

Ngram i SAOB

Karin Cavallin

The purpose of this study is to identify and eventually analyze reoccurring strings in the definitions of SAOB. Until recently, the digital version of SAOB consisted more or less of plain text, with some typographical information. Structure and informational categories in SAOB had not been marked in the xml code, and hence been practically unreachable for digital management. In order to improve readability, usefulness and adaptability of SAOB we have digitally identified different components of the text in the dictionary (Bäckerud 2016:102). During 2019 we, among other things, identified our language samples and the definitions. The definitions in SAOB contain several reoccurring strings. Within lexicography we know them as *collocations*. Since these strings usually lack other collocational features, we prefer to call them *reoccurring ngrams*. These ngrams can be quite formulaic, very long and differ from ordinary written and spoken Swedish. Ngram tools like *Ngram Statistics Package* (Banerjee & Pedersen 2003) and *The Sketch Engine* (Kilgarriff et al. 2004) have been used for a long time in lexicography to e.g. identify collocations. Such tools are here proposed to be used for purely lexicological purposes, on the dictionary text per se.

1. Inledning

Svenska Akademiens ordbok, SAOB, har redigerats under mer än hundra år. Ambitionen är att beskriva hela det nysvenska ordförrådet, från 1521 och framåt. Beskrivningen utgörs bland annat av etymologi, formredovisning (genus, plural, böjningar och dylikt), autentiska citat och deras respektive källa, samt definitioner.

En redigering som pågår under mer än ett sekel skapar givetvis ett eget sätt att beskriva betydelser av ord, med mer eller mindre formelartade uttryckssätt. Dessa uttryckssätt förändras också över tid, och kan skilja sig mellan olika redaktörer. För våra läsare och för nya redaktörer kan detta formelartade språk te sig ogenomträngligt. Olikheter i uttryckssätt kan också ge ett spretigt intryck när man jämför artiklar som behandlar likartade fenomen.

2. Syfte och problemställning

Lexikografens uppgift är att beskriva språket. I denna studie undersöks lexikografens eget språk, dvs. definitionsspråket och dess formler. Det primära syftet med studien är att få fram data som underlättar både den fortsatta redigeringen och den nära förestående revideringen av SAOB. Därtill tillkommer ett rent lexikologiskt syfte, nämligen att möjliggöra närstudier av definitionsspråket i SAOB.

3. Forskningsöversikt

Vanligtvis använder man korpusar för att få underlag vid redigering av en ordboksartikel. Här kommer istället SAOB själv, eller snarare SAOB:s definitionstexter, användas som en korpus. Mig veterligen har inga kvantitativa, *datadrivna*¹ studier gjorts på själva definitionsspråket i ordböcker. Nilssons licentiatavhandling (Nilsson 2019) är ett exempel på en *hypotesdriven* studie över hur fenomenet *bildlighet* beskrivs i just SAOB. Det har ännu inte gjorts någon studie över hur SAOB:s definitionsspråk skiljer sig från språket i en välbalanserad korpus. Det finns indikationer på att SAOB:s definitionsspråk är ganska avvikande.

Vid redigering av en ordboksartikel är det vanligt att man undersöker hur ordets närkontext ser ut. Till exempel kan översikter i kollokationsverktyg, som *WordSketches* i *SketchEngine* (Kilgarriff et al. 2004) eller KORP:s *ordbild* (Borin et al. 2012), användas. Kollokationsverktyg är datorprogram som utgår från frekvensberäkningar av vilka strängar som samförekommer med varandra. Dessa strängar motsvaras oftast av det vi känner som *ord*, men verktygen bryr sig egentligen inte om ord. Det som räknas är samförekomst av strängar avskilda med mellanrum eller skiljetecken. Dessa strängar kommer härefter kallas *ngram*.

