

5 Empirisk ordforskning som grund för vidareutveckling av *Svensk ordbok utgiven av Svenska Akademien*

Emma Sköldberg
Göteborgs universitet

Kristian Blensenius
Göteborgs universitet

Empirisk ordforskning är central för det lexikografiska arbetet med att uppdatera *Svensk ordbok utgiven av Svenska Akademien* (SO). Detta kapitel beskriver hur olika korpusar och verktyg, främst från Språkbanken Text och Kungliga biblioteket, används för att välja ut, analysera och revidera uppslagsord. Fokus ligger på böjningsuppgifter och betydelsebeskrivningar. Vi diskuterar metodologiska och praktiska utmaningar, inklusive digitaliseringens påverkan på lexikografiskt arbete. Genom moderna språkteknologiska verktyg har processen blivit mer vetenskaplig och effektiv, samtidigt som behovet av ytterligare textmaterial och metodutveckling kvarstår.

1 Inledning

Empirisk ordforskning är avgörande för det lexikografiska arbete som bedrivs inom projektet Svenska Akademiens samtidsordböcker vid Göteborgs universitet. I det här kapitlet redogör vi för hur sådan forskning gagnar vidareutvecklingen av definitionsordboken *Svensk ordbok utgiven av Svenska Akademien* (SO) som beskriver svenskt samtidspråk. Den senaste upplagan av SO är från 2021 och arbetet inför en uppdatering av ordboken pågår (se avsnitt 2.2). Vi diskuterar utifrån ett antal exempel hur olika slags korpusar och verktyg, vilka tillhandahålls av bl.a. Språkbanken Text, används inom ramen för urval och införlivande av nya uppslagsord, revidering av befintliga ordboksartiklar och utmönstring av uppslagsord ur ordboken. För att avgränsa oss fokuserar vi i hög grad på arbetet med två relativt olika typer av information i ordboksartiklarna, nämligen de informationskategorier som rör uppslagsordens böjning respektive betydelsebeskrivning.

Kapitlet är upplagt på följande sätt: i avsnitt 2 redogör vi för våra meta-lexikografiska utgångspunkter inklusive några centrala begrepp inom lexicografien. Vi presenterar även SO, dess bakgrund och ger exempel på hur en ordboksartikel i verket kan se ut. Avsnitt 3 innehåller redogörelser över hur Språkbanken Text ordforskningsplattform Korp respektive Kungliga bibliotekets textresurser används i det lexikografiska arbetet. Därpå, i avsnitt 4, redogör vi för de tre övergripande uppgifterna i samband med arbetet med SO, dvs. att hitta, välja ut och beskriva nya uppslagsord, att uppdatera och vidareutveckla beskrivningen av redan införlivade ord samt att mönstra ut ord. Avsnitt 5 består av två underavsnitt, ett om hur arbetet med att ge böjningsuppgifter bedrivs och ett annat om hur beskrivning av uppslagsordens semantiska egenskaper går till. Kapitlet avslutas med ett kortfattat slutord.

2 *Bakgrund*

2.1 *Metalexikografiska utgångspunkter*

Svenséns lexikografiska handbok (Svensén 2009: 12–38) innehåller en utförlig presentation av olika slags ordböcker. Med utgångspunkt i ordbokens yttre form brukar man exempelvis skilja på pappersordböcker och elektroniska ordböcker som ordboksappar och webbordböcker. I jämförelse med en tryckt ordbok står mer utrymme till lexikografens förfogande i en elektronisk ordbok och detta kan bl.a. påverka antalet uppslagsord och språkprov i verket. E-ordboken är dessutom mer dynamisk och går lättare att uppdatera.

Ordböcker klassificeras också efter antalet språk de behandlar. De kan vara enspråkiga (monolingvala), tvåspråkiga (bilingvala) eller flerspråkiga. En ordbok kan gälla det allmänna ordförrådet, ett visst fackspråk (såsom det juridiska språket), en geografiskt avgränsad varietet (såsom skånska) eller ord tillhörande en viss stilnivå (såsom slang). Därutöver kan en ordbok behandla många olika aspekter på ett visst ordförråd (t.ex. uppslagsordens uttal, böjning, semantik och etymologi) eller endast en viss aspekt, t.ex. de konstruktioner som uppslagsorden kan ingå i. Olika ordböcker behandlar skilda språkskeden, t.ex. fornsvenska eller modern svenska. Därtill kan de vara synkrona eller diakrona och ha mer eller mindre normativa ambitioner.

Vid praktiskt lexikografiskt arbete tar lexikograferna ofta sin utgångspunkt i de informationskategorier som vanligtvis återfinns i en ordboksartikel (jfr Svensén (2009: 6–8)). Informationskategorierna kan delas in på följande sätt, utifrån vad de bidrar med för uppgifter:

- stavning, uttal, böjning (formell information)
- ordklassstillhörighet, konstruktionssätt (syntagmatisk information)

- betydelsebeskrivning (semantisk information)
- språkexempel
- bruklighetskommentar (pragmatisk information)
- historik (etablering, ursprung och släktskap).

I detta kapitel fokuserar vi som sagt på formell information som rör böjning liksom semantisk information i betydelsebeskrivningarna.

Ordböcker är i hög grad att betrakta som bruksföremål. Lexikografen måste ha klart för sig vilka den aktuella ordboken i första hand vänder sig till eftersom dess form och innehåll ska anpassas efter målgruppens eller målgruppernas behov. En ordbok som vänder sig till vuxna modersmåls-talare behöver ha ett annat innehåll än en ordbok för vuxna inlärare på nybörjarnivå, skolbarn etc. (se t.ex. [Atkins & Rundell 2008](#): 28–33, [Tarp 2008](#), [Bergenholtz & Bergenholtz 2011](#)). Lexikografen behöver också ta ställning till vilken eller vilka funktioner ordboken ska fylla eftersom också detta påverkar innehåll och utformning i hög grad. Ska den t.ex. användas vid reception (dvs. vid förståelse av tal och skrift) eller vid produktion (dvs. då man vill formulera sig), eller ska den framför allt dokumentera ett visst språk? Exempelvis är betydelseangivelser viktiga vid reception medan uttals- och konstruktionsuppgifter är relevanta vid produktion. Vid dokumentation fyller historiska uppgifter om uppslagsorden en central roll (se t.ex. [Malmgren 2009b](#)).

2.2 SO och SO-redaktionens uppdrag

SO kan karakteriseras som en enspråkig, synkron, allmänspråklig och i huvudsak deskriptiv ordbok. Med sina ca 65 000 uppslagsord vänder den sig i första hand till användare som har svenska språket som modersmål men också till vuxna inlärare med goda kunskaper i svenska. Målsättningen är att ordboken ska tillgodose flera funktioner. Reception är den primära funktionen, men SO ska också stödja användarnas språkliga produktion (se vidare bl.a. [Malmgren 2009a,b](#)).

SO vilar på vetenskaplig grund. I samband med det redaktionella arbetet förhåller sig lexikograferna till såväl etablerad kunskap, återgiven i bl.a. [Svensén \(2009\)](#), [Durkin \(2015\)](#) och [Jackson \(2022\)](#), som till de senaste rönen inom lexikalisk och metalexikografisk forskning, t.ex. resultat av användarstudier (se bl.a. [Hult 2016](#)).

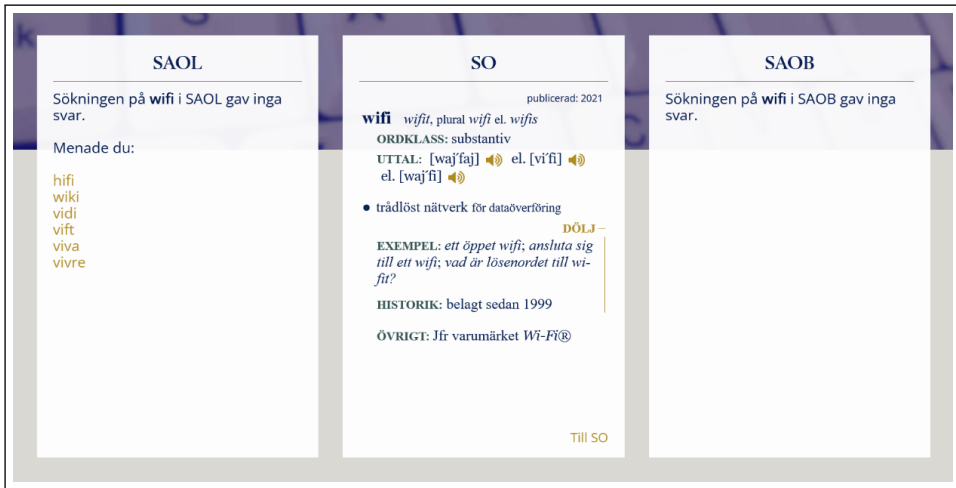
Ordboken bygger på en delmängd av en lexikalisk databas (på ca 200 000 ingångar) som förvaltas och vidareutvecklas vid Göteborgs universitet (GU). Arbetet med databasen påbörjades vid dåvarande Språkdata vid Göteborgs

universitet på 1970-talet i anslutning till projektet "Lexikalisk databas". Projektet hade två huvudsyften: dels att "etablera ett välstrukturerat lexikon lagrat i form av ett länkat nätverk (databas)", dels att "databasen skall läggas till grund för en ny svensk enspråkig definitionsordbok, Svensk ordbok" (Ralph m. fl. 1977: 3). Det mer teoretiska syftet var alltså lika centralt som det praktiska, att skapa en ordbok, vilket visar sig i den omfattande och teoretiskt genomarbetade dokumentationen av struktureringen av det lexikaliska innehållet. Ett exempel avser definitionsformatet i databasen: definitionen skulle vara fullständigt utbytbar mot definiendum, dvs. det ord som ska definieras, "i alla syntaktiska (och även morfologiska) kontexter, vid behov efter applikation av någon naturlig syntaktisk transformation" (Järborg 1989: 22). Detta krav medförde att olika ordklasser fick olika definitionsformat: ett substantiv skulle beskrivas med en obestämd nominalfras, ett adjektiv med en relativsats eller adjektivfras, ett verb med en infinitivfras etc. (se även Josephson 2022: 51, 254–255 om Lexikalisk databas och dess roll inom svensk språkvetenskap).

