

# NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Ordfrekvens och ordboksutrymme	
Forfatter:	Bo-A. Wendt	
Kilde:	Nordiska Studier i Lexikografi 11, 2012, s. 622-628 Rapport från Konferens om lexicografi i Norden, Lund 24.-27. maj 2011	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive">http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive</a>	

© Nordisk forening for lexicografi

## 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.

# Ordfrekvens och ordboksutrymme

*Bo-A. Wendt*

In a linguistic description there is normally a more or less conspicuous lack of balance between the space a phenomenon requires to be fully described and its frequency in actual language usage, here called a lack of *iconicity*. The article investigates to what degree some Danish and Swedish dictionaries are iconic in this sense, i.e. whether the space of a word in the dictionary corresponds with its actual frequency. As expected, iconicity is rather meager, the most frequent words being treated much shorter than their actual frequency would suggest and the less frequent ones the other way around, though the more comprehensive the dictionary the more iconic it normally is regarding the first category. Further, frequent nouns and adjectives can be shown to exhibit a more iconic description than frequent verbs, and compounds and derivations treated together with their base words often exhibit a (shorter and thus) more iconic description than those treated as their own entries (often iconically much too extensive). Iconicity is of course not the overall ideal state of good dictionary entries: frequent forms are often best treated rather generalized (and thus rather shortly), some less frequent words can only be described accurately if described rather extensively. Yet, it is argued that iconicity should not be dismissed altogether but to be considered as one factor (among many other ones) when deciding upon how much space a word is worth in a dictionary.

**NYCKELORD:** ordböcker, ikonisk, beskrivningsutrymme, frekvens, hörfrekventa ord, lårfrekventa ord, ordboksartiklars längd.

När jag arbetade som assistent inom projektet Svenska Akademiens grammatik (SAG), uttryckte någon gång redaktören Staffan Hellberg bryderi för att den grammatiska beskrivningen blev så oikonisk. Med detta avsåg han det förhållandet att de mest hörfrekventa grammatiska strukturerna ofta kunde beskrivas med bara några få ord, till exempel finitverbets andraplacering i svenska huvudsatser, medan ovanliga strukturer, undantagen från huvudregeln, gärna krävde mycket utförliga beskrivningar för att man skulle förmå ringa in dessas förutsättningar och giltighet. Ja, också rena gränsfall med avseende på grammatikaliteten över huvud taget kunde ibland kräva mer utrymme än den regel som gäl-

ler 99 fall av 100. Detta gjorde alltså grammatikframställningen oikonisk, med en ojämvikt mellan det man ser eller hör mycket av i verkligheten och det man ser mycket av vid läsningen: hög frekvens i språket – litet utrymme i texten, låg frekvens i språket – stort utrymme i texten. Den bild, ἡ εἰκὼν, som textens omfång ger läsaren svarar helt enkelt inte mot det avbildades vikt. Närmast ansluter väl Hellbergs benämning till Charles S. Peirce' användning av *icon* som semiotisk term, avseende ett tecken med likhet mellan form och betydelse (jfr Liszka 1996: 37).

Denna principiella synpunkt på språkvetenskapliga beskrivningar (eller rentav vetenskapliga beskrivningar över huvud taget) har sedan dess gjort sig påmind i mina tankar allt emellanåt. I min egen forskning har jag mest sysslat med stilhistoria eller åtminstone språkbruk i historisk belysning. Där är det kanhända enklare att åstadkomma detta att det vanligaste bruket också får störst utbredning i beskrivningen, men det är ändå lätt hänt också där att de udda fallen, ibland av avgränsningsskäl, måste ägnas jämförelsevis stort utrymme. Så småningom har jag tillika hamnat i lexikografiens våld och därvid ibland kommit att fundera litet kring hur det är med ordböckers ikonicitet, och det är detta som den följande framställningen kommer att avhandla, dels resonemangsvis (i avsnitt 1), dels på grundval av en liten stickprovsundersökning jag har gjort i några ordböcker, varvid jag försökt sätta siffror på graden av ikonicitet (i nyssnämnda mening, avsnitt 2).