Dessa frekvensberäkningar är generellt uppbyggda av ngramberäkningar där resultaten är viktade, utslätade (*smoothed*²) med hjälp av associationsmått, *Measures of Association*. Samförekomst mellan strängar som förekommer ofta och tillsammans med det mesta presenteras som mindre relevant än strängar som förekommer mer sällan, men mer eller mindre uteslutande tillsammans (Banerjee & Pedersen 2003:376). Utslätningen görs för att alltför frekventa strängar (såsom konjunktioner och prepositioner) annars lätt tar över som samförekommande och för att de mest semantiskt eller syntaktiskt relevanta samförekomsterna ska komma högt upp som kollokationsförslag.

¹ *Bottom-up* är kanske den mer gängse termen för *datadriven*.

² Det saknas tyvärr bra svenska termer för många begrepp inom språkteknologi och datavetenskap.

Denna ngramstudie skiljer sig från andra tillämpningar av ngram³ på ett par punkter:

- En ordboks definitionstext används som korpus.
- Det är väldigt höga, återkommande ngram, som är av intresse. Vanligtvis består kollokationer av upp till 3–4 ord.
- Utslätning tillämpas inte. Vid högre ngram än 4gram minskar frekvensen i sig, och gör utslätning tämligen irrelevant.⁴

4. Metod och material

För att undersöka definitionsspråket i SAOB krävs att SAOB:s definitionstext har märkts upp. Detta har gjorts under 2019 i samband med det stora digitaliseringsprojektet vid SAOB (beskrivet i Bäckerud 2016). Hela SAOB består av ca 200 miljoner tokens, medan definitionstexten utgör ca 9 miljoner tokens. En definition i SAOB kan bestå av noll⁵ till flera definitionsled. Vid en utförligare definition separeras varje definitionsled med semikolon. I denna studie undersöks endast ngram som ingår i eller utgör ett definitionsled.

Eftersom SAOB har bibehållit den redigeringstradition som grundlades redan för hundra år sedan skiljer sig SAOB i vissa avseenden från ordböcker som är skrivna under ett senare och kortare tidsspänn. Exempelvis saknas en enhetlig struktur och olika informationskategorier har inte nödvändigtvis en bestämd plats. För definitionsleden innebär detta att exempelvis etymologiparenteser och brukligheter kan ligga insprängda mitt i ett definitionsled. Detta medför att definitionsleden behöver rensas innan det utförs beräkningar på dem. Etymologiparenteser har plockats bort.⁶ Bruklighetsmarkörer har däremot inte plockats bort. Dessa är svåra att utskilja maskinellt från övrig definitionstext. Definitionsled som egentligen utgör en hänvisning⁷, såsom sådana som inleds med 'jfr', samt innehåller strängar längre än ett tecken som innehåller versaler, har rensats bort. Parenteser har tagits bort. Se vidare i avsnitt 4.2.

³ Den mest kända tillämpningen av ngram torde vara Google Books NGram Viewer, www.books.google.com/ngrams.

⁴ Personlig mejlkontakt med Ted Pedersen, 2016-08-02.

⁵ Det förekommer definitioner som är helt tomma, samt definitionsled som enbart utgörs av en hänvisning.

⁶ Eftersom etymologiparenteserna är omgärdade av hakparenteser ([]) var detta enkelt. I skrivande stund har etymologiparenteserna hunnit märkas upp inom ramen för taggningprojektet.

⁷ SAOB kallar det för hänvisning. I andra ordböcker kallas det exv. för korsreferens.

4.1. Perioder

SAOB kan delas in i fyra huvudsakliga perioder, baserade på bland annat omfång och ambitionsnivå (se vidare Larsson 2014). Den första perioden sträcker sig från SAOB:s början 1883 till ca 1920, period 2 från ca 1920–1960, period 3 1960–2002, och period 4 från ca 2002 och framåt. För att kunna studera eventuella förändringar och olikheter över tid har i princip denna indelning använts även i denna studie. För enkelhets skull har initial bokstav inom tidsintervallet delats in som hörande till en viss period, och inte exakt efter redigeringstidpunkten. Tidsintervallen motsvaras ungefär av artiklar med begynnelsebokstav A–E, F–Q, R–S respektive T–V.⁸

Antal tokens i definitionstexten skiljer sig åt mellan perioderna. Period 1 har ca 1,2 miljoner tokens, period 2 ca 3,5 miljoner, period 3 ca 3,5 miljoner och period 4 (till och med band 37, UTSUG–VRETA) ca 700 000 tokens. Detta är helt i linje med respektive periods ambitionsnivå.

