

Nyt fra Sprognævnet

UDGIVET AF DANSK SPROGNÆVN

NR. 4 · MARTS 1970

Hvor hyppigt optræder de forskellige bogstaver?

Hvilke bogstaver bruges mest i dansk? Er bogstaverne lige hyppige i forskellige slags tekster?

Spørgsmål af denne art rettes jævnligt til Dansk Sprognævn, i forbindelse med så forskellige formål som udformning af datamaskintastaturer og udarbejdelse af materiale for læsehæmmede. Men mens der i mange år har foreligget omfattende undersøgelser over ords hyppighed i dansk, foretaget med særligt henblik på skolernes retskrivningsundervisning, har det tilsyneladende simple spørgsmål om bogstavhyppigheder i danske tekster krævet nye undersøgelser.

At der er store uligheder i bogstavernes hyppighedsfordeling viser sig bl. a. ved bogtryk; da amerikaneren Samuel Morse i 1832 opfandt prikstreg-koden til telegrafi, hvor det er vigtigt, at et hyppigt bogstav beslaglægger kortere tid end et sjældent, søgte han hos en bogtrykker oplysning om, i hvilke antal denne anskaffede nye typer til erstatning for slidte. På dette grundlag fik *e* det korteste morsetegn, *t* det næstkorteste osv.

På skrivemaskiner, hvor hvert bogstav kun findes i ét eksemplar, slides de hyppigst brugte typer mest (dog spiller også bogstavernes form en rolle for graden af slid). I øvrigt er den sædvanlige placering af bogstaverne på skrivemaskiners tastatur oprindeligt fastlagt ud fra bogstavhyppigheder i engelske tekster.

Det er en ret udbredt opfattelse, at sprogstatistiske undersøgelseres værdi afhænger af deres omfang, sådan at resultaterne skulle være desto pålideligere, jo større tællinger man foretager. Dette er imidlertid ingen almengyldig sandhed. Det viser sig blandt andet, at selv meget anvendte bogstavers hyppighed svinger kendeligt efter teksternes art: I dansk skønlitteratur er *n* gennemgående hyppigere end *r*, mens forholdet er omvendt i avisartikler og faglitteratur. Denne forskel kan ved nærmere studium

vises at hænge sammen med en større brug af nutid – med udsagnsord endende på *-r* eller *-er* – i avis- og faglitteratur.

Hvis man for at få gennemsnitshyppigheder tæller i en blanding af 2 dele skønlitteratur og 1 del avis- og faglitteratur, får man følgelig et noget andet resultat, end hvis man tæller i 1 del skønlitteratur og 2 dele avis- og faglitteratur. Dette gælder, uanset hvor stor en tekstsamling man gennemtæller; og da der ikke med rimelighed kan tales om et »naturgivet blandingsforhold« mellem skønlitteratur, faglitteratur, avisartikler osv., er det indbildning at forestille sig »sande gennemsnitshyppigheder« for bogstaver i danske tekster som helhed.

Har man brug for nøjagtige værdier for en bestemt art af tekster, vil kun tællinger i denne art tekster være nyttige. Men til en del formål kan man nøjes med alment orienterende bogstavhyppigheder, hvis talværdi må betragtes med et vist forbehold. Procenttallene i tabellen skal opfattes på denne måde.

Ordningsnummer efter faldende hyppighed	Bogstav	Hyppighed i procent	Sammenlagt hyppighedsprocent
1	E e	16,6	
2	R r	8,0	Nr. 1–2: 24,6
3	N n	7,7	
4	T t	7,2	
5	D d	6,7	Nr. 1–5: 46,2
6	I i	5,8	
7	A a	5,6	
8	S s	5,5	
9	L l	5,1	
10	O o	4,5	Nr. 1–10: 72,7
11	G g	4,4	
12	M m	3,7	
13	K k	3,2	
14	V v	2,7	
15	F f	2,6	Nr. 1–15: 89,3
16	H h	2,1	
17	U u	1,6	
18	B b	1,4	
19	P p	1,3	
20	Å å	1,3	
21	Ø ø	0,9	
22	Æ æ	0,8	
23	Y y	0,6	
24	J j	0,6	
25	C c	0,1	
		100,0 %	

W w, X x, Z z, og Q q: hver mindre end 0,02 %.