Den andra upplagan av SO från 2021 är en vidareutveckling av en rad lexikografiska verk vilka har publicerats från 1986 och framåt (se t.ex. Malmgren 1992, 2009a; Malmgren & Sköldberg 2013: 125–126 och Borin & Holmer 2024 för en tillbakablick över dessa; se även Holmer m. fl. 2015). Den andra och senaste upplagan publicerades, till skillnad från tidigare upplagor, endast digitalt, i form av gratis nedladdningsbara appar och på Svenska Akademiens ordboksportal, Svenska.se. Se t.ex. Sköldberg (2022, 2024) för redogörelser för viktigare redaktionella ställningstaganden inför denna upplaga samt Trap-Jensen (2022) för en recension av verket.

Sedan år 2010 ägs orddatabasen som SO baseras på av Svenska Akademien och arbetet med denna regleras av ett avtal mellan Svenska Akademien och GU. Den vidareutveckling av databasen, som GU ska ombesörja, består bl.a. i att nya ord, ordformer och ordförbindelser ska införas i databasen så att SO kan utges i bokform och/eller i elektronisk form minst vart tionde år. Vidare ska så fullständiga upplysningar som möjligt beträffande stavning, uttal, böjning, stilvärde och betydelse, illustrerade med språkexempel, ges om alla viktiga ord och uttryck i det svenska ordförrådet, och information ska lämnas om ordens fraseologi och konstruktionssätt. Detta nödvändiggör även en översyn av befintliga ordboksartiklar. Det lexikografiska revideringsarbetet sker i Karp, Språkbanken Texts dataredigeringsplattform, i vilken det är möjligt att söka i och redigera olika slags lexikografiska resurser (se vidare Borin, Sköldberg m. fl. 2025 om plattformen).

I ordboksportalen Svenska.se kan man söka i Svenska Akademiens tre ordböcker, det mer normativa verket *Svenska Akademiens ordlista* (SAOL 14 2015), SO (2021) och den historiska ordboken *Svenska Akademiens ordbok*



Figur 1: SO-artikeln *wifi* i Svenska.se vid sökning i Svenska Akademiens tre ordböcker på Svenska.se

(SAOB 1898–2023) på samma gång (se vidare om Svenska.se i bl.a. Lönnroth 2018 och Bäckerud m. fl. 2020). I portalen kan man även söka i ett verk i taget och användarna har därmed också tillgång till en renodlad SO-webbsida. En fördel med just webbversionen av SO är att den är lättillgänglig för skribenter som arbetar vid datorn, men appversionerna har flera fördelar, t.ex. bättre och fler sökfunktioner liksom funktionen "Dagens ord". I figur 1 återges (något beskuret) webbsidans gränssnitt och sökresultatet för substantivet *wifi*. SO-artikeln återfinns i mittenspalten. Som synes är det aktuella uppslagsordet (ännu) inte upptaget i SAOL eller i SAOB.

Av figur 1 framgår att beskrivningen av uppslagsordet *wifi* är förhållandevis kortfattad. I fråga om formell information anges enbart *en* stavning. Däremot varierar ordet böjningsmässigt i de material som redaktionen konsulterat, vilket lett till att två pluralformer ges (*wifi* eller *wifis*). Som synes återfinns ett "el." (dvs. "eller") mellan formerna, vilket ska tolkas som att formerna är i princip likvärdiga (se s. X i inledningen till SO (2009); se även avsnitt 5.1).

Uttalet ges både i skrift och med länkade ljudfiler med inläst uttal. Som påpekas av bl.a. Sköldberg (2017: 126) är det, om man jämför med uppgifter om exempelvis ords stavning och böjningsmönster, svårare för lexikografen att bilda sig en uppfattning om hur ord uttalas. Detta beror inte minst på bristen på större svenska talspråskorpusar. SO-lexikograferna har här gjort bedömningen att ordet *wifi* uttalas på tre sätt i svenskan: [waj'faj] el. [vi'fi] el. [waj'fi].

När det kommer till syntagmatisk information meddelas att ordet tillhör ordklassen substantiv. *Wifi* har också endast en betydelse, vilken förmed-

las med definitionen "trådlöst nätverk" och definitionstillägget "för dataöverföring" (i mindre stil) (se vidare om betydelsebeskrivningarna i SO i avsnitt 5.2). Artikeln innehåller därutöver tre kortare språkprov, varav det sista utgörs av en fråga: *ett öppet wifi; ansluta sig till ett wifi; vad är lösenordet till wifi?*. Målsättningen är att exemplen ska illustrera uppslagsordets fraseologi och fungera som ett komplement till definitionen. Någon pragmatisk information ges inte i artikeln, vilket tyder på att ordet bedöms vara omarkerat med avseende på bl.a. stil och bruklighet. (Se Svensén 2009: 315 för genomgång av olika sätt att markera att ett uppslagsord avviker från huvuddelen av de lexikala enheter som beskrivs.) Till skillnad från många andra ordboksartiklar i SO innehåller artikeln *wifi* inga hänvisningar till betydelsemässigt relaterade, uppslagsord. Slutligen, angående etablering, ursprung och släktskap, anges att ordet är belagt i skrivna svenska texter sedan 1999, men det finns ingen explicit uppgift om var det första belägget har påträffats. Lite speciellt i just denna ordboksartikel är också upplysningen att uppslagsordet i hög grad sammanfaller med det registrerade varumärket *Wi-Fi* (se vidare bl.a. Svensén 2009: 428–429 och Agerbo 2018: 198–202 om varumärken ur ett lexikografiskt perspektiv).

I SO (2021) har redaktionen utnyttjat många av den digitala ordbokens fördelar och bl.a. utvecklat den interna sökstrukturen i form av ett stort antal nyttillagda klickbara länkar. Dessa underlättar för användarna att röra sig inom ordboken. De böjningsformer som anges är också fullt utskrivna. (I tidigare upplagor hade man, för att förkorta, ersatt ordets uppslagsform med ett tilde och skrivit "*~t*, plural *wifi* el. *~s*".) Det kan dock konstateras att SO ännu är starkt präglad av det tryckta formatet. I en elektronisk ordbok, och då i synnerhet i en webbordbok, behöver lexikografen inte använda sig av samma komprimeringsstrategier som i en tryckt ordbok, men många konventioner som utvecklades i samband med utarbetandet av tryckta ordböcker håller i sig ännu och kan vara svåra att frigöra sig ifrån (se vidare bl.a. Rundell 2015; se även Trap-Jensen 2002 och Holmer 2022: 76, 85–88). I artikeln *wifi* är det främst de länkade ljudfilerna med inläst uttal som tydliggör att det rör sig om en digital resurs. Men också huvudordet i definitionen, "nätverk" hade kunnat utgöra en klickbar länk till uppslagsordet *nätverk* så att ordboksanvändare vid behov lätt hade kunnat se vad det ordet betyder. Vidare hade vissa förkortade ord kunnat skrivas ut. Fler böjningsformer hade även kunnat presenteras, t.ex. genitivformerna *wifis* och *wifits*.

Den senaste upplagan når via Svenska.se ut till en bredare publik än de tryckta föregångarna. Dagens SO-användare har troligen skiftande erfarenheter av att utnyttja ordböcker och fler användare än tidigare torde utgöras av svenskinlärare.

I och med att SAOL och SO, som båda är samtidsordböcker, visas bredvid varandra i portalen har även deras respektive funktioner diskuterats inom den projektgrupp som står bakom dem båda. Målet har blivit att renodla SO och dess perspektiv. Målsättningen har också lett till att *SO* (2021) är mer deskriptiv än exempelvis 2009 års upplaga (se vidare [Sköldberg 2017, 2024](#)).

Värt att notera är emellertid att det är svårt att tala om renodlat deskriptiva eller renodlat normativa ordböcker. Det är snarare fråga om en gradskillnad dem emellan. [Sköldberg \(2017: 124\)](#) konstaterar att en normativ ordbok alltid bygger på en deskriptiv grund och att en deskriptiv ordbok alltid representerar lexikografens språkliga värderingar. Också deskriptiva ordböcker har således en normativ dimension (jfr även [Svensén 2009: 24](#)). [Trap-Jensen \(2002: 64–65\)](#) påpekar också att oavsett den enskilda ordbokens deskriptiva eller normativa inriktning, så betraktar användarna ofta det som behandlas i en ordbok som det korrekta.