4.2. Det statistiska verktyget Natural Statistics Package, NSP

För att beräkna ngrammen har definitionstexten matats in i verktyget Natural Statistics Package, NSP (Banerjee & Pedersen 2003). I skrivande stund är hänvisningarna i SAOB inte taggade, utan ligger inom räckvidden för definitionstexten. Dessa har därför hjälpligt filtrerats bort med heuristiska metoder⁹ med de i NSP inbyggda funktionerna *stop*, *token* och *non-token* i NSP. Under hösten 2019 planeras hänvisningarna att taggas. Vissa specifika förkortningar som förekommer i SAOB, som ”o.” och ”l.”, har också specialbehandlats.

NSP är inte helt anpassat för syftet att söka efter så höga ngram som studeras i denna studie. Frågan är om något ngramverktyg är anpassat för det. Vid höga ngram växer komplexiteten exponentiellt, och inte ens en kraftfull dator med stort minne klarar av att genomföra hela körningen.¹⁰

5. Resultat

Resultaten från denna studie presenteras för respektive ngram i form av 10-i-topp-listor i tabellform. Jag kommer särskilt redovisa 1gram, 3gram, 5gram och 7gram. Urvalet beror dels på utrymmesskäl, dels på att variationen mellan 10-i-topp-listorna inte visade sig vara särskilt stor vid en ökning med bara ett token. Det krävdes en ökning av minst två token för att det skulle bli tydliga och intressanta skillnader mellan tabellerna.

⁸ Larsson (2014) tar i sin studie hänsyn till att vissa artiklar på exv. E och F inte är skrivna i alfabetisk-kronologisk ordning. Det görs inte i denna studie.

⁹ Definitionsled som inleds med ”jfr” har filtrerats bort.

¹⁰ NSP kördes på en virtuell dator med 16 kärnor och 112 GB primärminne (D14_v2). Ett ngramverktyg med mindre minnesåtgång hade, om det finns, varit lämpligare.

Tabellen över 1gram visar förstås inga formler eller relevanta samförekomster, men visar ändå lite intressanta saker. Det första värt att notera är att de mest frekventa 1grammen genom perioderna är ganska lika. Värt att notera är att dödtecknet (†) är utanför 10-i-topp-listan under period 3. Det är också tydligt att materialet är olemmatiserat. ”Af” och ”av” syftar givetvis på samma semantiska enhet, men denna har genomgått en stavningsreform under 1900-talet.

	Hela SAOB	Period 1	Period 2	Period 3	Period 4
1	i 457984	i 49664	i 190802	i 178831	i 38687
2	om 207138	af 19299	om 77336	om 89348	om 22485
3	av 168829	om 17969	av 72814	av 79340	av 14652
4	som 154297	till 17527	som 65032	som 60179	med 14073
5	med 136652	med 16974	† 55224	med 53395	som 12732
6	till 127713	† 16608	till 53416	till 45932	se 11641
7	† 112921	som 16354	med 52210	på 39260	till 10838
8	ngt 107444	på 13068	ngt 48636	ngt 38385	ngt 10387
9	på 106175	en 12482	på 45612	att 36530	på 8235
10	att 86256	ngt 10036	en 36910	se 35397	† 7895

Tabell 1: 1gram.

I tabellen för 3gram börjar uttryck som liknar den så kallade saobitiskan dyka upp. Mest frekvent är flerordsuttrycket ”i fråga om”, som torde vara någorlunda, men inte lika, frekvent även i en vanlig, balanserad korpus. På resterande platser dyker det mest upp uttryck som är typiska för SAOB:s sätt att beskriva ords betydelser på. Vid 3gram har den absoluta frekvensen sjunkit avsevärt, men på grund av platsbrist redovisas inte siffrorna här.