Middelordlængden er 4,8 bogstaver (dvs. der optræder et ordmellemlum gennemsnitligt efter henvend 5 bogstaver).

Det skal fremhæves, at tabellens tal gælder bogstaver, altså *skriftsprogstørrelser* og ikke *udtalestørrelser*. Hyppigheder for udtalestørrelser (sprog-lyd) vil direkte kun kunne findes ud fra tale optegnet i lydskrift, og resultatet vil i høj grad være afhængigt af den benyttede lydskrifts art. Tilsvarende hører de angivne bogstavhyppigheder naturligvis sammen med en bestemt stavemåde (retskrivning); fx forrykkede indførelsen af bogstavet *å* i 1948 alle hyppighedstal for bogstaver i danske tekster. Tabellen her svarer til nutidig stavemåde.

For det andet skal det bemærkes, at tabellens tal gælder bogstavets hyppighed i *sammenhængende tekst*, ikke fx i en ordbog, hvor hvert ord kun forekommer 1 gang. Undersøger man ordforrådet i retskrivningsordbogen, finder man en mere ligelig fordeling af bogstaverne end i tekster, hvor hvert fjerde bogstav er et *e* eller et *r*, og hvor i gennemsnit hvert andet bogstav hører til gruppen *e, r, n, t, d, i*.

Den store overvægt af enkelte bogstaver i tekster hænger sammen med to ting. Dels udgøres de meget hyppige bøjningsendelser af få forskellige bogstaver, dels er bestemte bogstaver stærkt repræsenteret i de hyppigste småord (*d* fx i *det, den, de, der, med* og flere). Det er naturligvis forskelle i brugen – og dermed hyppigheden – af bestemte bøjningsendelser og bestemte ord, der fører til, at bogstavhyppighederne i nogen grad svinger med teksternes art og emne. I tekster med direkte tale vil ordene *jeg, ja, jo, nej* være væsentligt hyppigere end ellers, og hyppigheden af det sædvanligvis ret sjældne bogstav *j* vil tilsvarende være stærkt forøget.

I en artikel om zoologiske haver med deres zebraer, zebuer, zobler, gazeller og ozelotter vil bogstavet *z* antageligt selv i dansk kunne nå op på en nævneværdig hyppighed.

Hvis man specielt er interesseret i hyppigheden af *begyndelsesbogstaver* – fx for at tilrettelægge alfabetiske lister – kan allerede sideantallet under hvert begyndelsesbogstav i retskrivningsordbogen eller en anden ordbog tjene til orientering for det almindelige ordstofs vedkommende. Tilsvarende giver en telefonbog indtryk af hyppigheden af personer, hvis efternavn begynder med et bestemt bogstav.

Bogstavhyppigheder i tekster er meget forskellige fra sprog til sprog. I engelsk er *e* hyppigst ligesom i dansk, men med et lavere procenttal og nært fulgt af *t*. Også i svensk har *e* et lavere procenttal end i dansk og konkurrerer her med *a*. I tyske tekster overgår *e*'s hyppighed derimod gennemgående hyppigheden i danske tekster, med *n* på andenpladsen.

Henning Spang-Hanssen.

Oplysninger og råd fra Sprognævnet

Et udvalg af de ca. 2000 telefoniske forespørgsler som nævnet har besvaret i sidste halvdel af 1969.

7001 *Spørgsmål:* Skal man skrive *zigeuner* eller *sigøjner*? Jeg har lært at stave ordet med *z* og *eu*, men nu ser man ofte den anden stavemåde brugt.

Svar: Efter den nugældende officielle retskrivning staves ordet med *s* og *øj*: *sigøjner*.

Men fra 1904 til 1955 gav de autoriserede retskrivningsordbøger valgfrihed mellem de to stavemåder, og indtil 1904 skulle ordet skrives *Zigeuner*. I praksis foretrak langt de fleste at skrive ordet med *z* og *eu*, og stavemåden *zigeuner* har stadig en vis udbredelse.