3 *Resurser och verktyg vid lexikografiskt arbete*

Beskrivningar i ordböcker baseras på evidens som samlas ur stora mängder text, korpusar. I skrivande stund (september 2024) har ordboksprojektet vid Göteborgs universitet via Språkbanken Texts samlingar tillgång till bl.a. mer än 16 miljarder token med moderna texter. Tillgången till material är högst relevant med tanke på att SO, sedan starten, i stor utsträckning är korpusbaserad (se vidare [Malmgren 2009a: 16](#)). Lexikografen behöver ha tillgång till omfattande korpusar med olika slags texter från olika tidsepoker för att kunna hitta nya ord, undersöka ovanliga ord, studera ords spridning, analysera ords formella och innehållsmässiga utveckling, för att nämna några aspekter. Innehållet i textmaterialen ligger också till grund för morfologiska och syntaktiska språkexempel i ordboken. För att lexikografen ska kunna hantera all denna data och se mönster, krävs det olika slags språkteknologiska metoder och verktyg. Sådana metoder och verktyg leder till att det lexikografiska arbetet kan effektiviseras och bli mer objektivt.

I det följande beskrivs Korp och Kungliga bibliotekets Svenska tidningar med betoning på de delar som är viktigast för SO-redaktionens arbete.

3.1 *Korp – Språkbanken Texts ordforskningsplattform*

Som redan antytts spelar innehållet i olika slags referensmaterial sedan länge en avgörande roll inom lexikografisk praktik. För svenskans del är innehållet i ordforskningsplattformen Korp ytterst viktigt (se [Borin, Forsberg m. fl. \(2025\)](#) för en presentation av plattformen). Korp innehåller textsamlingar,

korpusar, med text från exempelvis dagstidningar, skönlitteratur och sociala medier, vilka är uppmärskade med bl.a. språkliga attribut. Det moderna materialet i Korp spänner över 1900- och 2000-talet (<https://spraakbanken.gu.se/verktyg/korp/anvandarhandledning>). Där finns även delkorpusarna Kubord (se avsnitt 3.2). Det empiriska material som legat till grund för flertallet äldre ordböcker har i huvudsak bestått av tryckt och redigerat skriftspråk i form av tidningstext, men även en del romantext. Det är emellertid mycket viktigt att lexikograferna har tillgång till många olika slags texter – såsom de har via Korp – för att de ska kunna ge en så rättvisande bild av hur språket faktiskt används (se vidare Holmer 2022: 85–91 om val av textunderlag vid ordboksarbete; se även Forsberg & Holmer 2024).

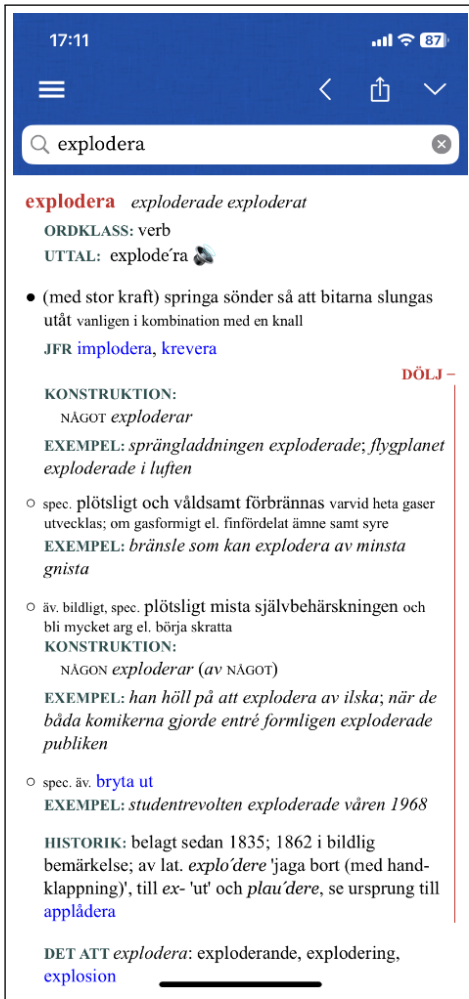
I Korps sökgränssnitt finns tre visningslägen: KWIC-konkordansvyn, Statistikvyn och Ordbildsvyn (Borin, Forsberg m. fl. 2025). Som utgångspunkt för en presentation av dessa tre visningslägen (och deras respektive användningsområden) väljer vi verbet *explodera*. Hur den aktuella ordboksartikeln ser ut i SO (2021) framgår av figur 2.

Av ordboksartikeln i figur 2 framgår bl.a. att ordet *explodera* böjs *exploderade* och *exploderat* i preteritum respektive i supinum. Vidare har ordet följande huvudbetydelse: '(med stor kraft) springa sönder så att bitarna slungas utåt'. Definitionen följs av definitionstillägget 'vanligen i kombination med en knall'. Parentesen runt "med stor kraft" betyder att just denna betydelsekomponent kan vara aktuell när ordet används men att det inte alltid är så. Därtill visas tre underbetydelser: 'plötsligt och våldsamt förbrännas (...)', 'plötsligt mista självbehärsningen (...)' och 'bryta ut'. Samtliga underbetydelser anges utgöra betydelsespecialiseringar i förhållande till huvudbetydelsen. Såväl den andra som den tredje underbetydelsen är därtill bildlig.

Vad kan man då få fram för information om verbet *explodera* i Korp? Enligt Atkins & Rundell (2008: 104) är KWIC-konkordanser centrala inom korpusbaserad lexikografi. En fördel med konkordanser är att man kan sortera dem på vänster- och högerkontext. De gör det också möjligt för lexikografen att överblicka ords kontexter och därigenom kunna urskilja återkommande mönster för hur ord används.

I figur 3 återges ett konkordansutdrag ur delkorpusen SVT Nyheter (2022) (på drygt 13 miljoner token; [Språkbanken Text u. å.](#)) i Korp. Utdraget gäller alla ordformer med grundformen *explodera* (med sammanlagt 288 belägg i materialet).

KWIC-vyn visar bl.a. olika böjningsformer av ordet *explodera*, främst infinitiv, preteritum och supinum. Vidare återfinns flera av de betydelser som verbet anges ha i SO. Exempelvis är huvudbetydelsen aktuell i ett flertal av konkordansraderna, t.ex. i "Husbil exploderade i Älvsjö" (rad 11), "[...] en kraftig bomb exploderade [...]" (rad 24) samt "Roboten exploderade strax



Figur 2: Artikeln *explodera* i SO (2021) (app för iOS)

efter att den avfyrats [...]” (rad 25). Att just huvudbetydelsen är så vanlig i denna korpus kan eventuell kopplas till att materialet består av nyhetstexter. En av underbetydelseerna är aktuell i fall som ”Vilket fick San Siro-publiken att explodera av glädje.” (rad 1) och “[...Anthony Modeste] stormade fram och nickade in 2-2, vilket fick hemmaarenan att explodera.” (rad 3).

Bland konkordansraderna finns det också en bildlig användning av *explodera* som inte behandlas i SO. Se t.ex. rad 16 med meningen ”Utbudet har exploderat på Hemnet, från 430 objekt för en månad sen till 610 i Uppsala [...]”. Verbet *explodera* kan då sägas betyda ’öka kraftigt’. En granskning

SVT NYHETER 2022	
Vilket fick San Siro-publiken att	explodera av glädje.
Mycket var bra han gjorde fina galoppombyten men	exploderade sen lite.
ste stormade fram och nickade in 2-2, vilket fick hemmaaren att	explodera .
kunde inte resa mig på grund av smärtan och mina ben höll på att	explodera i klättringen, säger van Vleuten.
Efter två mållösa perioder	exploderade Buffalo Sabres med sex mål i den sista perioden.
- Vi har fått uppleva hur damfotbollen har	exploderat på riktigt och vi har blivit förebilder och idoler - och kunnat driva vikt
även Roope Hintz med sitt tredje mål för matchen, varpå publiken	exploderade och kastade in sina huvudbonader på isen som hyllning till finländare
- Det	exploderade utan att jag hade räknat med det, säger han.
Den verkar ha haft gasolflaskor som kan ha	exploderat .
Vittnesuppgift: Svart väska	exploderade
• Husbil	exploderade i Älvsjö
Flaskan ska sedan ha	exploderat så att de fick ämnet över sig, vilket ledde till hudrodnad och klåda.
Flera medier rapporterar att vattenskotern	exploderade, men enligt polisen finns det inga uppgifter om detta.
- Det har	exploderat sedan årsskiftet.
Sen	exploderar rutan och en bit av en vägskylt tränger in i framrutan.
Utbudet har	exploderat på Hemnet, från 430 objekt för en månad sen till 610 i Uppsala, säger
- Jag blev så glad inombords och det kändes som om något	exploderade eftersom jag blev så glad när vi sågs.
Rasmus bil	exploderade på parkeringen: " Helvetet släpptes lös "
- Än så länge vet vi inte vad som	exploderat , men
ppsdelar fattas, säger Rasmus Sundström, som satt i bilen när den	exploderade i lördags i Enköping.
- För fem år sedan fick det en egen vägg i butiken, efter det har det	exploderat , säger Henry Brandt, butikschef på Partyland i Gränbystaden.
- Det var ett 20-tal elever och lärare vid stugan när det	exploderade och en gasbehållare började brinna.
Det rör sig troligtvis om en ballistisk robot som	exploderade relativt kort efter att den avfyrats, på en höjd under 20 kilometer, upp
56 människor har omkommit efter att en kraftig bomb	exploderade i en shiamuslimsk moské i provinshuvudstaden Peshawar i nordvästr
Roboten	exploderade strax efter att den avfyrats, enligt Sydkoreas militär.