	Hela SAOB	Period 1	Period 2	Period 3	Period 4
1	i fråga om	i fråga om	i fråga om	i fråga om	i fråga om
2	i sht i	i sht i	i sht i	i sht i	i sht i
3	i vissa trakter	i sht om	i vitter stil	i vissa trakter	i sht förr
4	i sht förr	i högre stil	i vissa trakter	i sht förr	numera i sht
5	i vitter stil	i sht förr	i sht förr	i vitter stil	fråga om att
6	i skildring av	numera i sht	i sht om	i skildring av	i sht om
7	numera i sht	för att beteckna	numera i sht	numera i sht	i skildring av
8	i sht om	på grund af	sht i vitter	sht i vitter	senare led i
9	sht i vitter	i motsats till	i högre stil	fråga om att	om person som
10	med tanke på	sht i skriftspr	för att beteckna	i sht om	om ngt sakligt

Tabell 2: 3gram.

Vid 5gram dominerar det formelartade språket. De mest frekventa 5grammen motsvarar framförallt det som används för att beteckna bruklighet, dvs. om eller när, eller i vilket sammanhang, ett beskrivet ord används. I tabellen förekommer bl.a. uttryck som betecknar inom vilken social eller regional sfär den beskrivna aspekten av ordet används, men även mer metaspråkligt språk, som ”stundom svårt att skilja från”. Det är inte självklart att skilja ett ords användningsområde från dess betydelse, varför konceptet *bruklighet* är problematiskt i sig. Det är därför inte givet att bruklighet bör filtreras bort, om det hade gått att göra på ett enkelt sätt.

Hela SAOB	Period 1	Period 2	Period 3	Period 4
i sht i vitter stil	i sht i fråga om	i sht i vitter stil	i sht i vitter stil	i sht i fråga om
i sht i fråga om	i sht i högre stil	i sht i fråga om	i sht i vissa trakter	i sht i skildring av
stundom svårt att skilja från	stundom svårt att skilja från	i sht i högre stil	i sht i skildring av	stundom svårt att skilja från
i sht i vissa trakter	i sht i vitter stil	stundom svårt att skilja från	stundom svårt att skilja från	i fråga om att ngt
i sht i högre stil	som har till yrke att	ofta svårt att skilja från	ngn gg i skildring av	i fråga om att ngn
i sht i skildring av	ofta svårt att skilja från	för att beteckna att ngn	numera i sht i skildring	numera i sht i skildring
numera i sht i vitter	person som har till yrke	om var särskild av de	i sht i fråga om	språkprov svårt att skilja från
numera i sht i skildring	i sht i religiös stil	för att beteckna att ngt	numera i sht i vitter	vissa språkprov svårt att skilja
ngn gg i skildring av	med syftning på en efterföljande	sht i vitter stil i	numera i sht i vissa	i vissa språkprov svårt att
sht i vitter stil i	ngn gg i högre stil	numera i sht i vitter	sht i vitter stil om	ersatt av bestämning inledd av

Tabell 3: 5gram.

Då denna studie primärt är en förstudie i väntan på en än mer uppmärkt SAOB, går jag inte in närmare i detalj på varje ngrams specifika frekvenstabell. Vid högre ngram än 7gram och större mängd tokens i periodfilerna räcker inte RAM-minnet till att köra igenom befintliga data. Endast de perioder som har minst antal tokens i sig kan hanteras av NSP för 7gram och uppåt. I appendix återfinns de 10 mest frekventa ngrammen för 7gram och 10gram.

Bland 10grammen finns ett tydligt exempel där redaktörerna brustit i konsistens. Talen *ett* till *tolv* borde beskrivas på snarlika sätt i SAOB. Bland 10grammens 10-i-topp-lista ser vi att beskrivningssättet varierar på två olika sätt: *tal som i serien av hela tal har sin plats*, respektive *som i serien av hela tal har sin plats mellan*.