## 1. Ordböckers ikonicitet i teorin

När det gäller förhållandet mellan frekvens och beskrivningsutrymme har ordböcker – där man beskriver ett ord i taget – givetvis det försteget framom grammatiken att de ovanligaste orden svårligen kan komma att påtagligt breda ut sig så mycket mer än de vanligaste. Däremot är det ingalunda självklart att de nödvändigtvis skulle få mindre utrymme heller, åtminstone näppeligen så mycket mindre som deras förhållandevis lägre frekvens borde givit upphov till – om nu beskrivningen skall vara någorlunda ikonisk i Staffan Hellbergs mening. Ikoniciteten härvidlag är förstås i hög grad avhängig av typen av ordbok. I en ordbok som bara redovisar stavning och böjning tar de allra flesta orden ungefär lika stort utrymme, vilket givetvis är oikoniskt i förhållande till frekvensen men ändå inte alls på det negativt korrelerande sätt som så ofta i grammatiken. Ett litet försteg har dessutom vissa av de vanligare orden i det att hög frekvens gärna medför större variation att redovisa. Vad sedan gäller ordböcker med fylligare lexikografisk beskrivning, omfattande också semantik, konstruktionsmönster och idiomatiska uttryck, så finns det här rimligtvis än mer att säga om framför-

allt de mer frekventa orden. Någon given korrelation finns det emellertid inte för de allra högfrekventaste formorden, till exempel konjunktionen *och*, kan – och bör – beskrivningen göras skäligen generell och därför sannolikt rätt kortfattad. Och de av utomspråkligt sakliga skäl ovanligare orden kan i gengäld ha en rätt vildvuxen formvariation som kan kräva nog så vidlyftigt utrymme, till exempel ord för gamla medicinalväxter (som *tibast* i SAOB). I just en ordbok som SAOB, som grundar sin beskrivning på ett material och dessutom uttryckligen försöker reglera förhållandet mellan detta material och beskrivningen i ett så kallat relationstal (som anger förhållandet mellan antalet belägg i materialet och antalet spalter i den färdiga ordboksartikeln), borde ju ikoniciteten ha riktigt hyfsade förutsättningar att visa till och med positiv korrelation.

Den förväntade ikoniciteten förutsätter dock två saker. För det första måste förstås materialet vara någorlunda frekvensviktat, och det kan vi nog krasst slå fast att detta inte är förhållanden vad gäller SAOB:s material. De grå funktions- och formorden är uppenbart underföreträdda i SAOBArkiv, och det beror givetvis på hur materialet är insamlat. Genom excerpering av bara några få utvalda belägg per sida har excerptisterna haft en tendens att föredra litet färgstarkare ord i sitt urval. Man kan som jämförelse tänka på hur materialet hade sett ut om det istället hade varit totalexcerperingar av ett antal texter. Jag skall här bara ge ett enda exempel, från ord jag själv haft om hand i ordboken: verbet *tro* har 1200 belägg i SAOB:s material, prepositionen *till* 9000 och prepositionen *under* 2100, vilket låter som ett betryggande övertag för de båda formorden, men hade materialet istället utgjorts av en totalexcerpering av det tidningsmaterial som NFO 2 grundar sig på, skulle dessa deras övertag var för sig ha varit mer än dubbelt så stora.

För det andra skulle en materialgrundad ordboks tilltänkta ikonicitet förutsetta likabehandling av alla ord oavhängigt deras plats i beskrivningens strukturhierarki. Här drar SAOB:s redigeringspraxis åt ett annat håll. Självständiga artiklar tillåts normalt breda ut sig förhållandevis mer med avseende på beskrivningsprecision och inte minst antal citerade språkprov än vad sammansättningar eller avledningar behandlade inne i en annan artikel brukar få göra. Återigen skall jag bara ge ett enda exempel från SAOB-praktiken: substantivet *under*, som givetvis är en egen artikel, har i materialet 410 belägg och beskrivs i den tryckta ordboken på 3 spalter; verbet *underkasta*, som behandlas i en sammansättningsramsa inuti artikeln om prepositionen *under*, har 396 belägg och beskrivs på 2 spalter, 69 % av vad det borde ha haft med hänsyn till antalet belägg i jämförelse med substantivet *under*; avledningen *underkastelse*, behandlad i en sekundär avledningsramsa under nyssnämnda sammansättning, har 52 belägg och beskrivs på en tiondedels (0,1) spalt, 38 % av vad det borde ha haft i jämförelse med sitt huvudord och bara 26 % i jämförelse med substantivet

*under*. Denna ojämvikt gäller alltså i förhållande till faktisk beläggsrikedom i SAOB:s material. Om vi istället ginge till faktisk frekvens i språket, blir ojämvikten än påtagligare, eftersom *underkasta* är drygt dubbelt så vanligt som substantivet *under* enligt NFO 2.

