

Sprog i Norden

Titel: Dansk sprognormering nu og i fremtiden
Forfatter: Dorthe Duncker
Kilde: Sprog i Norden, 2010, s. 19-32
URL: <http://ojs.statsbiblioteket.dk/index.php/sin/issue/archive>



© Dansk Sprognævn

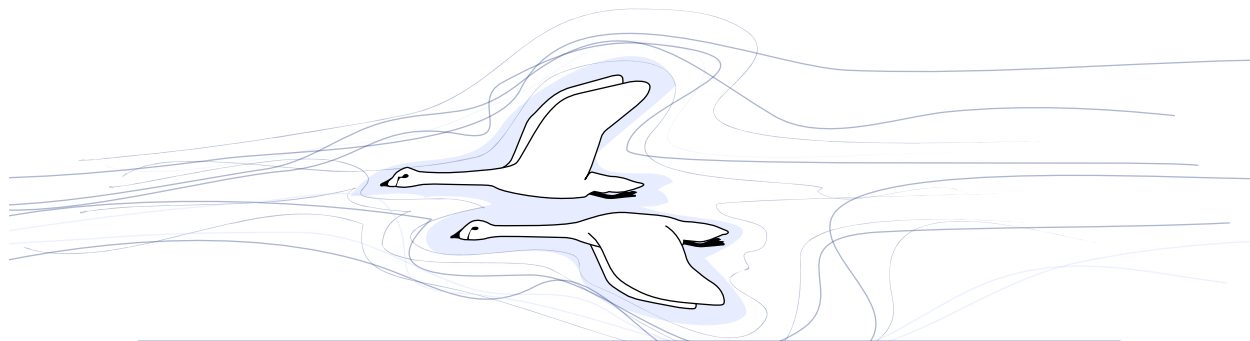
Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre numre af Sprog i Norden (1970-2004) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.



Dansk sprognormering nu og i fremtiden

Dorthe Duncker

Normeringsituationen for dansk skriftsprog er under forandring i disse år på grund af udviklingen på internettet og inden for korpuslingvistikken. Begge dele har konsekvenser for de principper som retskrivningen baseres på. Udviklingen på internettet har ført nye skribentmønstre med sig. Det nye er at lægpersoner kan læse massetilgængelige offentliggjorte tekster skrevet af andre lægpersoner, og at de gør det i stigende omfang. Dette medfører en anderledes dynamik i skriftsprogsnormen, og da Dansk Sprognævn har til opgave at følge udviklingen i dansk, må konsekvenserne af de ændrede skribentmønstre undersøges løbende. For at kunne følge med i hvordan skriftnormen praktiseres også i de digitale medier er der behov for et monitoreringskorpus hvor den enkelte teksts ortografiske status er kendt.

En stadig større del af de tekster som offentliggøres og læses på internettet, er skrevet af ikkeprofessionelle skribenter, og tekster af professionelle og lægfolk findes ofte mellem hinanden (fx blogs, nyhedsbreve og debatforummer). I det hele taget er flere og flere af de tekster der produceres og læses, digitale. Nye tekster publiceres på internettet, eller de publiceres samtidigt på flere platforme, både i trykte og digitale medier (From Jacobsen 2005). Udviklingen inden for tekstarter som kun produceres på nettet, er i eksplosiv stigning. På verdensplan fordobles antallet af blogs hvert halve år,¹ og en tilsvarende udvikling kan iagttages i Danmark. Ved udgangen af 2008 var der ca. 130.000 danske blogs,² hertil kommer en række gratis blogplatforme hos danske medier (bl.a. Urban-Blog, TV2 Blog, Avisen.dk) plus udenlandske blogtjenester som Google og Wordpress.com. I modsætning til endnu nyere medier som Facebook og ten-

¹ Jf. <http://www.sifry.com/stateoftheliveweb/>.

² Jf. <http://blog.overskrift.dk/2008/12/30/den-danske-blogosf%C3%A6re-ved-udgangen-af-2008/>

denser i retning af mikroblogging (fx Twitter hvor et indlæg ikke må overskride 140 tegn), så udmærker den klassiske blog sig ved generelt længere indlæg.

Denne udvikling medfører en anderledes dynamik i skriftsprogsnormen, og for Dansk Sprognævn, som har til opgave at følge udviklingen i dansk, hører det med at ændringen i sammensætningen af offentligt tilgængelige tekster medtages i normeringens empiriske grundlag.

