående trenden har stannat av under de senaste åren.

Återigen är det oklart om det går att göra någon utsaga om framtiden. Den observerade avmattningen kan tyda på att uppkomsten av *hen* som könsneutralt alternativ för *hon/han* och *henne/honom* har kommit in i sin slutfas (S-kurvan har nått taket, som ligger mycket lågt). Går det att veta om det inte följer ett genombrott till? Man kan hoppas att få någon insikt i detta genom att lära sig mer om den sociolingvistiska situationen i den nuvarande fasen. Är det så att några få individer har gått över till *hen* och använder det konsekvent eller är det snarare så att ett större antal individer använder *hen* då och då eller bara i vissa kontexter? Finns det fler individer som går över till *hen*? Brukar de som redan har börjat använda ordet göra det mer aktivt? Hur ser åldersfördelningen ut? För att kunna säga något om framtiden behöver vi först besvara sådana frågor.

5 Grammatik

Till sist undersöker vi en potentiell pågående språkförändring inom svensk grammatik. Sedan 1990-talet har språkvårdare uppmärksammat att infinitivmärket *att* börjat falla bort efter det futurala hjälp verbet *kommer* (se Olofsson 2008 för en översikt). Exempel (15) illustrerar *kommer* med infinitivmärke och (16) *kommer* utan infinitivmärke framför infinitiv på Twitter.

(15) Detta *kommer att göra* skillnad för Borlängebon!

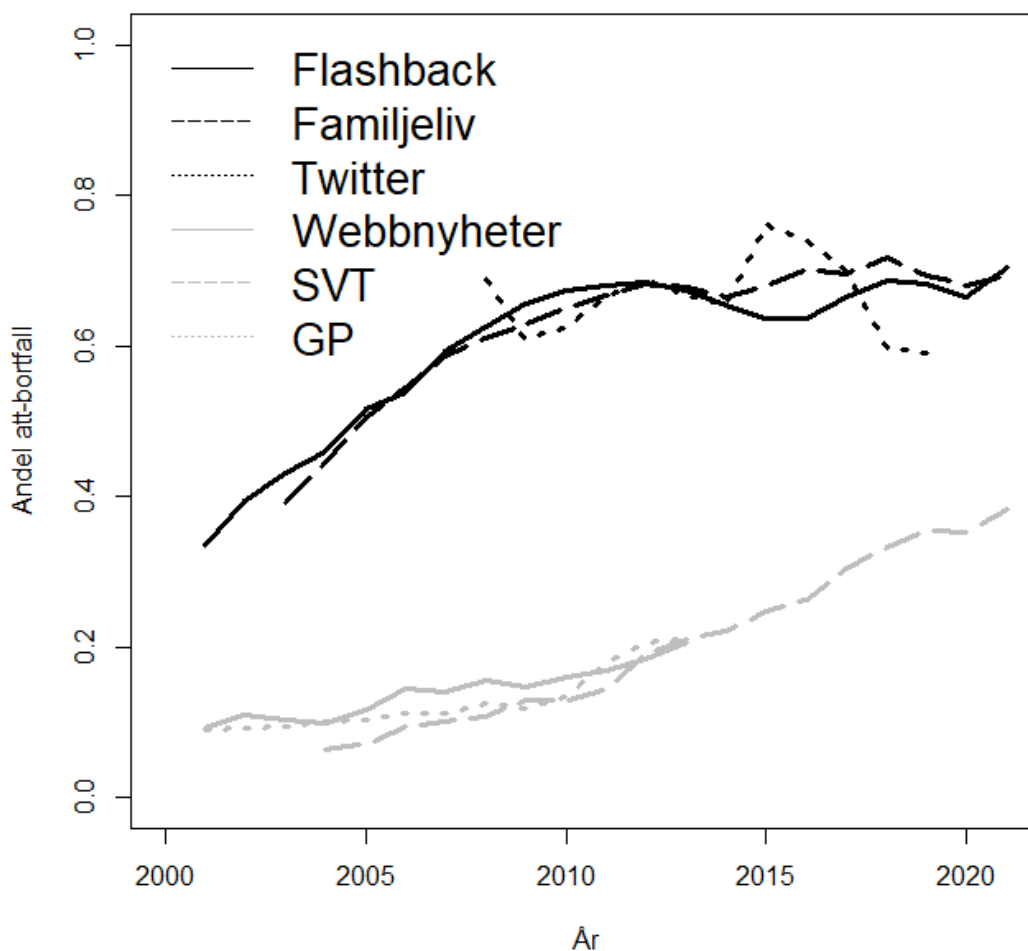
(16) Många pratar men det *kommer* nog aldrig bli så

Delsing (1993) och Olofsson (2008) observerar att *att*-bortfall är vanligare när det förekommer en eller flera satsdelar mellan *kommer* och infinitivfrasen, som i (16). Språkriktighetsboken (2005:357) påpekar att *kommer* utan *att* är vanligare när infinitivfrasen består av flera verb, som i (17).