Vi har fået ordet via tysk, og det er herfra stavemåden med *z* og *eu* stammer. Men da ordet ikke er af tysk oprindelse, og da det på dansk udtales med *s* (og ikke med *ts*) som første lyd, er fordanskningen af stavemåden velbegrundet.

7002 *Spørgsmål:* Hvorfor er ordene *chok* og *check* behandlet forskelligt i Retskrivningsordbogen, der foreskriver *k* i *chok*, men *ck* i *check*?

Svar: Når *check* skal staves med *ck*, er det fordi ordet er lånt fra engelsk og man har valgt at følge en af de to stavemåder der anvendes i dette sprog (hvor både *check* og *cheque* forekommer, den første især i amerikansk, den sidste især i britisk engelsk).

Ordet *chok* er derimod lånt fra det franske ord *choc*. I dansk optrådte ordet i det 19. århundrede med højst forskellige stavemåder, men da det første gang blev optaget i den autoriserede retskrivningsordbog (i 1904), var det med den franske stavemåde: *Choc*. I de følgende udgaver (1909 og senere) blev der givet valgfrihed mellem *Choc* og den delvis fordanskede skrivemåde *Chok*, og fra 1918 har stavemåden med *k* været den eneste officielle. Når ordet ikke sjældent ses stavet *chock*, skyldes det vel især indflydelse fra *check*.

7003 *Spørgsmål:* Er det ikke rimeligere at skrive *can-d. psyk.* (og *stud. psyk.*) end *can-d. psych.* (og *stud. psych.*)? Vi skriver jo *psykologi*, *psykolog* osv. med *k* i moderne dansk.

Svar: De korrekte stavemåder er *can-d. psych.*, *stud. psych.* Ikke desto mindre ses stavemåden *psyk.* hyppigt; selv i officielt sprog kan den forekomme.

Akademiske titler af denne type er forkortelser af rent latinske ordforbindelser (*candidatus psychologiae, studiosus psychologiae*), ikke af de tilsvarende danske ord, og staves derfor i overensstemmelse med de latinske ords stavemåde. Jævnfør forkortelser som *phil.* (men *filosofi*), *theol.* (men *teologi*), *oecon.* (men *økonomi*), *pharm.* (men *farmaceut*), *merc.* (men *merkantil*).

Spørgsmål: Da jeg gik i skole, skrev man *Sverrig*. Nu er der flere og flere der skriver *Sverige*. Hvilken af de to former er den rigtigste? Og udtales de ens?

7004

Svar: Stednavne er ikke taget med i den retskrivningsordbog som Sprog-nævnet udgav i 1955. Det var de derimod i dens forgænger, Dansk Ret-skrivningsordbog, som gav valgfrihed mellem *Sverrig* og *Sverige*. Begge disse skrivemåder er stadig anvendelige, men det tilrådes at skrive *Sverige* da det er denne form der bruges i svensk. Denne stavemåde er efterhånden blevet langt den almindeligste i dansk, og den er den eneste der er angivet i de nordiske sprognavns liste »Stater, hovedstæder, nationalitetsbetegnel-ser« (1961).

Uanset skrivemåden er udtalen i to stavelser, altså uden et *-e* i slutningen, den almindelige i dansk.

Spørgsmål: Udtales ordet *koks* med en åben å-lyd, dvs. med samme vokal som ordet *kok*, eller med en mere lukket vo-kal, altså som *u*'et i ordet *bukser*?

7005

Svar: Begge udtaler er acceptable.

Tidligere blev udtalen med den mere lukkede selvlyd (som i *bukser*) af mange betragtet som den bedste, og denne udtale blev i 1928 anført som hovedudtalen i Ordbog over det danske Sprog. Nu er udtalen med den åbne å-lyd (som i *kok*) sikkert den almindeligste.