Før internettet var den ikkeprofessionelle skriftbårne kommunikation ikke masse- eller netværkskommunikation, men kommunikation i en til en-relationer. Massekommunikation foregik tidligere i en til mange-relationer hvor én professionel instans producerede tekst trykt på håndgribelige tekstbærere (bøger, aviser, magasiner) som blev læst af mange. I forhold til skriftnormen medfører fremkomsten af de sociale medier på internettet at almindelige mennesker som ikke er professionelle skribenter, nu i hidtil ukendt omfang kan læse tekster skrevet af andre almindelige mennesker og professionelle i mange til mange-relationer. Både skribentmønsteret og skriftmønsteret kompliceres herved og medfører at det fremover bliver umuligt at opretholde illusionen om at der findes én norm for dansk skriftsprog.

Skriftlighed og kommunikationsteknologier

Historien viser at hver gang en ny kommunikationsteknologi er blevet indført, har det haft afgørende konsekvenser for skriftligheden. I 1500-tallet betød tryk-
keteknologien at man ikke længere skulle håndkopiere bøger i ét eksemplar ad gangen, men at mange identiske kopier kunne distribueres til et endnu større antal læsere. Denne udvikling afgjorde skriftnormens stabilisering og undertrykte den samtidige håndskriftnorm som hurtigt blev overdøvet af tryknormen (Duncker 2001).

Efter indførelsen af den almindelige skolepligt i 1700- og 1800-tallet (med skolelovene i 1720, 1740 og 1814) steg antallet af folk som kunne læse og skrive dansk, og privatpersoner udvekslede indbyrdes håndskrevne breve frem til den næste kommunikationsteknologi, telefonen, som fra de første årtier af det 20. århundrede og frem bevirkede en støt nedgang i den private korrespondance idet distancekommunikation nu kunne klares og blev klaret gennem telefonsamtaler. Helt frem til 1990'erne var skriftbilledet domineret af trykte medier, aviser, bøger og magasiner, skrevet og redigeret af professionelle skribenter, og ikkeprofessionelle sprogbrugere havde groft sagt kun adgang til hinandens huskesedler, julekort og opslag i det lokale supermarked.

Fra midten af 1990'erne ændres dette billede. Internettet, der fra begyndelsen af 1980'erne havde været et ret lukket netværk mellem computere fra for-

skellige forskningsinstitutioner, vokser voldsomt og får i løbet af 1990'erne mange brugere uden for forskningsmiljøerne, og i samme takt får e-post større udbredelse. Samtidigt bliver biltelefoner til mobiltelefoner, og man kan sende små tekster, sms'er, til hinanden via telefonen. Udbredelsen af både e-post og sms betyder at almindelige mennesker begynder at skrive til hinanden igen, ikke på papir denne gang, men på computer og mobiltelefon. Der produceres meget mere tekst, men denne del af skriftkommunikationen foregår fortsat, ligesom det gamle privatbrev, i en til en-relationer (bortset fra muligheden for gruppe-sms og mail til flere modtagere). Indflydelsen på skriftnormen i bred forstand er tilsvarende begrænset. Til gengæld opstår der hurtigt lokale kommunikationsnormer og skriftkonventioner i tilknytning til de nye medier, bl.a. særlige forkortede stoveformer og humørikoner (Laursen 2005; Hjarvard 2004, 2008).

I 1996 brugte 12% af danskerne internet i gennemsnit to timer om ugen. I 2001 havde 54% af den danske befolkning ifølge Danmarks Statistik adgang til internettet fra hjemmet, i 2009 var det 86%.³

Fra omkring 2005 tager internettet en ny drejning som reducerer graden af envejskommunikation fra én hjemmesidejer til mange brugere. Sociale medier hvor ligemænd kommunikerer indbyrdes (peer-to-peer), vinder frem i stadig større udstrækning, og ikkeprofessionelle skribenter kan publicere og læse hinandens tekster uden nogen central redaktionel infrastruktur. Denne drejning får i høj grad konsekvenser for udviklingen inden for skriftsprogsnormen, og den stiller ændrede krav til overvågningen af denne udvikling.

Det er her væsentligt at bemærke at dette ikke handler om – eller i hvert fald ikke primært handler om – en eventuel påvirkning fra de sproglige særtræk som nogle af de nye medier har ført med sig (medialekter, jf. Hjarvard 2004), men om praksis inden for ganske almindelig normalprosa, blot under ændrede produktionsvilkår.

Det nye er at en skribentgruppe hvis praksis ikke tidligere har indgået i grundlaget for den officielle danske ortografi, nu har fået direkte adgang til det offentlige rum og er blevet deltagere i netværkskommunikationsprocessen.