Figur 3: Del av konkordansutdrag för *explodera* i alla ordformer i SVT Nyheter 2022 i Korp

av fler användningar av verbet i samma korpuser visar att det finns fler fall där ordet tycks betyda detsamma, bl.a. "Skulderna har exploderat i storlek [...]" och "svenskarnas poddlyssnande exploderar". Detta faktum talar för att ännu en bildlig användning av ordet ska inkluderas i nästa upplaga av ordboken.

Enligt Statistikvyn är det ordformen *exploderade* som är vanligast förekommande i delkorpuser. Den här typen av resultat kan ligga till grund för beslut om vilken eller vilka böjningsformer som kan vara lämpliga att använda i ordbokens språkexempel för att dessa ska spegla bruket av ordet så bra som möjligt. Via denna vy kan man relativt lätt också jämföra bruket av olika stavningsvarianter av ett och samma ord och, om man söker på ordet som för- eller efterled, studera vilka sammansättningar och avledningar som ett ord ingår i.

Även Ordbildsvyn är ett viktigt verktyg för lexikografen. I en ordbild visas det sökta ordet tillsammans med andra ord, vilka ordet har olika syntaktiska relationer till, i den aktuella korpuser. För ett verb visas bl.a. listor över återkommande subjekt och objekt och för ett substantiv visas vanliga adjektiv och verb. Ordbildsvyn bidrar till att lexikografen bättre kan överblicka vad

explodera (verb)

subjekt	explodera	objekt	adverbial
1. bomb	16	1.	526
2. petflaska	6	2. bombladdningar	2
3. gasflaskor	6	3. tok	1
4. intresse	10	4. smällare	1
5. flaskbomber	4	5. arena	1
6. efterfrågan	6	6. gång ²	1
7. flaska	3		
8. krutförråden	2		
9. techsektorn	2		
10. poddlyssnande	2		
11. flaskbomb	2		
12. fyrverkeriraket	2		
13. bilbomber	2		
14. megaakvariet	2		
15. militärflygbas	2		
		1. i luft	6
		2. i luft ²	6
		3. vid trappuppgång	3
		4. i huvudstadsregionen	2
		5. i kramkalas	2
		6. till skrotdelar	2
		7. i mångmiljon-popularitet	2
		8. i förbränningsugnarna	2
		9. på vägmina	2
		10. i inferno	2
		11. vid företag	2
		12. gasflaskor	2
		13. på hemnet	2
		14. år	8
		15. till lägenhetshus	2

Figur 4: Ordbild av verbet *explodera* i delkorpusen SVT Nyheter 2022 i Korp.

som förekommer i det sökta ordets kontext, vilket är nödvändigt för att hen ska kunna åstadkomma en så rättvisande bild av ordets syntagmatiska egenskaper som möjligt. Men ordbilder kan även tydliggöra olika betydelser hos ett ord (se avsnitt 5.2). I figur 4 återges en ordbild av verbet *explodera* (baserad på korpusen SVT Nyheter 2022).

Av figur 4 framgår bl.a. att *bomb*, *petflaska* eller *gasflaskor*, men också *intresse* och *efterfrågan* är vanliga subjekt till *explodera* i den aktuella korpusen. De tre första orden kan kopplas till huvudbetydelsen i SO och de två senare till den ännu inte beskrivna underbetydelsen. Vidare listas vanligare adverbial som t.ex. *i luft* och *vid trappuppgång*. Mittenkolumnen visar att verbet oftast inte tar objekt (de övriga fallen synes vid en närmare kontroll vara felparsade subjekt).

Ur lexikografisk synpunkt kan flera av de ord som uppträder i ordbilden för SVT Nyheter 2022 vara med i SO-artikeln *explodera*, men för att beskrivningen av det aktuella uppslagsordet ska bli så heltäckande som möjligt krävs det att flera olika slags texter, bl.a. nyhetstexter, skönlitteratur och texter publicerade på sociala medier, ligger till grund för ordbildsvisningarna (se vidare om Korps ordbild ur ett lexikografiskt perspektiv i [Sköldberg 2022](#)).

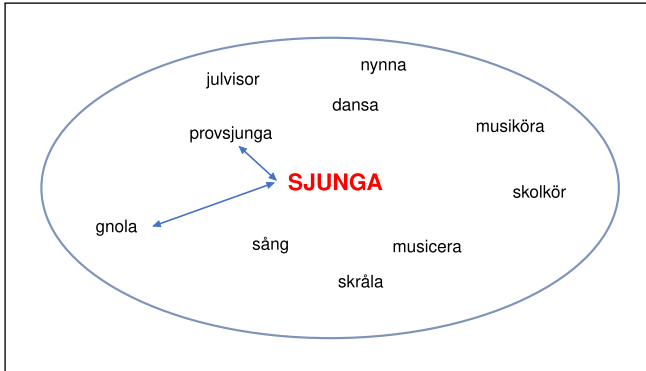
3.2 KB:s Svenska tidningar och Kubord

Ytterligare en viktig resurs i det lexikografiska arbetet är de tidningsmaterial som tillhandahålls av Kungliga biblioteket (KB). Tjänsten Svenska tidningar innehåller digitaliserade tidningar och tidskrifter från 1645 till idag. I dagsläget (september 2024) rör det sig om material från närmare 2 000 olika tidnings- och tidskriftstitlar.

SO-lexikograferna söker i KB:s tidningsmaterial främst vid datering av nya uppslagsord, dvs. för att finna första belägg i skrift. Med tiden har det också blivit uppenbart att många av de befintliga angivelserna i SO som rör uppslagsordens förstabelägg kan revideras efter slagningar i KB:s material, och det beror på att den allmänna tillgången till skrivna material är en helt annan idag än under tidigt 1990-tal då många av årtalen skrevs in i SO-databasen. Inte sällan uppmärksammas redaktionen av intresserade ordboksanvändare på att en angivelse om första belägg kan revideras. Att uppdatera alla uppslagsord på denna punkt är en mycket omfattande uppgift och den prioriteras inte just nu inom ordboksprojektet. Men några av de uppslagsord som kommer att få uppdaterade årtalsuppgifter i nästa upplaga av SO är *barbecue* (nuvarande förstabelägg i SO: 1989; uppdaterat förstabelägg: 1868), *espresso* (nuvarande angivelse: ca 1960; uppdaterat förstabelägg: 1924), *illustrativ* (nuvarande förstabelägg: 1906; uppdaterat förstabelägg: 1863) och *tsunami* (nuvarande förstabelägg: 1973; uppdaterat förstabelägg: 1942).

Av upphovsrättsliga skäl är tillgången till de tidningstexter som KB förvaltar begränsad, och detta vållar bekymmer för lexikografer och andra språkforskare som undersöker det svenska ordförrådet, liksom för exempelvis historiker och samhällsvetare. KB-labb vid Kungliga biblioteket och Språkbanken Text har av detta skäl tagit sig an uppgiften att skapa datasamlingar baserade på samtida tidningstexter som kan stödja ordforskning utan att för den sakens skull kränka upphovsrätten. Samarbetet mellan KB-labb och Språkbanken Text har bland annat resulterat i datasamlingarna Kubord 1 och Kubord 2, som är fritt tillgängliga via Korp. Därtill är ännu en datasamling, Kubord-fasttext, på väg att bli publik (se bl.a. [Forsberg & Holmer 2024](#) om dessa samlingar).

Kubord 1 bygger på drygt 80 delsamlingar med olika pressmaterial, sammanlagt lite mer än 3 miljarder token. Tidningsmaterialen är publicerade mellan åren 2000–2021 i bland annat nationella morgon- och kvällstidningar liksom i landsortspress. På grund av upphovsrätten kan Kubord 1 inte göras sökbar på samma sätt som andra korpuser i Korp. Det går att slå upp ord, men förekomsterna visas inte i konkordansformat. Som redan framgått är tillgången till ords kontext viktigt vid lexikografiskt arbete (se avsnitt 3.1).



Figur 5: Vektorrymd med några grannar för verbet *sjunga* (exemplet är hämtat ur Forsberg & Sköldberg, under utgivning)

Det sammanhang som orden används i behövs för att lexikografen ska kunna avgöra ett ords betydelser och fastslå i vilka konstruktioner ordet typiskt uppträder i. I Kubord 1 kan man dock få bl.a. detaljerad frekvensinformation om de ord man vill undersöka och detta räcker långt i samband med vissa lexikografiska uppgifter. Exempelvis kan man, på basis av Kubord 1, undersöka likheter och skillnader i ordförrådet mellan två årgångar av en och samma tidning. Sådana jämförelser är användbara vid det nyordsarbete som bedrivs inom ordboksprojektet (se avsnitt 4.1). Innehållet i Kubord 1 kan även fungera som stöd för utmönstring av uppslagsord (se avsnitt 4.3).