Detta föranledde mig att skärskåda samtliga i SAOB beskrivna tal¹¹ och jag kunde då finna ännu mer variation. *TVÅ*, *TRE*, *SJU* och *TIO* beskrivs med *tal som i serien av hela tal har sin plats mellan*, medan övriga beskrivs på ett antal olika sätt:

- EN/ETT ”det första grundtalet (i skrift ofta betecknat 1 l. I)”
- FYRA ”det tal som i talräckan har sin plats mellan tre o. fem”
- FEM ”det tal som i talräckan har sin plats mellan fyra o. sex”
- SEX ”tal som i talserien har sin plats mellan fem o. sju”
- NIO ”det tal som i talserien har sin plats mellan åtta o. tio”
- ELVA ”grundtal som följer närmast efter tio”
- TOLV ”grundtal angivande det antal som med en enhet överstiger elva; tal som i talserien har sin plats mellan elva o. tretton o. som utgör bas för duodecimalsystemet”

6. Diskussion och analys

Bruklighetsmarkörer är ofullständigt uppmärkta i dagsläget och skymmer sikten för den rena definitionstexten i ngrammen. I princip hade dessa kunnat manuellt sorteras bort från 10-i-topp-listorna, men då skulle resultaten bli svåra att upprepa. Dessutom skulle parallelliteten med 1gram och 2gram bli skev. Som nämndes i avsnitt 5 är det inte helt självklart var gränsen går mellan en regelrätt betydelsebeskrivning och bruklighetsmarkör. Denna studie tydliggjorde en oklarhet inom SAOB som var mer omfattande än man tidigare insett.

Det är en klar nytta med att beräkna ngram på SAOB:s definitionsspråk. Typiska formler tydliggörs och kan därmed underlätta vid upplärning av nya redaktörer. Framöver skulle man även, givet att en jämförande ngramanalys görs på en balanserad korpus, kunna jämföra språket i SAOB med vanlig skriven text.

NSP byggdes inte för syftet att beräkna långa ngram, utan för kortare ngram, 1–4gram. Det är sällan någon intresserat sig för längre ngram än så. I denna studie har jag tittat på de tio mest frekventa 1–10grammen. I många fall kan det troligtvis räcka med att titta på resultat från de lägre ngrammen i SAOB, och därigenom hitta längre återkommande ngram.

¹¹ Alla grundtal utom ÅTTA är till dags dato beskrivna i SAOB.

7. Avslutande kommentarer

Det som återstår att studera är relationen mellan definitionstext och annan text. Ju mer uppmärkt allt blir, desto mer korrekt kan man se fördelningen av olika informationskategorier. Man måste också fundera på hur relevant och användbart det är att veta att en viss formulering återkommer så få gånger som bara två. Är orden semantiskt nära besläktade är det förstås motiverat.

För att kunna ta ett samlat grepp vid betydelsebeskrivningarna vid revideringen kommer SAOB-redaktionen förmodligen framförallt att ta hjälp av interna hänvisningar för att hitta de artiklar som är relaterade till varandra. Jag tror dock att ngramanalysen kan vara ett komplement till detta, genom att man med den kan dels hitta fler relaterade artiklar, dels skapa en utgångspunkt för hur vår betydelsebeskrivning ser ut språkligt.

Det torde även vara möjligt att ta bort vissa innehållstunga ord ur ngramsökningarna, för att få fram ännu fler typiska formler.