Alle aldersgrupper deltager i denne proces, men det er ikke overraskende de yngre årgange som fører an. Til gengæld er der tale om et demokratisk projekt. Den uddannelsesmæssige baggrund spiller ikke nogen stor rolle, og det gør den demografiske fordeling heller ikke. Der er en markant stigning i antallet af brugere som har læst eller selv skrevet blogindlæg fra 2008 til 2009, og alt taler

³ Jf. <http://www.statistikbanken.dk>, FABRIT1 og BEBRIT01.

for at denne tendens vil fortsætte. Ifølge Danmarks Statistik læste 22% af befolkningen i 2008 blogs og 9% skrev blogindlæg; i 2009 var tallene hhv. 33% og 18%.⁴

Dansk Sprognævnets opgaver

I forhold til arbejdet med skriftsprognormen repræsenterer udviklingen på internettet med fremkomsten af nye skribentmønstre en stor udfordring. Af lovgrundlaget for Dansk Sprognævns arbejde fremgår det at:

Dansk Sprognævn har til opgave at følge det danske sprogs udvikling, at give råd og oplysninger om det danske sprog og at fastlægge den danske retskrivning. (*Lov om Dansk Sprognævn*, § 1, stk. 1)

Dansk Sprognævn skal arbejde på videnskabeligt grundlag. I sit arbejde skal nævnet tage hensyn til sprogets funktion som bærer af tradition og kulturel kontinuitet og som spejl af samtidens kultur og samfundsforhold. (*Lov om Dansk Sprognævn*, § 1, stk. 3)

Den retskrivning, der fastlægges af Dansk Sprognævn, og som offentliggøres i den officielle danske retskrivningsordbog, skal fastlægges i overensstemmelse med de i stk. 2-4 nævnte hovedprincipper for retskrivningen.

Stk. 2. Hovedprincipperne for dansk retskrivning er traditionsprincippet og sprogbrugsprincippet.

Stk. 3. Efter traditionsprincippet skrives ord og ordformer i dansk i overensstemmelse med den praksis, der har været gældende siden bekendtgørelse nr. 24 af 27. februar 1892, og som siden 1955 er kommet til udtryk i de retskrivningsordbøger, som Dansk Sprognævn har udgivet. Efter traditionsprincippet ligger stavemåderne af det eksisterende ordforråd principielt fast, bortset fra justeringer som følge af sprogbrugsprincippet. [...]

Stk. 4. Efter sprogbrugsprincippet skrives ord og ordformer i dansk i overensstemmelse med den praksis, som følges i gode og sikre sprogbrugerers skriftlige sprogbrug. (*Bekendtgørelse om Dansk Sprognævns virksomhed og sammensætning*, § 1).

⁴ Jf. <http://www.statistikbanken.dk>, BEBRIT04.

Dansk Sprognævn har som sin primære kerneopgave at følge udviklingen i dansk. Nævnet skal arbejde på videnskabeligt grundlag og må derfor gå empirisk til værks. Det rejser to spørgsmål: Hvor og i hvilken form findes empirisk basis i form af spor efter frembragt dansk? Og hvordan overvåger man i praksis forandringer i denne basis systematisk?

Normeringens principper og empiriske basis

Hvor den empiriske basis har skullet søges, har hidtil været afgjort med dens form, nemlig som publicerede trykte tekster. Den genremæssige spredning og relative fordeling af kilderne har varieret noget gennem tiden. Dansk Sprognævn begyndte i 1955 at opbygge sit ordkartotek ved systematisk at indsamle nye ord, ordforbindelser og ordanvendelser. Gennem hele perioden frem til i dag har aviserne udgjort den største kilde, i gennemsnit næsten 60%, mens bøger, ugeblade og andre kilder hver bidrager med en mindre andel. Først fra 2008 indgår internetkilder og indtil videre kun med 1%. – Se Jarvad og Halskov 2010 for en detaljeret redegørelse af kildefordelingen.