## 2. ... och i praktiken

Jag skall nu gå över till en mindre stickprovsundersökning som jag gjort i några danska och svenska ordböcker, där jag försökt mig på att sätta siffror på ikoniciteten vad gäller ordfrekvens och ordboksutrymme. De här talen måste tas med en stor nypa salt, eftersom de är resultat av ett skäligen fyrkantigt tillvägagångssätt. Ordböckerna är DDO och ODS för danskan, SAOL, SO och SAOB för svenskan, och för ordfrekvenserna har jag använt Bergenholtz 1992 och NFO 2 (1971). Detta senare val är förstås inte så rättvisande för de historiska ordböckerna, men i gengäld har jag bara valt ut basala ord, dels de allra vanligaste i språket och de vanligaste substantiven, verben och adjektiven, dels ett antal lågfrekventa – och med tanke på SAOB endast ord på A–T. Jag har därvid utgått från NFO 2, som ger tal för lemmatiserade ord, och sedan tagit motsvarande ord i den danska frekvensordboken, som endast redovisar graford. För danskans del har jag därför fått lämna därhän vissa av orden, eftersom de uppenbart döljer homografer. De utvalda orden har sedan frekvensviktats inbördes, och så har jag som utgångspunkt för jämförelser tilldelat ord vars frekvens är en 1000-del eller mindre av det allra högfrekventaste ett ikoniskt omfång om en rad i SAOL och tre rader i DDO och SO, ord vars frekvens är en 5000-del eller mindre av det högfrekventaste sex rader i SAOB samt ord vars frekvens är en 10000-del eller mindre därav tre rader i ODS. Den sistnämnda särbehandlingen beror på att de ovanligare danska orden är förhållandevis ovanligare jämfört med de vanligaste än i svenskan, vilket sannolikt sammanhänger med att det är de svenska orden som varit utgångspunkt – med följd att vissa av de danska motsvarigheterna helt enkelt inte är normalsynonymen där.

Denna myckna sifferexercis har givit det utfall som redovisas i tabellen.

*Tabell. Relativ andel av ikoniskt ordboksutrymme för vanliga och ovanliga ord utifrån en tilldelning av en rad i SAOL, tre rader i DDO, ODS och SO och sex rader i SAOB åt de lågfrekventaste orden (taget som en snävare kategori i ODS och SAOB).*

De vanligaste orden						Några ovanliga ord					
Ord	DDO	ODS	SAOL	SO	SAOB	Ord	DDO	ODS	SAOL	SO	SAOB
<i>den</i>	3	5	3	2	27	<i>entenlantingen</i>	400	170	150	600	500
<i>og/och</i>	5	6	1	3	6	<i>førend/førrän</i>	400	4000	50	100	130

<i>i</i>	12	11	1	3	10	<i>långs/längs</i>	650	400	<b>100</b>	216	<b>100</b>
<i>at/att</i>	8	8	1	2	19	<i>nogen-/någonting</i>	-	1000	<b>100</b>	233	<b>100</b>
<i>en</i>	23	9	2	3	9	<i>unden/undan</i>	-	1200	<b>150</b>	283	500
<i>på</i>	30	26	1	6	<b>105</b>	<i>somme, pl./somlig</i>	667	340	400	333	<b>130</b>
<i>till/till</i>	32	36	2	26	44	<i>hvorvidt/huruvida</i>	400	360	200	600	<b>130</b>
<i>med</i>	25	18	1	7	31	<i>inbyrdes/inbördes</i>	467	950	300	267	867
<i>aflav</i>	26	9	2	6	8	<i>ligesådan/likadan</i>	467	2600	200	300	333
<i>som</i>	25	38	2	5	13	<i>enhver/envar</i>	344	<b>133</b>	300	233	500
substantiv:											
<i>dag</i>	553	160	19	201	325	<i>kro/krog</i>	533	400	<b>100</b>	367	667
<i>del</i>	344	<b>117</b>	19	<b>96</b>	315	<i>provinsby/ landsortsstad</i>	267	<b>140</b>	200	33	<b>50</b>
<i>gang/gång</i>	234	79	20	142	173	<i>pistol</i>	767	280	<b>100</b>	467	1000
<i>land</i>	431	200	45	<b>103</b>	<b>91</b>	<i>silhouet/siluett</i>	700	900	300	367	1300
<i>tid</i>	455	206	33	<b>81</b>	171	<i>serart/särart</i>	-	400	<b>100</b>	200	300
verb:											
<i>ha</i>	...	...	2	16	21	<i>bemestre/bemästra</i>	233	2400	<b>100</b>	267	200
<i>kunne/-a</i>	24	15	1	16	13	<i>förudsigel/förutspå</i>	800	300	<b>100</b>	167	250
<i>skulle/skola</i>	54	32	3	16	61	<i>harmonisere/-a</i>	567	3000	200	367	500
<i>få</i>	...	...	2	34	57	<i>skimre/skimra</i>	-	2000	<b>100</b>	633	2000
<i>blive/bli</i>	<b>60</b>	17	5	12	<b>93</b>	<i>mala</i>	...	...	400	733	3000
adjektiv:											
<i>gammell/-al</i>	188	105	19	112	<b>88</b>	<i>barnlig/barnslig</i>	567	333	<b>100</b>	267	4000
<i>god</i>	141	109	80	220	573	<i>farverig/färgglad</i>	433	<b>100</b>	<b>100</b>	133	<b>100</b>
<i>ny</i>	121	<b>60</b>	5	35	34	<i>karrig/karg</i>	567	2000	200	333	1000
<i>stor</i>	143	148	4	41	131	<i>kynisk/cynisk stereotyp</i>	433	400	300	333	1000
							400	600	300	267	300