7.1. Lemmatisering och normalisering

Definitionstexten i SAOB är tämligen enhetlig vad gäller former. Oftast skrivs definitionerna i presens, och det är sällan ord böjs i bestämd form eller plural. Dock förekommer det förstås en del böjningar. Det förekommer exempelvis plurala verbformer ända in på 1970-talet i definitionstexten i SAOB, trots att sådana mer eller mindre försvann ur svenska språket i övrigt under 1940-talet. Därtill tillkommer exempelvis ett antal i dag föråldrade stavningar som *hv* (*hvarpå*), *fv* (*öfverläggning*) och *f* (*af*) där vi i dag endast skriver *v*, och *dt* (*landtman*), där vi i dag bara skriver *t*. Det förekommer föråldrade former som *konung* (*kung*), *fader* (*far*), *moder* (*mor*). I onormaliserad text räknas *mor* och *moder* som två olika strängar. Strängarna *av* och *af* förekommer frekvent. En normalisering skulle gett en mer rättvisande bild av exempelvis frekvensen för alla de ngram som *af* eller *av* ingår i. För denna studies del hade det varit önskvärt med en normalisering av redaktionstexten, men det hade varit ett svårt och mycket omfattande arbete.

7.2. Användningsområden

Det finns tre huvudsakliga användningsområden för resultaten i denna studie. Dessa är att:

- strama upp definitionstexten vid revideringen,
- träna upp nya redaktörer som snabbare kan tillgodogöra sig viktiga formler,
- rent lexikologiskt kunna studera hur SAOB har sett ut genom åren.

Förfaringssättet är givetvis tillämpligt på andra ordböcker, under förutsättning att definitionsspråket i den aktuella ordboken är tillräckligt uppmärkt. Ett sidoreultat från studien har redan bidragit till att strama upp brukligheterna i SAOB inför revideringen, då denna studie synliggjort att SAOB-redaktionen haft oklara gränser mellan vad som ska läggas i den rena definitionstexten och vad som bör placeras som en uppgift av mer kronologisk, frekventiell, sociologisk eller regionalpråklig natur, alltså om *bruket*.

Framöver skulle man kunna jämföra med en balanserad korpus och därmed kunna se mer exakt hur svenskan i SAOB skiljer sig från svenskan i stort. Detta kan dels ge rent lexikografiska insikter, dels ge uppslag till vad man kanske bör förändra för att underlätta för läsaren vid revideringen. Det skulle kunna röra sig om en uttolkning av vissa uttrycksätt, eller en förändring eller förenkling av vissa uttrycksätt. Därtill tillkommer det rent pedagogiska syftet. Nya redaktörer får chans att få en överblick av definitionsspråket i SAOB, och kan därmed tillgodogöra sig uttrycksättet i SAOB snabbare.

Litteratur

Ordböcker

SAOB (1889–) = *Ordbok över svenska språket utgiven av Svenska Akademien*. 1–37. Lund, Sverige.

Annan litteratur

Banerjee, Satanjeev & Ted Pedersen (2003): The Design, Implementation, and Use of the n-gram Statistic Package. I: *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, s. 370–381. Mexico City, Mexiko.

Borin, Lars, Markus Forsberg & Johan Roxendal (2012): *Korp – the corpus infrastructure of Språkbanken*. Proceedings of LREC 2012, s. 474–478. Istanbul, Turkiet.

Bäckerud, Erik (2016): Indexeringen av SAOB. I: *Nordiska Studier i Lexikografi* 13, s. 101–109. Köpenhamn, Danmark.

Kilgarrieff, Adam, Pavel Rychly, Pavel Smrz & David Tugwell (2004): The Sketch Engine. I: *Proceeding of Euralex*, s. 105–116. Lorient, Frankrike.

- Larsson, Lennart (2014): En ”mer l. mindre stor” stor ordbok – om variationerna i SAOB:s omfång och ambitionsnivå. I: *LexicoNordica* 21. Tidskrift om leksikografi i Norden, utgitt av Nordisk forening for leksikografi. s. 61–80.
- Nilsson, Pär (2019): *Bildliga betydelser i SAOB: Om beskrivningen av betydelseutvecklingsmekanismer analyserad ur ett kognitivt semantiskt perspektiv*. Licentiatavhandling 2019, 150 s. Lund, Sverige.