Frem til 2004 harmonerer denne fordeling med overvægten af avistekster ret godt med befolkningens læsevaner,⁵ men den afspejler ikke den nyeste udvikling i brugen af internettets medier. Der foreligger ikke aktuelle danske undersøgelser (ud over den ovenfor nævnte som viser en markant stigning i antallet af danskere som læser eller skriver blogs), men en nordisk undersøgelse fra 2007 viser at befolkningerne i Norge, Sverige og Finland bruger dobbelt så lang tid på internettet som de bruger på at læse avis.⁶

Forklaringen på kildegrundlagets sammensætning i nævnets excerpering skal blandt andet søges i de to hovedprincipper for dansk retskrivning: traditionsprincippet og sprogbrugsprincippet (jf. uddragene af loven og betænkningen ovenfor). Hvis traditionen blev fulgt til punkt og prikke, var stavemåde og ordforråd statisk, og dermed ville der heller ikke være nogen sproglig udvikling at følge. Forandringen kommer via sprogbrugsprincippet, men traditionsprincippet ligger implicit i sprogbrugsprincippet, idet der her kun skal tages hensyn til den ortografiske praksis "som følges i gode og sikre sprogbrugerens skriftlige sprogbrug". En god og sikker sprogbruger er hermed en skribent som holder traditionsprincippet i hævd. Og denne implikation sætter uvilkårligt en effektiv bremse på forandringstakten.

Gruppen af privilegerede empirileverandører efter sprogbrugsprincippet er overvejende blevet sat lig med kildens type, med velrenommerede publikati-

⁵ Jf. <http://www.statistikbanken.dk, KVVU11, 2004>.

⁶ Jf. http://www.nordicom.gu.se/common/stat_xls/1240_10510_media_time_shares_2007.xls.

onskanaler med redaktørordning hvor personer der ernærer sig ved deres pen, får mangfoldiggjort deres tekster. Dén antagelse holder ikke længere. Dels er den redaktionelle infrastruktur svækket idet arbejdet med teksterne på mange dagblade foregår under stigende tidspres med samtidig publicering til flere medier. Dels kan alle som lyster, nu komme uredigeret til tastaturet på internettet – og blive læst. Professionelle og amatører mellem hinanden. Der er dermed brug for en empirisk valid operationalisering af sprogbrugsprincippet under hensyntagen til de ændrede vilkår.

Traditionsprincippet og sprogbrugsprincippet

Traditionsprincippet er i praksis en samlebetegnelse for en række andre principper som bruges i normeringsarbejdet. Udredningen *Sådan staver vi* (Jervelund 2008) udfolder ti principper som er indeholdt i traditionsprincippet: det fonematiske princip, fordanskningsprincippet, det internationale princip, det nordiske princip, tegnkonstansprincippet, det etymologiske princip, det distinktive princip, det dekorative princip, det pædagogiske princip og systematikprincippet. Principperne informeres fra to sider: internt fra en systemisk betragtning af retskrivningsnormen og eksternt via sprogbrugsprincippet. Ændringer i stavningen og/eller *Retskrivningsordbogens* lemmainventar kan udgå alene fra et norminternt ønske om at skabe eller opretholde konsekvens i ordbogens oplysninger i overensstemmelse med et eller flere af principperne. Men ændringer kan også komme udefra, fra tekster skrevet på dansk, og i så fald bringes sprogbrugsprincippet i anvendelse.

Sprogbrugsprincippet er som de øvrige principper i sidste ende underordnet traditionsprincippet, men der gælder det særlige for dette princip at en sprogbrugsform kun i realiteten kodificeres såfremt 1) den pågældende (dobbelt) form ikke kolliderer med de øvrige principper, og 2) formen er udtryk for ”en bred og demokratisk uvidenhed der går på tværs af alder, uddannelse og erhverv” (Hansen 1975, s. 94). Den sidste betingelse går under navnet ”Den hemmelige præmis” og er formuleret af Erik Hansen i hans normeringskanoniske artikel af samme navn.

Fra ordsamling til ordtrawler

Spørgsmålet om hvordan man i praksis systematisk kan overvåge forandringer i den empiriske basis, kan besvares bagudrettet og fremadrettet. Før pc’ens tid blev den sproglige udvikling fulgt ved at nævnets sprogforskere læste kilderne og manuelt registrerede ”nye ord, nye betydninger, nye bøjningsformer, nye

stavemåder osv.” (Galberg Jacobsen 2005, s. 93). Teksterne med de nye former blev indsamlet og opbevaret i nævnets omfattende seddelsamling. Den observerede form blev indstregt i teksten og skrevet på en seddel med oplysning om findested og dato, og sedlen med tekst blev anbragt på alfabetisk plads efter den nye form. Senere hen blev registreringen foretaget elektronisk, og med udgangen af 2009 er hele samlingen blevet digitaliseret (Kirchmeier-Andersen 2009). I samme tidsrum er en væsentlig del af kilderne blevet tilgængelig i elektronisk form, navnlig i kraft af den store avisdatabase Infomedia.