Kubord 2, som i stora drag innehåller samma material som Kubord 1, ger användaren tillgång till statistik över enstaka ord men också över ordpar som står i syntaktisk relation till varandra, såsom verb–subjekt eller substantiv–attribut. Denna kontextuella information möjliggör ordbilder för de ord som förekommer i korpusarna.

Den tredje Kubord-resursen, dvs. Kubord-fasttext, spelar en viktig roll i samband med SO-redaktionens studier av ordvektorer i det lexikografiska arbetet (se Forsberg & Sköldberg 2022, Bouma m. fl. 2024 och Forsberg & Sköldberg under utgivning). Arbetet med ordvektorer ger information om ord vars språkliga kontexter liknar varandra. På så sätt kompletterar ordvektorerna de uppgifter lexikografen får fram via konkordanser och ordbilder. Figur 5 illustrerar en vektorrymd för verbet *sjunga*. På bilden återges ett urval ord som hamnar nära, eller som man säger, utgör grannar.

Bland grannarna i en vektorrymd återfinns till exempel ofta semantiskt besläktade ord till det ord som studeras. Dessa grannord kan ligga till grund för fler hänvisningar till synonyma, antonyma och kohyponyma ord inom

ordboksartiklarna. Studier av vektorrymder kan därmed leda till tillägg i ordboken som förtydligar kopplingarna mellan dess uppslagsord. Bland grannarna finns det också många sammansättningar som kan tjäna som morfologiska språkprov i ordboksartiklar. Därtill kan ordvektorer ringa in olika slags semantiska fält och ge information om ords värdeladdning. Bland grannarna till verbet *sjunga* återfinns sammansättningar som *provsjunga*, mer eller mindre synonyma ord som *skråla*, och ord som kan föras till samma semantiska fält, t.ex. *musicera*. Där finns även för detta verb typiska objekt som *sång*. De aktuella grannarna står således i olika relation till det undersökta verbet. Grannarnas avstånd från *sjunga* ger också en indikation om graden av likhet i fråga om ordens kontext. Grannen *provsjunga* ligger närmare *sjunga* än ett annat verb, *gnola*. Det innebär att de sammanhang där *provsjunga* används liknar mer de sammanhang där *sjunga* förekommer, jämfört med de kontexter där *gnola* används (se bl.a. Bouma m. fl. 2024 och Forsberg & Sköldberg under utgivning för fler exempel).

I nuläget sträcker sig Kubord-materialen som sagt fram till och med år 2021, men materialen kommer successivt att kompletteras med nyare tidningstexter, vilket är avgörande för ordforskare som inriktar sig på bl.a. nyord.

4 Tre övergripande uppgifter i arbetet med SO

Det redaktionella arbetet med SO kan delas in i tre huvuduppgifter: 1) nyordsjakt, urval av uppslagsord och författande av nya ordboksartiklar, 2) revidering av befintliga ordboksartiklar och 3) utmönstring av ordboksartiklar.

4.1 Nyordsjakt, urval av uppslagsord och författande av nya ordboksartiklar

SO-redaktörerna hyser av förståeliga skäl stort intresse för nya svenska ord. Med nya ord menas här ord som skapats och spridits i svenska texter på senare år, och därmed kan sägas ingå i svenskan. Ett exempel är substantivet *lårskav* ('skavsår på lårets insida orsakat av långvarig friktion') som används allt mer sedan 2018.¹ Med nya ord avses också (fack)ord som på senare tid ökat i användning i allmänspråket, t.ex. växtbeteckningen *jätteloka* som återfinns i svenska texter sedan åtminstone 1961. Nya svenska ord kan antingen bli uppslagsord eller fungera som språkexempel i form av sam-

1 Ordet är med på Språkrådets nyordslista för året 2018 (se <https://www.isof.se/download/18.1e4309991774c3fe50b6ea40/1612877399497/Nyordslistan%202018.pdf>; se även Karlsson 2021). Ordet behandlas sedan år 2021 i SO.

mansättningar eller avledningar i ordboken. Utöver den här typen av ord är redaktörerna också intresserade av äldre ord som av någon anledning inte fångats upp och införlivats i tidigare upplagor av SO, t.ex. *torrvara*, som är belagt i svenska tidningar åtminstone sedan 1730-talet och som nu är inlagt i SO-databasen inför nästa uppdatering. Nyare ord intresserar för övrigt inte bara lexikografer utan även andra språkvetare, språkteknologer, historiker och samhällsvetare, eftersom ett språks ordförråd och dess utveckling säger något om det omkringliggande samhället och hur det förändras.

För att bli varse nya ord som kan läggas in i databasen använder sig medarbetarna inom ordboksprojektet av en metod som i korthet går ut på att ordförrådet i en viss korpus jämförs med ordförrådet i en annan korpus och så tar de fasta på de ord som finns i den ena textsamlingen men inte i den andra. De ord som är nyskapade eller har ökat i bruk jämförs sedan med de ord som redan behandlas i ordboken och så får redaktionen en lista över s.k. nyordskandidater att välja ur (se [Holmer m. fl. 2024](#)). En jämförelse t.ex. mellan de ord som används i Göteborgs-Posten år 2020 och 2021 visar bland annat att följande ord har ökat i användning mellan de två åren: *barnfridsbrott, charkbricka, fucking, gastropub, hjärndimma, höghöjdsbana, incel, intimitetskoordinator, kulturtolk, miso, mockumentär, mobildata, nagelsalong, naturhus, nettonollutsläpp, näringsterapeut, parallellsamhälle, powerwalk, preklinisk, prosecco, reduktionsplikt, sexsomni, skidalpinism, skyddsperson, streetfood, trygghetspension, utsläppssiffra, verkställighetshinder* och *yes*. Många av dessa ord illustrerar hur flertalet svenska ord bildas. De säger även något om vår samtid, om förändringar i samhället, om nya vetenskaper och ny teknik, om mer eller mindre lyckliga händelser etc. (se [Karlsson 2021](#)).

Kubord-materialen kan även jämföras med innehållet i andra slags texter i Korp. Ett flertal väletablerade, men ofta mer informella, svenska ord som ännu inte finns med i SO, har tydliggjorts genom jämförelser mellan Kubord och Poeter.se, en webbsida där skribenterna kan ladda upp sina egna alster. Exempel på sådana ord är *bokmärka, healer, lost, photoshoppa, repeat* och *vänförfrågan*. I arbetet med SO, som ska beskriva modern svenska, är det viktigt att bruket av också mer vardagliga ord undersöks så att även de orden kan komma i fråga som uppslagsord.

Vid själva urvalet av uppslagsord beaktas i hög grad aspekter som frekvens i allmänspråket och spridning i olika texter. När det kommer till sammansättningar spelar ordens grad av semantisk genomskinlighet en viktig roll. Lexikaliserade sammansättningar tas med i SO, men mer genomskinliga sammansättningar behandlas bara i begränsad omfattning (se t.ex. s. IX i inledningen till [SO 2009](#)). Användningen av de ord som väljs ut analyseras grundligt och de aktuella orden förses sedan, i enlighet med avtalet mellan GU och Svenska Akademien, med en rad upplysningar (se avsnitt [2.2](#)).

4.2 Revidering av befintliga ordboksartiklar

Som redan framgått har innehållet i SO utvecklats under decennier. En mycket stor andel av SO:s artiklar tillkom på 1980-talet inför den första upplagan av verket och dessa artiklar behöver av olika skäl bearbetas. En anledning är att språket har förändrats sedan ordboksartiklarna skrevs. Exempelvis kan, som vid fallet *explodera*, en betydelse ha tillkommit. Ett ytterligare sådant exempel är *tänka* som, troligen under inflytande från engelskan, numera också betyder 'tycka' (se t.ex. [Blensenius 2015](#)). Ett annat skäl är att dagens lexikografer har tillgång till större korpusar och till korpusar med olika slags språk. Exempelvis kan större korpusar med texter från sociala medier innehålla stavningsvarianter som man inte finner i mindre korpusar baserade på redaktionella texter. Här kan förhållandet mellan stavningsvarianterna *spray* och *sprej* nämnas. I tidigare upplagor, inklusive [SO \(2009\)](#), finner man angivelsen "**sprej** el. **spray**", men i [SO \(2021\)](#) har varianternas inbördes ordning ändrats. Angivelsen "**spray** eller **sprej**" är helt enkelt mer rättvisande när det gäller hur ordet stavas i olika texter av idag. En tredje orsak till att äldre ordboksartiklar behöver revideras är att SO nu enbart är digital och att redaktionen så långt det är möjligt vill utnyttja det digitala formatet och bl.a. skapa tydligare samband inom ordboken med hjälp av hänvisningar i form av klickbara länkar.