(17) Men vart ska jag jobba asså *kommer* ju *behöva jobba* på Subway igen... önska bara de fanns jobb inom musik typ vad som

Här kommer vi att fokusera på en eventuell ökning av *att*-bortfall i nyhetstexter och sociala medier sedan millennieskiftet. Olofsson (2008) och Malmgren (2017) undersöker andelen *att*-bortfall i tidningstexter sedan 1960-talet. Fram till millennieskiftet förekommer *kommer* utan *att* ganska sällan, men därefter verkar spridningen av fenomenet ta fart. Malmgren (2017) rapporterar att *att*-bortfall i nyhetstexter från GP ökar från 3% år 2003 till 14% år 2013. Anderssons spaning från 2017 i (3) antyder att spridningen av *att*-löst *kommer* har fortsatt därefter. Blensenius & Rogström (2020) finner att *att*-bortfall är mindre vanligt i nyhetstexter jämfört med sociala medier.

Vi sökte efter ordformen *kommer* (skiftlägesoberoende) i den syntaktiska annoteringen antingen följd av en underordnad infinitiv eller följd av ett underordnat *att*, som i sin tur har en underordnad infinitiv. Det får inte förekomma fler än fem ord mellan *kommer* och infinitiven eller *att*. Sedan räknade vi antal fall där *att* är utelämnat.



Figur 4. Andel *att*-bortfall (frekvens av *kommer* utan *att* vs. *kommer* med *att*). Det totala antalet observationer är ca 13 miljoner, varav ca 8 miljoner med *att*-bortfall.

Figur 4 visar att fenomenet *att*-bortfall gradvis ökar i frekvens i alla våra korpusar. Vi ser dock att nyhetskorpusarna Webbnyheter, SVT och GP (de grå trendlinjerna längst ner) skiljer

sig markant från de sociala medierna Flashback, Familjeliv och Twitter (de svarta trendlinjerna högst upp). Andelen *att*-bortfall ligger genomgående 25–50 procentenheter lägre i nyhetstexter jämfört med sociala medier. Andelen *att*-bortfall ligger initialt under 10% i nyhetstexterna och kryper sedan sakta upp mot 35–40% år 2021. Diskussionsforumen Flashback och Familjeliv uppvisar redan från början en relativ hög andel av *att*-bortfall, på över 30%. Vi ser en kraftig ökning av *att*-löst *kommer* fram till nästan 70% år 2010. Därefter ligger *att*-bortfall ganska stabilt i diskussionsforumen, runt 65-70%. Trendlinjerna i diskussionsforumen skulle alltså kunna visa den mittersta och sista delen i en S-kurva, där *att*-bortfall efter en snabb expansionsfas nu har kommit in i en slutfas. Trendlinjen för Twitter uppvisar ingen S-kurva. Efter en (något ostadig) uppgång av *att*-bortfall, ser vi en ganska kraftig tillbakagång från år 2015.

Vår undersökning av olika korpusar visar alltså att sociala medier ligger före i utvecklingen av *att*-bortfall de senaste tjugo åren jämfört med nyhetstexter. Den uppåtgående trenden i sociala medier har mattats av på senare tid medan vi inte ser en sådan avmattning (än) i nyhetstexter. *Kommer* utan *att* verkar alltså ha etablerats som ett vanligt alternativ till *kommer* med *att* inom sociala medier, utan att den *att*-lösa varianten har tagit över helt.

Våra data stämmer alltså överens med Anderssons observation i (3). Precis som med *dom* och *hen* är det dock också intressant att fundera över om vi vågar säga något om framtiden. Andersson själv förutsäger att bortfallet av *att* kommer att bli absolut, se (18).