Denne ordsamling indeholder belæg på usædvanlige forekomster, men er ikke indsamlet med henblik på en dækning af den helt almindelige praktisering af skriftsprogsnormen eller af det relative forhold mellem det almindelige og det usædvanlige. Og med den dagligt voksende tekstmængde på internettet er det heller ikke menneskeligt muligt at holde øje med bare det usædvanlige.

Heldigvis er udviklingen inden for korpuslingvistikken og sprogteknologien fulgt med den øvrige teknologiske udvikling, og det er derfor blevet muligt at automatisere en del af det arbejde som tidligere fordrede menneskelige læsere. I dag høster man således automatisk tekster fra avisdatabasen og fra internettet ved hjælp af et særligt søgeprogram, en ordtrawler, og gennemsøger teksterne systematisk for typiske nyordsindikatorer (fx citationstegn, kursiv skrift, ”såkaldt”), for mulige påvirkninger fra engelsk, ligesom søgeloggene fra netudgaverne af *Retskrivningsordbogen*⁷ og *Nye ord i dansk*⁸ leverer data om nyordskandidater (Halskov 2008, dias 4). Ordtrawleren returnerer lister med nyordskandidater som derefter evalueres manuelt (Halskov & Jarvad 2010).

Fremover vil registreringen af nye ord og tekster primært foregå elektronisk, men der er en række forhold man må tage højde for i forbindelse med automatiserede søgninger og indsamling af digitale tekster. For det første kan maskiner ikke læse og forstå tekster, det kan kun personer. For det andet kan et søgeprogram ikke umiddelbart se på en vilkårlig hjemmeside, blog eller en avisartikel for den sags skyld, om den er skrevet af en god og sikker skriftsprogsbruger i bekendtgørelsens forstand, og derfor kan dens indhold ikke umiddelbart indlemmes i normens empiriske basis da man ikke kender tekstens ortografiske status. Man kan godt søge og finde blandt andet nydannelser efter bestemte principper automatisk, men den omgivende teksts status er uklar. Nye ord som ikke samforekommer med typiske indikatorer, nye betydninger, nye bøjningsformer, nye stavemåder osv. skal fortsat kunne registreres, og til dette arbejde må både korpuslingvistiske, sprogteknologiske og manuelle metoder bringes i

⁷ <http://www.retskrivningsordbog.dk/ro/ro.htm>.

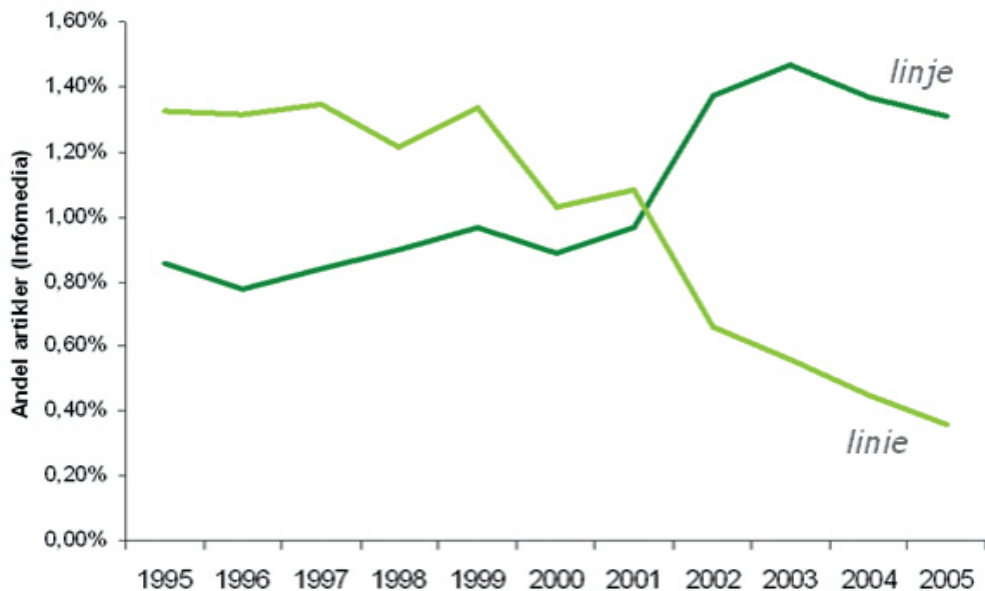
⁸ <http://nyeordidansk.dk/noid/noid.htm>.