Spørgsmål: Om bøjningen af ordet *stencil* oplyser Retskriv-ningsordbogen kun at man kan sætte *en* foran, og at det hedder *stencils* i flertal. Kan man ikke bruge den bestemte form *stencilen* og flertalsformen *stenciler*? Og er det i så fald rigtigere at skrive disse former med to *l*'er?

7006

Svar: Når Retskrivningsordbogen ikke angiver bestemt form og kun nævner flertalsformen *stencils*, er det sikkert i overensstemmelse med sprogbrugen i begyndelsen af halvtredserne da ordbogen blev redigeret. Siden er der

sket sådanne ændringer i sprogbrugen at der ikke længere kan indvendes noget imod formerne *stencilen* og *stenciler*.

Der er endda grund til at foretrække flertalsformen *stenciler* for *stencils*, da en flertalsform på *-er* stemmer bedre med danske bøjningsmønstre (der kan fx umiddelbart dannes en bestemt flertalsform: *stencilerne*). Flere nyere ordbøger (fx Nudansk Ordbog, fra 3. udgave, 1962, og Gyldendals Fremmedordbog) nævner begge flertalsformerne.

Det tilrådes (i overensstemmelse med sprogbrugen og med de nævnte ordbøger) at skrive de bøjede former med enkelt *l*: *stencilen*, *stenciler* (jævnfør at Retskrivningsordbogen har ét *l* i *pencilen*, *penciler*).

7007

Spørgsmål: I edb-terminologien bruges det engelske ord *bit* som betegnelse for et binært ciffer, og i informationsteorien anvendes ordet som enhed for informationsmængden.

Hvordan bøjes ordet på dansk?

Svar: Ordets bøjning ligger endnu ikke helt fast i dansk. Den eneste fremmedordbog der har optaget ordet (Gyldendals Fremmedordbog), angiver flertalsformen *bits*, men flere af de fagfolk der har beskæftiget sig med databehandlings terminologi, er gået ind for at bruge en uændret form i flertal: *flere bit*.

Følgende bøjning tilrådes: *en bit*, *bitten*, *flere bit*, *bittene*.

I det forslag til dansk terminologi på edb-området som Dansk Standardiseringsråd offentliggjorde i oktober 1969, er flertalsformen *bit* anvendt.

7008

Spørgsmål: Er *hygsom* et anerkendt ord? Jeg har ikke kunnet finde det i Retskrivningsordbogen.

Svar: Ordet er ikke med i Retskrivningsordbogen og findes heller ikke i Nudansk Ordbog, i Ordbog over det danske Sprog eller i selv de største ordbøger fra dansk til et fremmed sprog.

Der er på den anden side ingen tvivl om at *hygsom* er et ord der anvendes i et vist omfang og dermed kan siges at være »anerkendt« i sprogbrugen. De ældste eksempler i nævnets samling er fra aviser fra begyndelsen af 1940'erne, men ordet er sandsynligvis ældre. I de senere år er det blevet benyttet af forfattere som Tom Kristensen, Børge Outze, Leif Panduro og Sven Holm.

Ordet kan vistnok kun bruges spøgefuldt og måske let afstandtagende, og det er formentlig opstået ved at man for at opnå en let humoristisk virkning har udskiftet afledningsendelsen *-elig* i ordet *hyggelig* med endelsen *-som* (jf. *morsom*, *grusom* m. fl.).

Spørgsmål: Ordet *konfitterknas* bruges jo om lækre ting, lækkerier og lignende. Mon det kommer af fransk *confit d'oise* (gås, henkogt og opbevaret i fedt)?

7009

Svar: Ordbog over det danske Sprog har ikke denne form af ordet men *kramfitterknas* (med sideformen *komfitterknas*) og nævner at *kramfitter-*, *komfitter-* måske er spøgende omdannelser af *konfiture* (= konfekt) eller *konditor*.

Det er ikke sandsynligt at ordet kommer af *confit d'oise* (men de franske ord *confit* og *confiture* er naturligvis nært beslægtede).