För att säkra SO:s kvalitet är det nödvändigt att närma sig ordboks innehållet ur flera perspektiv. Granskningen av materialet kan med fördel ske genom 1) läsning av ordboksartiklar från bokstäverna A–Ö, 2) närstudier av en informationskategori i taget (t.ex. uttal, böjning, bruklighetskommentar) och 3) tematiska genomgångar av olika domäner (juridik, medicin etc.). Som exempel kan nämnas att redaktionen inför [SO \(2021\)](#) gjorde en omfattande genomgång av många tusen konstruktionsuppgifter, som finns typiskt i verbartiklar (men även i ett stort antal substantiv- och adjektivartiklar) och som har till uppgift att ange hur uppslagsord konstrueras med andra ord (se [Blensenius 2019](#)). En annan omfattande genomgång gällde ord som kan föras till ämnet pedagogik. Mellan olika upplagor har det omgivande samhället, inklusive det svenska utbildningssystemet, förändrats. Dessa förändringar kan också avläsas i språkbruket såsom det återspeglas i bl.a. Språkbanken Texts korpusar. Ett konkret fall är uppslagsordet **rektor**. Jämför hur substantivet beskrivs i [SO \(2009\)](#) respektive i [SO \(2021\)](#) (de båda artiklarnas innehåll återges i exempel 1 och 2):

- (1) **rektor** [...]
 • [...] (titel för) högste ledare för skola eller högskola {JFR **studierektor**}: *rektorsexpedition; bråkstakarna skickades till ~; ett enskilt samtal med ~*
 KONSTR.: ~ (för el. på el. vid ngt)
 HIST.: sedan 1612; av lat. *re'ctor* 'styresman'; till **regera** (SO 2009)
- (2) **rektor** [...]
 • [...] (titel för) högsta administrativa och pedagogiska ledaren för skola, universitet, högskola eller annan utbildningsinstitution
 JFR **studierektor**
 SAMMANSÄTTN./AVLEDNING: *rektorsexpedition; rektorsutbildning*;
 KONSTRUKTION:
 ▶ *rektor* (för/på/vid NÅGOT)
 EXEMPEL: *eleven bad om ett enskilt samtal med rektor; som rektor för skolan har hon både personal- och budgetansvar; en biträdande rektor med ansvar för den övergripande pedagogiska planeringen*
 HISTORIK: belagt sedan 1612; av lat. *re'ctor* 'styresman'; till **regera** (SO 2021)

Av ordboksartikeln från 2021 framgår det bl.a. att rektorer med tiden har fått förändrade arbetsuppgifter och att avståndet mellan rektor och elever har minskat.

Ytterligare en specialsatsning gjordes på s.k. kontroversiella ord med koppling till olika diskrimineringsgrunder. Några sådana ord är *hora, blatte, rödskinn, funktionsvariation, miffo, bög* och *hedersrelaterad* (se t.ex. [Pettersson & Sköldberg 2020a,b](#)). Arbetet gick bl.a. ut på att revidera uppsättningen med uppslagsord, att omformulera betydelsebeskrivningar och att modernisera språkprov så att de blev mer könsneutrala. Till arbetet kom även att uppdatera bruklighetskommentarer i artiklarna, t.ex. "kan uppfattas som nedsättande" och "stark nedsättande". [Pettersson & Sköldberg \(2020a: 112\)](#) lyfter dock de begränsningar som innehållet i korpusar och andra textdatabaser har i just detta sammanhang:

En relevant fråga [...] är hur lexikografer hittar de ord som är – eller kan uppfattas som – diskriminerande i den databas som de arbetar med och ska revidera. [...]. [S]pråket utvecklas snabbt och ett ord som tidigare varit förhållandevis neutralt kan fort bli kontroversiellt [...]. En central fråga är också hur lexikografen avgör vad ett ord har för laddning. Det kan finnas källor att utgå från, såsom t.ex. RFSL:s ordlista, men de passar inte alltid lexikografens syfte.

Det är alltså uppenbart att innehållet i korpusar utgör ett bättre stöd i samband med vissa lexikografiska uppgifter, såsom vid angivelse av ett ords böjning, än vid andra arbetsuppgifter.

4.3 Utmönstring av uppslagsord

Ordförrådet utvecklas hela tiden och det är naturligt att vissa av orden i olika upplagor av SO, inklusive den andra från 2021, inte längre hör till det samtida svenska ordförrådet. Dessa ord riskerar att mönstras ut. Förteckningar med så kallade utmönstringskandidater utarbetas genom att listan med uppslagsord i SO jämförs med ordförrådet i korpusar över samtida språk i Korp. Den uppsättning korpusar som SO-redaktionen på senare tid valt att jämföra uppslagsorden med består av bl.a. tidningstexter och nyhetstexter från SVT (båda materialen är från åren 2010–2021), innehållet i den svenska versionen av Wikipedia samt texter publicerade på sajten Poeter.se med användargenererat innehåll. Allt som allt innehåller korpusen ca 3 miljarder token. En jämförelse mellan uppslagsorden och ordförrådet i korpusarna visar att SO innehåller ett stort antal ord som inte förekommer – i någon av deras respektive böjningsformer – i de aktuella korpusarna (jfr [Holmer m. fl. 2024](#) för liknande jämförelser mellan förteckningen över uppslagsord i SAOL och innehållet i samma textmaterial). Exempel på sådana uppslagsord är *kabinettspolitik*, *kakelugnsnisch*, *kakelugnssättare*, *kallhus*, *kalstammig*, *kardioskleros*, *kartusianerorden*, *kassaköp*, *katalysatorrenad*, *katetrisera*, *kautschukparagraf* respektive *raderkniv*, *radiokol*, ¹*rallentando*, ²*rallentando*, *realkatalog*, *reprografi*, *rimsalta*, *rousseauan*, *rönnbärsfilosofi*.

En mer kvalitativ genomgång av listan med utmönstringskandidater visar att många av dessa utgörs av beteckningar på föremål, företeelser, aktiviteter etc. som är ovanligare, knappt används eller är inaktuella i det svenska samhället av idag (t.ex. *kakelugnssättare* och *raderkniv*). Bland utmönstringskandidaterna finns också många fackord, exempelvis det medicinska ordet *kardioskleros* liksom de homografa musiktermerna *rallentando* (adverb och substantiv). Eventuellt ska ord som dessa inte vara med i allmänspråkliga SO utan listas i fackordböcker av olika slag (jfr avsnitt 2.1).

En annan lexikografisk arbetsuppgift är att granska det inte obetydliga antal ord som är försedda med bruklighetsbeteckningarna "något ålderdomligt" och "ålderdomligt". I SO (2021) finns det ca 1 000 ordboksartiklar som innehåller etiketten "något ålderdomligt", t.ex.: *barnförlamning*, *bröstsocker*, *grosshandlare*, *hemlighus*, *judekörs*, *malträtera*, *mentalhygien*, *sedlighetsbrott*, *själöspilling*, *stilettkäpp*, *tankomat*, *tattare*, *tvivelsjuk*, *törhända*, *vanartig*, *vidlåda*, *översiggiven*. Därtill finns det ca 600 ordboksartiklar med etiketten "ålderdomligt", exempelvis *dårhus*, *fjärdingsväg*, *huskors* och *rucklare*. Även

bland dessa ord finns beteckningar på äldre föremål och produkter, t.ex. *bröstsocker* och *stilettkäpp*. Det finns också beteckningar (t.ex. *barnförflamning* och *judekörs*) som av olika anledningar kommit att ersättas av andra svenska beteckningar (i detta fall *polio* och *japansk lykta*).

Många uppslagsord i SO skulle därmed kunna utmönstras men frågan är om det är nödvändigt. Eftersom också nästa upplaga enbart blir digital finns det, åtminstone i teorin, oändligt med plats i också den (se t.ex. [Rundell 2015](#)). Detta innebär att flertalet ord kan vara kvar i SO, även om de är mindre typiska för modernt språkbruk. Å andra sidan, och med tanke på visningen av de tre ordböckerna i Svenska.se, är det inte självklart att synkrona SO, som ska spegla samtida svenska, behöver redogöra för ord som är utförligt beskrivna i SAOB, t.ex. *antvarada*, *banesår*, *jungfrudom*, *kormoran*, *kruserlig*, *ostärkt* och *paltig*. Användare som är intresserade av ord som inte längre används kan förutom att konsultera SAOB söka i online-resursen SAOLhist.se vilken bl.a. innehåller [SO \(2009\)](#).

5 *Två informationskategorier i fokus*

I det följande resonerar vi kring två relativt olika informationskategorier i SO och hur lexikografen använder olika empiriska material och verktyg i samband med arbetet med dessa.

5.1 *Information om uppslagsordens böjning*

I SO redovisas böjningsformer på följande sätt för böjliga ord (se [SO 2009: XII](#)): för substantiv ges i regel bestämd form singular och obestämd form plural (t.ex. *hunden hundar*), för verb i regel preteritum- och supinumformerna (t.ex. *hoppade hoppat*) och för adjektiv vanligtvis neutrum- och pluralformerna (t.ex. *gult gula*). Sedan finns det en rad avvikelser från huvudmönstret. Många substantiv är mer eller mindre oböjliga, och då anges "ingen böjning", t.ex. för *fanders*, enstaka verb böjs ogärna i vissa former, t.ex. *stinga*, där upplysningen "preteritum undviks" ges, och vissa adjektiv böjs ogärna i neutrum, t.ex. *rigid*, vilket också anges i böjningsinformationen.