Siffrorna i tabellen anger procentuell andel av det i förhållande till frekvensen i faktiskt språkbruk ikoniska ordboksutrymmet (givet det tilldelade radutrymmet för de lågfrekventaste orden); idealet är med andra ord 100 – då är utrymmet helt ikoniskt. (Tal mellan 50 och 150 = någorlunda ikoniska värden utmärks med fet stil i tabellen.) Nu skall inte dessa tal läsas som ”hur väl man lyckats” i detta hänseende, eftersom detta slags ikonicitet knappast har varit något medvetet mål alls i någondera ordboken (annat än möjligtvis ytterligt medelbart i SAOB genom dess relationstal (se avsnitt 1 ovan)); det man inte har någon tanke på att eftersträva kan man inte heller lyckas eller misslyckas med att uppnå. Närheten till eller avståndet från det ikoniska 100 är istället huvudsakligen ett utfall av andra omständigheter och överväganden kring redigeringen av de utvalda orden.

Som väntat finner vi överlag låga för att inte säga mycket låga tal för språkets allra vanligaste ord och lika väntat allra lägst i ordlistan SAOL, högst i SAOB.

Också ODS och i än större utsträckning DDO uppvisar ett antal tvåsiffriga tal, och särskilt DDO mycket jämnare fördelade sådana än de svenska. Möjligen återspeglar detta senare att DDO bygger på en datoriserad korpus med totalexcerperade texter (där givetvis faktisk frekvens mycket tydligare får fullödigt genomslag). I SAOB återfinns dessutom en ytterst påtaglig avvikare i sällskapet: prepositionen *på* med nästintill ikoniskt omfång, en följd av den enastående detaljrikedom som denna artikel uppvisar (i sin tur väl delvis en följd av de tilläggsexcerperingar som redaktören uppges ha gjort i arbetet med detta ord). Av de vanligaste innehållsorden håller sig sedan substantiven överlag väl framme, också i ordlistan (jämförelsevis i alla fall). I SAOB har *dag* och *del* till och med tre gånger större utrymme än vad frekvensen borde lett till, och i DDO *dag*, *land* och *tid* hela fyra-fem gånger större utrymme. De högfrekventa adjektiven har påfallande jämnt fördelat och rätt ikoniskt utrymme i de två danska ordböckerna. Här närmar sig rentav också SAOL full ikonicitet för *god* (detta tack vare ovanligt frikostig redovisning av idiomatiska uttryck samt en särupplysning om komparativens och superlativens användning). I SAOB har *god* fått nästan sex gånger ”för stort” utrymme. Bland de ovanligare orden har förstås många lika oikoniskt stort utrymme, eller ännu mycket större. Detta gäller särskilt i de två stora ordböckerna, ODS och SAOB, med SAOB:s *barnslig* som allra ”värst”: 40 gånger större än sitt ikoniska omfång.

Ordet *barnslig* är avledning av *barn* men behandlas i SAOB i en egen artikel (liksom *barnlig* i ODS). Å andra sidan är det, mot bakgrund av vad som sades i avsnitt 1 ovan, värt att lägga märke till att det enda ordet bland de ovanligare som underskrider sitt ikoniska omfång är flerledssammansättningen *landsortsstad* i SAOB, där redovisat i en sekundär sammansättningsramsa (under *landsort*, i sin tur i ramsa under *land*). Också i ODS gynnar sådan osjälvtändig ställning i beskrivningen ikonicitet bland de lågfrekventa orden: både *provinsby* och *farverig* är beskrivna i sammansättningsramsor inuti respektive förleds artikel.