anvendelse. De overvældende, men lettilgængelige digitale tekstmængder inviterer endvidere til at lade sprogbugsprincippet komme til udtryk gennem et frekvensprincip til støtte for et eller flere af de øvrige normeringsprincipper, med kvantitative argumenter for eller imod bestemte fænomeners indlemmelse i den officielle norm. Imidlertid udnytter man kun en brøkdel af den viden om skriftnormens opretholdelse og dannelse der kan hentes ud af teksterne, hvis man kun har blik for én ting ad gangen. Man får således ikke at vide hvilke fænomener der følges eller ikke følges ad, og i hvilke slags tekster det sker.

Frekvensprincippet som støtte for sprogbugsprincippet

Et eksempel på samspillet mellem frekvensprincippet og andre principper findes i begrundelsen for at afskaffe stavemåden *linie* som dobbeltform til *linje*. I *Dansk Retskrivningsordbog* (DRO), der udkom fra 1923 til 1946, var der mange dobbeltformer med *i* og *j*. Siden *Retskrivningsordbogen* af 1955 (RO 55) er dobbeltformernes antal reduceret i overensstemmelse med ændringer i udtalen (det fonematiske princip), fx ophørte valgfriheden i DRO mellem *Smedie el. Smedje* med RO 55 hvor *smedje* er anført som eneform. Hovedreglen for fordelingen af *i* og *j* i tryksvag stavelse efter konsonant har siden RO 55 været den at der skrives *j* hvis vokalen i den foregående stavelse er kort, og *i* hvor vokalen er lang (og har stød). I overensstemmelse med denne praksis gør *Retskrivningsordbogen*, 3. udgave, 2001 (RO 01) *linje* og *kastanje* med *j* til eneformer, svarende til de gængse udtaler med kort vokal: [ˈlinjə], [–ˈsdanjə]. Og efter samme praksis er *konkylie* med *i* gjort til eneform, svarende til den gængse udtale med lang vokal (stød): [-HkyZljE]” (Schack 2001, s. 14). Konsekvenserne af ændringen i stavningen af *kastanje* og *konkylie* er begrænsede da ingen af ordene indgår i andre opslagsord i RO 01. Anderledes forholder det sig med *linje* som indgår i næsten 50 sammensatte eller afledte lemmaformer, bl.a. *linjevogter*, *skudlinje*, *linjere*. Denne ændring har altså omfattende konsekvenser, og for at hente støtte til den aktiveres sprogbugsprincippet i form af frekvensprincippet:

”At billedet nu er ved at vende [til fordel for formen med *j*], kan man forvisse sig om ved at søge i danske avisers elektroniske tekstarkiver på internettet. Enkelte aviser har ganske vist stadig en vis forkærlighed for formen med *i*. [...] I nogle aviser er fordelingen af de to staveformer stort set fifty-fifty. [...] I mange aviser står formen med *j* imidlertid stærkt.” (Schack 2001, s. 14).



Figur 1 Søgning i avisdatabasen Infomedia viser at formen *linie* forekommer i flest artikler før 2001, og at formen *linje* efter 2001 er den mest brugte. Før 2001 var *linie* og *linje* sideordnede ortografiske varianter, men med *Retskrivningsordbogen*, 3. udg., 2001 blev *linje* eneform. Diagrammet viser det kvantitative størrelsesforhold mellem de to formers forekomster, men afslører ikke formernes ortografiske kontekster. Man kan således ikke se om fx *linie* før 2001 fortrinsvis forekommer i tekster som i øvrigt følger den gældende retskrivning, men efter 2001 snarere optræder i tekster hvis ortografiske praksis i øvrigt er afvigende. For at kunne besvare dette og lignende spørgsmål er det nødvendigt at kende de enkelte teksters ortografiske status.

I 2001 blev søgningerne foretaget i de enkelte avisers egne onlinearkiver, men siden 2002 har man kunnet søge i flere aviser på en gang med Infomedia. En Infomedia-søgning for perioden 1995-2005 på formen med hhv. *i* og *j* viser at retskrivningsændringen havde en perfekt timing i forhold til praksis idet *linje* med *j* klart kommer i overtal fra 2001 (jf. figur 1).