Vissa ord böjs oregelbundet (t.ex. verb som *gå gick gått* och adjektiv som *bra bättre bäst*), men dessa utgör sällan några utmaningar i ordboksarbetet eftersom det främst är gamla ord som kan föras till denna kategori. Böjningsmönstren kan emellertid förändras över tid även för befintliga ord, som när svaga böjningsmönster tar över från starka böjningsmönster. Ett exempel på detta är mönstret *simma sam summit* som konkurrerar med det i dag betydligt frekventare *simma simmade simmat* (att *sam summit* är mindre

frekventa former markeras i SO av att de föregås av "äv."). Regionala varianter som preteritumformen *lös* till *lysa* är också ett exempel på ord där böjningsförändringar över tid är möjliga och därmed särskilt viktiga att kunna följa.

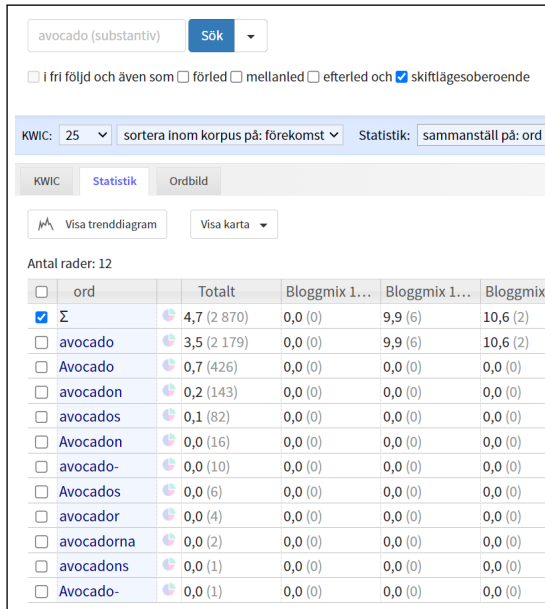
För att ta reda på hur ordens böjningsmönster ser ut i faktiskt språkbruk används Korp och då främst Statistikvyn på ordformsnivå. De korpusar som konsulteras är framför allt nyhetskorpusar, men även skönlitteratur, akademisk text och vissa sociala medier. Eftersom orden ofta är lemmatiserade i lexikonresursen Saldo (se t.ex. [Borin & Holmer 2024: 44–45](#)) och därmed innehåller samtliga brukliga böjningsmönster, räcker det ofta att söka på en viss grundform med alla ordformer i en viss korpus. I statistikvisningen ges både frekventa former och ibland eventuella variationer i bruket. SO har som nämnts en deskriptiv inriktning (se avsnitt 2.2), vilket medför att frekvens i hög grad styr hur eventuella variantformer presenteras: om en viss pluralform är frekventast, kommer den att redovisas först.

Utöver Korp kan andra textdatabaser konsulteras för kontroll av specifika ordformer, exempelvis Retriever/Mediearkivet, där de senaste tidnings- och tidskriftstexterna finns representerade och sökbara. Här finns t.ex. även facktidsskrifter och (transkriberade) radioprogram representerade, där man kan göra stickprov.

Något lexikografen ibland funderar på i arbetet med böjningsangivelser är ord som böjs på flera olika sätt i bruket. Hur många varianter ska då ges i en mer deskriptiv ordbok som SO, och hur ska dessa förhålla sig till dem i den mer normativa ordlistan SAOL? Ett exempel är uppslagsordet *papper*, för vilket SAOL i artikelhuvudet ger böjningsformerna *papperet* el. *pappret*; plural *papper*, bestämd plural *papperen* el. *pappren*. I SO redovisas fler böjningsformer än så, i alla fall för bestämd plural: *pappret* el. *papperet*, plural *papper*, bestämd plural *pappren* el. *papperna* el. *papperen* el. *papprena*. Hur många alternativa böjningsformer som ska ges och hur de ska redovisas i förhållande till SAOL är fortfarande en öppen fråga.

En annan kategori av ord som böjs på flera olika sätt utgörs av ord som ansluter till flera böjningsmönster, framför allt sådana som har inspirerats av andra språk. Ett sådant fall utgör s-pluralen, som i *enchilada-s* och *wrap-s*. Pluraländelsen berör framför allt nyare ord. Om man skulle lägga in ord som *boomer*, *pinata/piñata* och *youtuber* i SO, skulle man få anledning att ta reda på hur dessa ord böjs i bruket. Här kan man ta hjälp av t.ex. ordformsfrekvenser i Kubord-korpusarna. Det verkar som om s-pluralen dominerar klart om man söker efter ord som börjar på t.ex. "boom" och "pinat". I fallet *pinata/piñata* används förvisso *pinator/piñator* också i viss mån.

Ett annat intressant exempel är de ord där en viss böjningsform sticker ut frekvensmässigt. Att pluraler som *britter* och *pengar* tenderar att vara



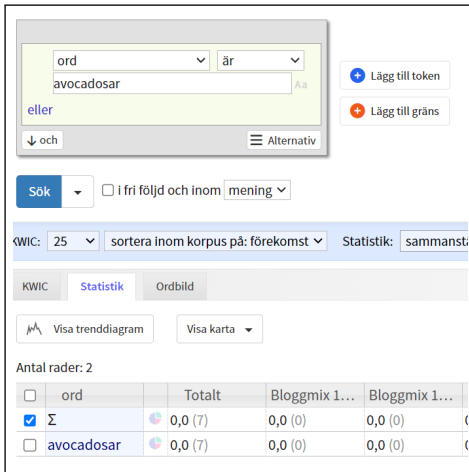
Figur 6: Statistik över sökträffar för *avocado* i alla ordformer som ges i Korp (korpusen Bloggmix)

frekventare än uppslagsformerna *britt* och *peng* är välkänt och anges med anmärkningen "vanligen plural" i SO. [Blensenius & Martens \(2020\)](#) finner emellertid fler användningar av pluralformer för substantiv som avviker i förhållande till användningen av grundformen, i deras fall i ett material bestående av diskussionsforum. Pluralen *räkor* är där betydligt frekventare än *räka*, *trakter* klart frekventare än *trakt* och *vittnesutsagor* klart frekventare än *vittnesutsaga*. Denna typ av uppgifter går såklart att få fram genom statistik i Korp, men det hade varit bra att integrera någon form av enkel frekvensvisning av ordformer vid varje uppslagsord i redigeringsgränssnittet.

Lexikografen behöver vidare vara medveten om att ordformer kan saknas i ett visst böjningsparadigm för en grundform i Korp. Då kan hen behöva söka efter specifika ordformer. Ett exempel är den vardagliga pluralformen *avocadosar/avokadosar* (till *avocado/avokado*), som inte återfinns genom sökning efter *avocado/avokado* i någon böjningsform. I figur 6 ges exempel med sökning efter *avocado* i alla ordformer som inkluderats i böjningsparadigmet.

Det finns också andra sätt att söka efter potentiella ordformer, t.ex. att sökuttrycket "Börjar med" *avocado* i Utökad sökning. En sökning efter specifikt *avocadosar* i samma korpus visar t.ex. att *avocadosar* förekommer 7 gånger i korpusen (se figur 7).

Fall som *avocado/avokado*, som inte uppvisar någon större stamvariation,



Figur 7: Statistik över sökträffar för ordformen *avocadosar* i Korp (korpusen Bloggmix)

är alltså lätta att söka fram på olika sätt i Korp, men vissa former kan vara svårare att få fram. Ett exempel på en sådan form är den regionala preteritumvarianten *lös* till *lysa*, som togs upp ovan, som inte verkar finnas i *lysa:s* böjningsparadigm i Korp (*lysa* och *lysa*² i Saldo).

5.2 Information om uppslagsordens betydelser

Svenséns lexikografiska handbok (Svensén 2009) innehåller ett fylligt kapitel om hur betydelsebeskrivningar i olika ordböcker kan se ut. I boken finns också utförlig information hur själva arbetet med att fastlägga betydelser går till och vad som kan påverka den betydelsebeskrivning som lexikografen väljer att använda (se även t.ex. Atkins & Rundell 2008: 263–315). Svensén (2009: 205–211) konstaterar bl.a. att innan lexikografen ger sig i kast med att beskriva ett uppslagsords betydelse måste hen ta ställning till om ordet ska anses ha en enda betydelse eller flera delbetydelser. Enligt författaren grundar sig detta ställningstagande i regel på korpusdata, vilket bidrar till att betydelsebeskrivningarna blir mer precisa och rättvisande. Olika slags kriterier, såsom morfologiska, syntagmatiska, semantiska, paradigmatiska och pragmatiska, styr också om analysen resulterar i att det aktuella uppslagsordet tilldelas en eller flera betydelser.

Lew (2013: 287) lyfter å sin sida fram två motsatta strategier, kända som *lumping* och *splitting*, vad gäller betydelseindelning i enspråkiga ordböcker. Den första strategin går ut på att minimera antalet betydelser så att var och en

av dem täcker så stort semantiskt innehåll som möjligt. Den andra strategin tenderar att resultera i ett större antal finindelade betydelser. Lexikografens val av strategi beror ofta på den typ av ordbok som hen arbetar med och ordbokens målgrupp (Atkins & Rundell 2008: 267–268). Den semantiska beskrivningen i en ordbok som främst vänder sig till inlärare på nybörjarnivå kännetecknas ofta av lumping medan den semantiska beskrivningen i en ordbok som vänder sig till lingvister etc. kännetecknas av splitting.