I øvrigt kendes der også en sideform *flitterknas* (overskrift i Information 4.7.1968: Finkulturelt flitterknas eller meningsfuld kunst), hvis første led kan være overført fra et ord som *flitterstads*, jævnfør at Ordbog over det danske Sprog oplyser at *kramfitterknas* også kan bruges som ubestemt udtryk for noget urigtigt, galt.

Ordbogen nævner en tredje betydning af dette ord: 'erotisk tilnærmelse', 'samleje'. I denne anvendelse er det muligt at navneordet *en fitte* ('kvindens (ydre) kønsdele') optræder i ordets midterste stavelser, jf. Kaj Bom: Slangordbogen, 1957, s. 67 (hvor det desuden oplyses at der fra 1920'erne er noteret yderligere en sideform: *kramfilerkomknas*).

Alle disse forskellige ordformer viser at sproget her bliver brugt som legetøj; se om dette emne nærmere i den lige nævnte bog, s. 25-28.

Spørgsmål: Kan følgende formulering accepteres: *Han var ikke i stand til at kunne gennemføre sine planer?*

7010

Svar: Nej. Formuleringen *var i stand til at kunne gennemføre* er en sammenblanding (en såkaldt kontamination) af *var i stand til at gennemføre* og *kunne gennemføre*.

Da *være i stand til* her betyder det samme som *kunne*, bliver resultatet af sammenblandingen et udtryk der rummer en overflødig gentagelse (en såkaldt pleonasme). I mindre omhyggeligt sprog kan der forekomme dobbeltudtryk af samme type, fx

Vi må kunne være i stand til at ...

Vi har mulighed for at kunne ...

Det var ham umuligt at kunne ...

Jeg fik besked på at skulle ...

Man gav ham tilladelse til at måtte ...

Sådanne formuleringer bør undgås.

Spørgsmål: Hvor sættes der komma efter den grammatiske kommatering i følgende sætninger med *forudsat* og *uanset*:

Han kommer forudsat at han får tid.

Hun rejser uanset om han er der eller ej.

Forudsat at han kan få fri, holder vi møde i morgen.

Uanset hvordan vejret var, gik han en tur hver dag.

Svar: Den grammatiske opbygning af disse sætninger kan opfattes på flere måder. Det synes at være rimeligt at betragte ordforbindelserne *forudsat at*, *uanset om* osv. som bindeordsled der indleder en sætning, og derfor må det her tilrådes ikke at sætte komma efter *forudsat* og *uanset*. Derimod må der efter den grammatiske kommatering sættes komma foran *forudsat* og *uanset* når de indleder en sætning og der står en overordnet sætning foran, som i de to første eksempler.

Sprognævnets publikationer

Stavning af russiske navne (96 s.; 19,15 kr.) udkommer 3. april. Den måde russiske navne staves på i de sprog der bruger det latinske alfabet, er højest uensartet og ofte inkonsekvent. Nu er det lykkedes sprognævnene i Danmark, Norge og Sverige i samarbejde med slavister i de tre lande at udarbejde et fællesnordisk system der skulle gøre det muligt at nå frem til ensartede stavemåder i aviser, håndbøger, atlasser osv. Foruden en redegørelse for det fællesnordiske transskriptionssystem og baggrunden for det, rummer bogen en navneliste med ca. 2000 russiske efternavne, fornavne og geografiske navne.

Turistord i Norden (95 s.) udkommer i april eller maj. I bogen er en lang række ord som danske turister kan få brug for når de rejser i de andre nordiske lande, opført på dansk, norsk, svensk, finsk, islandsk og færøsk.

Nyt fra Sprognævnet udgives af Dansk Sprognævn og udsendes 2 gange årlig. Abonnement kan tegnes ved indbetaling af et engangsgebyr på 2,00 kr. til Dansk Sprognævn (postkonto 10 17 20). Tilmeldingsgebyret er det samme hvis der bestilles flere eksemplarer til samme abonnent. I øvrigt er abonnementet gratis. – Nr. 3 kan fås endnu.

Dansk Sprognævn, Vester Voldgade 115³, 1552 København V. Postkonto: 10 17 20. Telefon (01 28) BYen 705. Ekspeditionstid mandag–fredag 13–15.