(18) ”I framtidens språkbruk kommer infinitivmärket (*att*) saknas efter *kommer* – men inte hos oss som lärde oss det som barn. Hos oss sitter *kommer å* i ryggmärgen.” (Lars-Gunnar Andersson i Göteborgs-Posten 2017-01-21)

Extrapoleringen av den nuvarande trenden i sociala medier tyder på att förändringen kan ha nått taket, men vi vill igen framhäva att en linjär extrapolering inte nödvändigtvis är pålitlig. Om den nuvarande situationen t ex beror på att *kommer* med *att* uteslutande finns hos äldre talare (som har det i ryggmärgen), medan yngre talare redan har övergett den varianten helt, så är (18) med stor sannolikhet korrekt. Om det dock är så att även många yngre talare behåller *att*, är det mindre klart vad som kommer att ske i framtiden. Man bör därför titta närmare på vilka som behåller *att* och i vilka kontexter de gör det.

6 Avslutande diskussion

Vi har undersökt tre påstådda fall av språkförändring med hjälp av korpuslingvistiska metoder. I två av de tre fallen kan vi påvisa en pågående språkförändring där en ny variant sprider sig som alternativ till en eller flera befintliga varianter. Vår fallstudie av det könsneutrala pronomenet *hen* visar att bruket av *hen* ökade i frekvens från 2012 i både nyhetstexter och sociala medier. Vår undersökning av *att*-bortfall efter hjälp verbet *kommer* visar att varianten utan *att* blev vanligare efter millennieskiftet, först i sociala medier och sedan i nyhetstexter. Någon bestående ökning i bruket av *dom* som uttalsnära stavningsalternativ för *de* och *dem* kan vi inte se i vårt korpusmaterial. Det är intressant att notera att ökningen av *hen* och *att*-bortfall verkar ha stannat av i några av våra korpusar.

Vår undersökning visar att språkförändring sprider sig olika fort i nyhetstexter och sociala medier. Det könsneutrala *hen* införs samtidigt i både nyhetstexter och sociala medier år 2012, men sprider sig betydligt fortare i sociala medier. *Att*-bortfall följer samma mönster. Vi ser att användare av sociala medier genomgående utelämnar *att* i högre utsträckning än skribenter av

nyhetstexter. Det antyder att sociala medier ligger i framkant vad gäller pågående språkförändring i nutida svenska, jämfört med nyhetstexter som har en mer återhållsam inställning till nya språkvarianter. Detta är sannolikt en följd av att sociala medier, och framför allt diskussionsforum, är mindre redigerade än nyhetstexter och närmare talspråket (se även Wiktorsson 2018 om bloggtext). Vi behöver dock undersöka fler fall av språkförändring för att se om den generaliseringen håller.

Vi har jämfört de observerade trendlinjerna med den klassiska S-kurvan. Dessa jämförelser är dock rent visuella, utan någon formell analys. Dessutom kan man nästan argumentera att vilken kurva som helst liknar ett S, eftersom vi möjligen bara ser ett fragment av kurvan, och eftersom en S-kurva kan se ut på olika sätt, då den kan vara avbruten eller med tak (och golv) på olika nivåer. Observera också att vi använder S-kurvan snarare som ett deskriptivt begrepp, när det finns teorier om varför kurvan uppstår (Labov 1972; Blythe & Croft 2012). Man kan i princip testa dessa teorier genom att analysera mer noggrant hur förändringen sprids i samhället.

Vi vill avsluta med två huvudpoänger. För det första hoppas vi ha visat att storskaliga korpusstudier ger forskare verktyg som kan användas för att testa, komplettera och förtydliga intuitioner och påståenden om språkförändring i nutida svenska (jämför påståendena i 1–3). För det andra anser vi att verktygen inte räcker till för att göra välmotiverade uttalanden om framtiden. Det är lockande att säga utifrån figur 2 att andelen *dom* kommer att fortsätta att minska, men om vi hade tittat på Flashbacks linje i grafen 2002, hade vi då kunnat gissa att vi stod vid en vändpunkt och att den observerade ökningen skulle avslutas? Vi påstår inte att spådomarna (10, 18) är felaktiga. Vi påstår istället att vi inte kan veta detta, och att språkvetenskapen i dag saknar verktyg för att rigoröst utvärdera påståenden om framtiden. Projektet Cassandras huvudsyfte är just att börja utarbeta sådana verktyg genom att noggrant analysera sociala och språkliga faktorer i flera olika fall av språkförändring, samt testa om förändringens senare fas kan förutsägas utifrån tidigare faser.