Frekvensprincippet kan således bruges til at generalisere over og bekræfte en fornemmelse i forhold til sprogbrugen. Men hvor overbevisende et sådant søgeresultat end kan forekomme for en umiddelbar betragtning, så er det samtidig mangelfuldt. Resultatet gælder kun netop formerne *linie* og *linje*, ikke hele lemmaet inklusive bøjningsformer, og det omfatter heller ikke sammensætninger og afledninger, og deres bøjede former. Det ville også være omfattende og urimeligt tidskrævende manuelt at skulle fremsøge alle disse former en ad gangen. Desuden får man ikke gennem selv nok så mange kvantitative opgørelser af denne art noget at vide om hvordan skriftnormen i øvrigt praktiseres i tekster

med den ene eller den anden ortografiske variant, eller om denne praksis ændrer sig i løbet af observationsperioden. Man kan ikke aflæse noget om den ortografiske kontekst varianterne indgår i, fx om *linie* før 2001 fortrinsvis forekommer i tekster som i øvrigt følger traditionsprincippet, men efter 2001 snarere forekommer i tekster som i øvrigt afviger fra dette princip. Selv om teksterne er tilgængelige i deres helhed, resulterer sådanne søgninger efter enkeltord ikke i noget helhedsbillede af normens praktisering. Og endelig er den genremæssige dækning aldeles ensidig hvis der kun søges i avistekster.

Korpus til monitorering af dynamikken i dansk skriftsprogs norm

Med de eksplosivt voksende digitale tekstmængder bliver den kerneopgave som er pålagt Dansk Sprognævn, at følge dansk sprogudvikling, stadig tungere, og det bliver fremover helt nødvendigt at anvende automatiske metoder eksplorativt for at løfte den. Der er her brug for et monitoreringskorpus (monitor corpus) repræsentativt sammensat mht. til kilder og genrer under hensyntagen til de nye digitale skribentmønstre. Korpusset skal være åbent i den forstand at det løbende opdateres med nye tekster, men sådan at man hele tiden nøje ved hvad det er for en tekstmængde man aktuelt undersøger.

Et monitoreringskorpus som tager højde for teksterne i deres helhed, vil være en værdifuld støtte til arbejdet med at overvåge sprogudviklingen, og det vil i sig selv udgøre en væsentlig forskningsresurse. Ved at etablere et sådant korpus får man en velbeskrevet basis til at kvalificere øvrige søgeresultater på, og man kan undgå de bivirkninger som frekvensprincippet fører med sig, idet man kan fokusere på distributionen og samforekomsten af træk på tværs af teksterne og dermed i stedet operere med et udbredelsesprincip. Men først og fremmest opnår man indsigt i hvordan den stabile del af skriftsprogsnormen er konventionaliseret, og hvilke mekanismer der er virksomme i dens forandring.

For at kunne uddrage viden om en bestemt forms relation til skriftnormen må den tekst formen forekommer i, som helhed kunne relateres til normen. Det kan ske ved at klassificere den enkelte tekst i forhold til et referencekorpus som repræsenterer traditionsprincippet i overensstemmelse med den gældende retskrivning (ortografisk guldstandard). Herved opnår man en eksplicit kvalificering af sprogbrugsprincippet, og idet den enkelte tekst i kraft af relationen til referencekorpusset er placeret på en skala mellem en retskrivningsnær og en retskrivningsfjern pol, kender man også dens ortografiske status. Ved at søge efter bestemte afvigelsestyper eller konkurrerende former i et korpus organiseret på denne måde får man ikke blot oplysning om typens forekomsthypig-

hed, men også om dens relation til traditionsprincippet, om den fortrinsvis forekommer i tekster hvis ortografiske praksis i øvrigt ligger nær på eller fjernt fra retskrivningsnormen. Selve klassifikationsprocessen kan udnyttes til at spore aktuelle ortografiske tendenser, og den vil desuden give overblik over hvor stor del af et repræsentativt udsnit af tekster skrevet på dansk der på et givet tidspunkt faktisk følger traditionsprincippet. Hvis andelen af tekster der overholder traditionsprincippet, daler over tid, dvs. hvis flere og flere af de høstede tekster udviser en praksis som afviger fra referencekorpussets, vil det være udtryk for at skriftnormen er i bevægelse.

Ifølge bekendtgørelsen kan ortografiske justeringer som følge af sprogbrugsprincippet kun foretages på grundlag af tekster som ellers overvejende følger traditionsprincippet. Ved hjælp af et klassificeret monitoreringskorpus vil man på et empirisk velfunderet grundlag kunne etablere et skopus for sprogbrugsprincippet og sondre mellem stavemåder og ordformer som forekommer i tekster der er klassificeret som retskrivningsnære, og former som nok findes i skriftnormen, men ikke i den del af den som skal fastlægges i den officielle retskrivning.