Om vi jämför ordböckerna parvis över språkgränserna, kan vi se att DDO uppvisar påtagligt bättre ikonicitet än SO för de vanligaste formorden men å andra sidan har oikoniskt stort utrymme för de vanligaste substantiven, där SO överlag ligger nära det ikoniska idealet. Den viktigaste iakttagelsen vad gäller ODS och SAOB är kanhända den att ODS överlag har en litet jämnare nivå i respektive grupp av vanliga ord – ingalunda alltid ikonisk men jämförelsevis jämn, där SAOB:s variation däremot är rätt påfallande. I gengäld har SAOB en jämnare nivå i gruppen lågfrekventa formord, där det till och med är så att hälften har fått ett någotsånär ikoniskt utrymme.

All denna sifferexercis skall nu tas för vad den är, och den tillåter knappast att dra alltför stora växlar på. Dels är dess utfall tillkommet med en grovhuggen metod vars validitet kan ifrågasättas, dels bevisar den något som vi i dess huvud-

drag resonemangsvi kunde räkna ut på för hand, dels mäter den ju något som bara är en sidoverkan av helt oavhängiga ställningstaganden i den lexikografiska praktiken. Huruvida den inte bara borde vara denna oavsedda sidoverkan får bli ämnet för det avslutande avsnittet.

### 3. Ikonicitet alls önskvärt?

Är då det större mått av ikonicitet som ändå legat som ett slags uttalat eller rentav halvt uttalat idealtillstånd bakom den hittillsvarande framställningen alls något att sträva efter, ens i en stor och uttömmande lexikografisk beskrivning som den i ODS eller SAOB? Fullständig ikonicitet kan vi genast avfärda som rätt eller snarare alldeles orimlig (liksom den givetvis är det också i en grammatisk framställning – något annat menade naturligtvis inte heller Staffan Hellberg). Det verkar minst sagt långsökt att SAOB:s artikel om konjunktionen *och* skulle göras 20 gånger längre bara för att motsvara ordets enastående vanlighet och väl än orimligare att språkets någorlunda grundläggande men ändå inte så väldigt frekventa ord, av typen *mala*, *karg* och *barnslig*, skulle behöva avstå från en fullödig beskrivning bara för att hovsamt hålla sig inom sina ikoniska skrankor. Likafullt vill jag hävda att det kan finnas skäl att emellanåt ha med detta slags ikonicitet i bakhuvudet; annars hade jag förstås inte alls tagit till orda i ärendet. Användningen av detta som argument för en fylligare, eller för den delen mindre fyllig, beskrivning måste givetvis lämpas efter läglighet, men ett ords centralitet i ordförrådet borde vara en självständig faktor, bland många andra förvisso (och delvis samspelande), när man avgör hur mycket man vill avsätta av ordbokstexten åt det. Att utrymmet i förhållande till frekvens är så väldigt olika inom var sina grupper med frekvensmätigt snarlika typer av ord som det ändå är i SAOB bör till exempel trots allt kunna verka som ett litet memento i sammanhanget. (Det återstår ju dock att jämföra dessa förhållanden också med materialets omfång för respektive ord.) Åtminstone vill jag så här till sist slå ett slag för att ett ords beskrivning såsom en självständig artikel eller såsom en underordnad del av en annan artikel inte i sig borde ha någon betydelse för hur stort utrymme det tillåts få, det vill säga hur ikoniskt dess omfång blir.

#### KÄLLOR OCH LITTERATUR:

Bergenholtz, Henning. 1992: Dansk frekvensordbog baseret på danske romaner, ugeblade og aviser 1987–1990. København.



- DDO = Den danske ordbog, hovedred.: Ebba Hjorth & Kjeld Kristensen, udg. af Det Danske Sprog- og Litteraturselskab, sex band. København 2003–2005.
- Liszka, James Jakób, 1996: A General Introduction to the Semeiotic of Charles Sanders Peirce. Bloomington & Indianapolis.
- NFO 2 = Allén, Sture, 1971: Nusvensk frekvensordbok baserad på tidningstext 2. Lemman. Stockholm.
- ODS = Ordbog over det danske sprog, grundlagt af Verner Dahlerup, udg. af Det Danske Sprog- og Litteraturselskab, bind 1–27 + supplementbind 1–5. København 1919–1954, 1992–2005.
- SAOB = Ordbok över svenska språket, utg. av Svenska Akademien, 1893 ff. Lund.
- SAOL = Svenska Akademiens ordlista över svenska språket, trettonde uppl. Stockholm 2006.
- SO = Svensk ordbok, utg. av Svenska Akademien, utarbetad vid redaktionen för Svenska Akademiens samtidsordböcker, Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet, två band. Stockholm 2009.

*Bo-A. Wendt*

Svenska Akademiens ordboksredaktion  
bo.wendt@svenskaakademien.se