Efterord

Projektet Cassandra finansieras av Marcus och Amalia Wallenbergs Minnesfond (MAW 2020.0060) 2021–2024. Författarna har alla bidragit till artikelns koncept, dataanalys och text. EC har även stått för litteraturstudien och en större del av skrivandet. AB har extraherat och visualiserat data.

Litteratur

Andersson, Lars-Gunnar. 2017. Kommer ”att” vara kvar? Göteborgs-Posten, 2017-01-21.

Blensenius, Kristian & Lena Rogström. Att hantera grammatisk förändring i en deskriptiv ordbok. Nordiska studier i lexikografi 15. S. 81–90.

Blythe, Richard & William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language* 88. S. 269–304.

Borin, Lars, Markus Forsberg & Johan Roxendal. 2012. [Korp – the corpus infrastructure of Språkbanken](#). Proceedings of LREC 2012. Istanbul: ELRA. S. 474–478.

Artikel inlämnad för granskning den 31 oktober 2022.

Cassirer, Peter. 1976. Formerna av pronomenet *de* i södra Sverige. *Nysvenska studier* 55–56. S. 251–290.

Delsing, Lars-Olof. 1993. *Kommer utan att*. I: *Flyktförsök: kalasbok till Christer Platzack på femtioårsdagen 18 november 1993 från doktorander och dylika Institutionen för nordiska språk*. Lunds universitet.

Ewerth, Sten. 1978. Ska dom eller skall de? *Språkvård* 1978(2). S. 25–30.

Ewerth, Sten. 1992. Tidningstexter, om *dom* igen. *Språkvård* 1992(3). S. 17–19.

Ewerth, Sten. 2008. Dom skriver fortfarande dem. *Språktidningen* december 2008.

Hammarstedt, Martin, Anne Schumacher, Lars Borin & Markus Forsberg. 2022. [Sparv 5 User Manual](#). Teknisk rapport.

Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3). S. 199–244.

Labov, William. 1972. *Sociolinguistic patterns*. University of Pennsylvania Press.

Ledin, Per & Benjamin Lyngfelt. 2013. Olika *hen*-syn. Om bruket av *hen* i bloggar, tidningstexter och studentuppsatser. *Språk och stil* 23. S. 141–174.

Lind Palicki, Lena & Susanna Karlsson. 2020. Till doms över dom. Och över dom andra. En språkideologisk analys av en språkvårdsdebatt. *HumaNetten* 44. S. 175–206.

Lövestam, Sara. 2022. ”Dom” slår snart ut ”de” och ”dem”. *Svenska Dagbladet*, 2022-02-22.

Malmgren, Sven-Göran. 2017. Hur upplever du *hen*? Nio lexikala resor från 1965 till 2015. 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. S. 19–35.

Milles, Karin, Karin Salmson & Marie Tomicic. 2012. Det behövs ett nytt ord i svenska språket. *Svenska Dagbladet*, 2012-01-20.

Milles, Karin. 2013. En öppning i en sluten ordklass? Den nya användningen av pronomenet *hen*. *Språk och stil* 23. S. 107–140.

Nevalainen, Terttu & Helena Raumolin-Brunberg. 2016. *Historical sociolinguistics: language change in Tudor and Stuart England*. Taylor & Francis.

Nevalainen, Terttu. 2015. Descriptive adequacy of the S-curve model in diachronic studies of language change. I: Sanchez-Stockhammer, Christina (red.): *Can We Predict Linguistic Change? Studies in Variation, Contacts and Change in English* 16. <https://varieng.helsinki.fi/series/volumes/16/nevalainen/>

Olofsson, Arne. 2008. Framtid i förändring. Hur länge kommer *att* dröja sig kvar? *Språk och stil* 18, 143–155.

Artikel inlämnad för granskning den 31 oktober 2022.

Svenska Språknämnden. 2015. Språkriktighetsboken. Stockholm: Nordstedts.

Svensson, Anders. 2021. Vad hände med ”hen” egentligen? Dagens Nyheter, 2021-01-24.

Wiktorsson, Maria. 2018. How hybrid is blog data? A comparison between speech, writing and blog data in Swedish. *Nordic Journal of Linguistics* 41(3). S. 367–377.