Foruden den viden som kan hentes ud af selve klassifikationen, vil denne fremgangsmåde også afkaste værdifulde data om den helt upåfaldende praktisering af skriftnormen – som alt andet lige er den mest udbredte, og som udgør den nødvendige baggrund for at blive opmærksom på det usædvanlige.

Den automatiske analyse af de høstede tekster og af teksterne i referencekorpusset kan i vid udstrækning foretages med sprogteknologiske værktøjer og metoder som allerede er udviklet. Der findes et bredt udvalg af metoder til morfologisk analyse, opmærkning af navne og kontekstuelle og fraseologiske analyser som sammen med andre tekstlige analyser kan bruges til at opstille en profil af målepunkter på grundlag af referencekorpusset og til at analysere og opmærke de nye tekster så de kan indplaceres i forhold til referencekorpusset. Nogle af metoderne er velafprøvede og kan umiddelbart implementeres, fx morfologisk analyse, mens andre endnu er i en udviklingsfase eller ikke har været afprøvet på større mængder dansk tekst; det gælder fx de kontekstuelle og fraseologiske analyser hvor der er flest erfaringer fra engelsksprogede korpusser og i mindre grad fra danske.

Litteratur

Duncker, Dorthe, 2001: *Dynamikken i ortografien*. I: Peter Widell og Mette Kunøe (red.) 8. *Møde om Udforskningen af Dansk Sprog*, s. 73-82. Århus: Aarhus Universitet.

- From Jacobsen, Peter, 2005: *Fagpressens flermediale fremtid. En rapport om muligheder og udfordringer ved udgivelse i flere mediekkanaler*. Århus: Center for Journalistik og Efteruddannelse.
- Galberg Jacobsen, Henrik, 2005: *På sproglig grund. Dansk Sprognævn fra a til å 1955-2005*. Odense: Syddansk Universitetsforlag.
- Halskov, Jakob, 2008: It-støttet excerpering og registrering af nye ord og ordforbindelser. Foredrag d. 30. oktober 2008 i Selskab for Nordisk Filologi, København.
- Halskov, Jakob og Pia Jarvad, 2010: Human og maskinel excerpering af neologismer. I: *NyS, Nydanske Studier*, bd. 38. s. 39-68.
- Hansen, Erik, 1975: Den hemmelige præmis. Argumenter og skinargumenter i sprogrigtighedsspørgsmål. I: Klaus Kjøller (red.) *Analysen af sprogbrug. Argumentation*, s. 88-102. København: Borgen.
- Hjarvard, Stig, 2004: The Globalization of Language. How the Media Contribute to the Spread of English and the Emergence of Medialects. *Nordicom Review* 2004, no. 1-2, pp. 75-97.
- Hjarvard, Stig, 2008: *En verden af medier. Medialiseringen af politik, sprog, religion og leg*. København: Samfundslitteratur.
- Jervelund, Anita Ågerup, 2008: *Sådan staver vi – om ortografi og stavefejl*. København: Dansk lærerforeningens Forlag og Dansk Sprognævn.
- Kirchmeier-Andersen, Sabine, 2009: Dansk Sprognævn – Fremtidens udfordringer og planer. I: Dorthe Duncker, Anne Mette Hansen & Pia Quist (red.) *Studier i Nordisk 2006-2007*, s. 113-127. København: Selskab for Nordisk Filologi.
- Laursen, Ditte, 2005: Svar udbedes! Besvarelsenormen i unges sms-kommunikation. I: Peter Widell og Mette Kunøe (red.) *10. Møde om Udforskningen af Dansk Sprog*, s. 238-256. Århus: Aarhus Universitet.
- Schack, Jørgen, 2001: Ændringer i opslagsords staveform eller ordform. *Nyt fra Sprognævnet* 2001/4, s. 12-17.

Summary

The development of the Internet and of corpus linguistics is changing the conditions under which an orthographic norm for Danish can be maintained. Both

developments affect the principles upon which the orthographic norm is based. The development of the Internet, for example, has introduced new patterns of authorship. The novelty being that non-experts now read texts written by other non-experts, and that these texts are published and disseminated. These changes are making the written language norm more dynamic, and since it is a key objective of the Danish Language Council to monitor developments in the Danish language, the consequences of the new patterns of authorship must be analyzed continuously. In order to track changes in the written language norm of new digital media, one must compile a monitor corpus in which the orthographical status of the individual texts is known.