Appendix

Hela SAOB

numera i sht i vitter stil i 61
 i vissa språkprov svårt att skilja från 50
 numera i sht i vitter stil om 50
 i vissa fall utan bestämd avgränsning från 47
 sht i vitter stil i fråga om 44
 i sht i vitter stil i fråga 44
 i vissa fall svårt att skilja från 41
 numera i sht i vitter stil till 38
 i sht i vitter stil i utvidgad 33
 i sht i vitter stil i mer 31

Tabell 4: 7gram, Hela SAOB.

Period 1	Period 2
med syftning på innehållet i en hel 8	i sht i vitter stil i fråga 18
med syftning på en efterföljande relativ bisats 8	sht i vitter stil i fråga om 18
syftning på innehållet i en hel sats 8	i sht i vitter stil i mer 17
numera i sht i högre stil i 6	i fråga om det ljud som uppkommer 16
den relativa bisatsen motsvarar en nutida konsekutiv 5	i sht i vitter stil i överförd 15
bisatsen motsvarar en nutida konsekutiv att sats 5	i sht i vitter stil som har 15
hvarvid den relativa bisatsen motsvarar en nutida 5	i sht i vitter stil i utvidgad 13
relativa bisatsen motsvarar en nutida konsekutiv att 5	numera i sht i vitter stil i 13
† i ställning efter det styrda ordet 5	i fråga om det förhållande att ngt 12
närmande sig karaktären af ett gm sin 4	numera i sht i fråga om utländska 12

Tabell 5: 7gram, Period 1 och Period 2.

Period 3	Period 4
numera i sht i vitter stil i 47	i vissa språkprov svårt att skilja från 45
i vissa fall utan bestämd avgränsning från 47	i vissa språkprov utan klar avgränsning från 11
numera i sht i vitter stil om 42	förled i ssgr betecknande ngt som har 8
i vissa fall svårt att skilja från 30	det som betecknas med det styrda ledet 6
numera i sht i vitter stil till 30	om ngt som vittnar om sådan egenskap 6
sht i vitter stil i fråga om 25	som i ngt avseende påminner om en 6
i sht i vitter stil i fråga 25	ssgr betecknande ngt som har avseende på 5
i sht i vitter stil i utvidgad 20	i ssgr betecknande ngt som har avseende 5
i sht i vitter stil om person 19	man får hand om ngt för att 4
med objektet ersatt av bestämning inledd av 18	det som anges med det styrda ledet 4

Tabell 6: 7gram, Period 3 och Period 4.

Period 1	Period 4
hvarvid den relativa bisatsen motsvarar en nutida konsekutiv att sats 5	göra så att man får hand om ngt för att 4
mindre närmande sig karaktären af ett gm sin ställning i 4	tal som i serien av hela tal har sin plats 3
närmande sig karaktären af ett gm sin ställning i satsen 4	om växt vars blad innehåller ett enzym som kan användas 3
sig karaktären af ett gm sin ställning i satsen betonadt 4	utgörande det led som i sitt sammanhang betecknar detta okända 3
så stor mängd af ngt som man kan omfatta med 3	som i serien av hela tal har sin plats mellan 3
af alla de enskilda om hvilka den tillagda bestämningen gäller 3	simmare som gjort de näst bästa resultaten vid proven till 2
bygdemålsfärgad framställning i fråga om vissa af seden föreskrifna föräringar 2	det ljud som uppkommer då luften sätts i rörelse av 2
jordbruk som icke är förenadt med boskaps-skötsel af ngn betydenhet 2	på att tidrymden omfattar den tidpunkt som anges med rektionen 2
i olika färger allt efter som ljuset kommer från ena 2	en plats som är belägen på nästan samma höjdnivå som 2
på det steg som utfyller det kolumnmellanrum på ett ark 2	om sladd som förbinder telefonlur med den övriga delen av 2

Tabell 7: 10gram.

Karin Cavallin
 fil. mag., redaktör och språkteknolog
 Svenska Akademiens ordbok
 Dalbyvägen 3
 SE-224 60 Lund
 karin.cavallin@svenskaakademien.se