Svensén tar också upp de två olika sätt som betydelserna hos polysema ord kan struktureras. Enligt författaren (2009: 211–212) kan de olika betydelserna ordnas *linjärt*, dvs. som ett antal diskreta enheter i en sekvens, eller *hierarkiskt*, som ett antal huvudbetydelser, till vilka grupper av underbetydelser är associerade. För att förtydliga kan ett ord som *aptitretare* nämnas. Detta ord har dels en mer bokstavlig, dels en mer bildlig betydelse. När polysemistrukturen beskrivs linjärt i ordboken är dessa två betydelser listade som betydelse 1 och betydelse 2 och när polysemistrukturen är hierarkisk listas de två betydelserna som betydelse 1 och betydelse 1a eftersom den andra betydelsen anses vara en bildlig underbetydelse till den första.

I SO ordnas betydelserna, i enlighet med riktlinjerna i Ralph m. fl. (1977) och Järborg (1989), hierarkiskt. I ordboken specificeras dessutom förhållandet mellan huvudbetydelserna och deras respektive underbetydelser (i termer av betydelseutvidgning, betydelsespecialisering, metaforisk användning etc.). I SO är analysen och distinktionen mellan olika betydelser relativt "finkornig". SO-lexikograferna är med andra ord snarare splitters än lumpers.

Svensén (2009: 214ff.) tar därtill upp olika typer av ordförklaringar, t.ex. lexikografiska definitioner och synonymförklaringar. Som redan framgått innehåller bl.a. Ralph m. fl. (1977) och Järborg (1989) information om definitionsformat i SO (se avsnitt 2.2) och de riktlinjer som uttrycks där är i många avseenden giltiga än idag. Men innehållet i skrifterna är inte heltäckande. Exempelvis finns det inget skrivet om hur interjektioner ska beskrivas (se Landqvist & Sköldberg 2024). Den nuvarande SO-redaktionen försöker följa traditionen och påtalade principer, men väljer ibland att frånga vissa direktiv bl.a. för att förenkla, inte minst med tanke på nya användargrupper av SO och deras behov.

Eftersom de semantiska uppgifterna i SO är så centrala utgör en viktig del av ordboksarbetet att undersöka om betydelserna hos uppslagsorden har förändrats – och i så fall *hur*. Lexikografen upptäcker att den befintliga beskrivningen behöver modifieras, att nya betydelser behöver läggas till m.m. i samband med olika arbetsmoment, t.ex. vid undersökning av nyordskandidater. Som exempel kan nämnas att det bland de senaste nyordskandidaterna till SO finns ett antal sammansättningar som slutar på substantivet *ring*, bl.a.

pedofilring, *narkotikaring* och *reko-ring*. Slagningar i korpusar visar också på fraser i vilka samma betydelse hos *ring* är aktuell, t.ex. *en ring av pedofiler*. Denna betydelse, som kan ha sitt ursprung i engelskans *circle* och som kan jämföras med en betydelse hos ordet *krets*, är inte registrerad i SO sen tidigare. En figurativ underbetydelse till ordet *ring* behöver läggas till i ordboken.

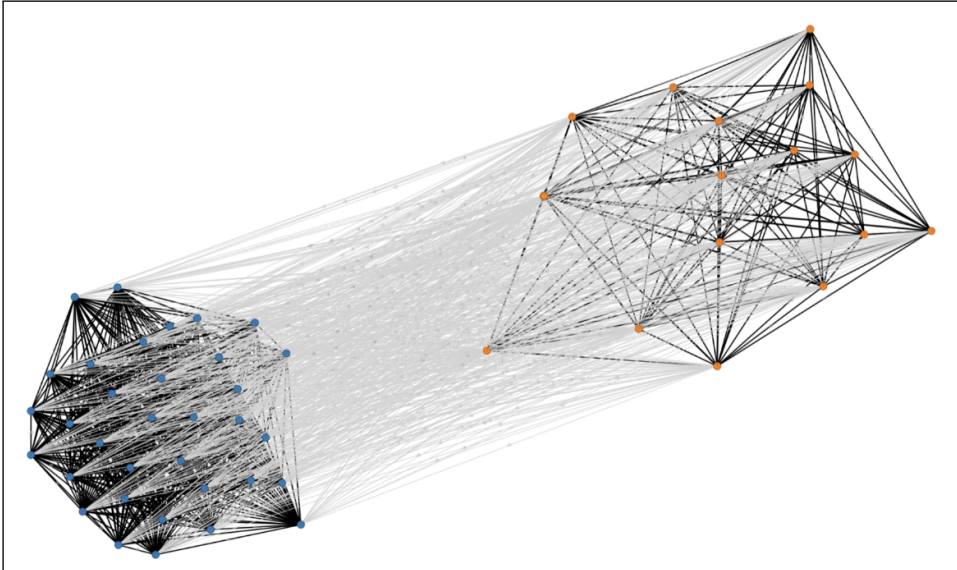
Vidare kan lexikografen bli varse nya betydelser när hen använder sig av Ordbildsvyn i jakt på kollokationer och andra återkommande ordkombinationer. Ett exempel är uppslagsordet *riktmärke* som enligt SO (2021) betyder 'märke som man tar sikte på för att hålla rak kurs t.ex. vid segling'. En ordbildsvisning i Kubord 2 visar att det bland de vanligare framförställda attributen till ordet finns adjektiv som *vetenskaplig* och *ideologisk*. Bland de efterställda attributen finner man ord som *asylmottagande*, *gruppsstorlek* och *löneökning*. Dessa ord indikerar att ordet *riktmärke* också används i bildlig bemärkelse, vilket stöds av sökningar i andra korpusar. I exempelvis SVT 2020, som erbjuder KWIC-visningar, uppträder ordet i bl.a. följande meningar (vår fetstil):

- (3) - Det är två meters avstånd som är **riktmärket** för att kunna minska smittspridningen.
- (4) Redan 2016 kom Skolverket med ett **riktmärke** på 6–12 barn för grupper med 1–3-åringar.
- (5) Enligt Trafikverket kommer de arbeten som utförs dagtid inte heller att överskrida **riktmärken** för buller- och stomljud i närliggande bostäder.

Allt talar för att substantivet, utöver sin nuvarande betydelse, kommer att ha en bildlig underbetydelse i nästa upplaga av SO.

Andra slags närläsningar bär också frukt. Exempelvis har flertalet interjektionsartiklar reviderats som en följd av en jämförelse mellan innehållet i befintliga interjektionsartiklar och hur interjektioner används i mer tal-språkna texter i Korp (se Landqvist & Sköldberg 2024). Undersökningarna har bl.a. lett till att flera interjektioner i nästa upplaga kommer att beskrivas utifrån den eller de funktioner som de fyller istället för med traditionella definitioner med vars hjälp man försöker återge interjektionens betydelse (se även Svensén (2009: 241)). Ett exempel på en sådan interjektion är *heja* som i nuvarande upplaga har betydelsebeskrivningen 'må det gå bra för viss deltagare i tävling eller dylikt'. I nästa upplaga av ordboken kommer uppslagsordet istället att följas av beskrivningen "uttryck för önskan om framgång för viss deltagare i tävling etc."

Redaktionen uppmärksammas också på svagheter i betydelsebeskrivningarna av ordbokens användare. Ett exempel på detta är den användarkom-



Figur 8: DUREls semantiska kluster för adjektivet *enkelspårig*

mentar som gällde uppslagsordet *fruga* med betydelsen 'hustru till viss man'. Liksom många andra ord som försetts med mer inkluderande betydelsebeskrivningar inför [SO \(2021\)](#) behöver också denna betydelse omformuleras så att den blir mer könsneutral (se [Petersson & Sköldberg 2020a](#)).

Därtill är SO:s redaktionsmedlemmar delaktiga i studier som rör datorstödda metoder för att upptäcka semantisk förändring på ett systematiskt sätt. Bl.a. har [Sköldberg m. fl. \(2024\)](#) genomfört olika slags experiment i anslutning till annoteringsverktyget DUREl som använder sig av språkmodeller för automatisk semantisk analys av ord (se vidare [Schlechtweg m. fl. 2018, 2024](#) om verktyget). I DUREl kan man studera ett ords betydelser genom att titta på den semantiska likheten mellan olika meningar som innehåller ordet. Programmet skapar en visualisering med kluster av lika meningar, som är tänkta att motsvara fokusordets eventuella olika betydelser. Inom just dessa experiment är det SVT-korpusen (cirka 200 miljoner tokens; [Språkbanken Text u. å.](#) med texter publicerade mellan 2004 och 2021 som använts).

För att illustrera hur gränssnittet i DUREl fungerar kan adjektivet *enkelspårig* nämnas. I SO har detta ord dels en huvudbetydelse ('som har bara ett spår'), dels en bildlig underbetydelse ('ensidig, ytlig'). Efter att ha laddat upp 50 slumpmässigt valda meningar ur SVT-korpusen som innehåller ordet i DUREl, utfördes alla ytterligare steg i verktygets gränssnitt. Figur 8 visar de två resulterande DUREl-klustren för *enkelspårig*, ett med blå prickar och

ett annat med orange prickar, motsvarande de två betydelserna hos detta uppslagsord i SO.

I DUREls front-end-gränssnitt kan man, genom att klicka på de blå prickarna, se korpusbelägg som "Den enkelspåriga järnvägen mellan Motala och Hallsberg är idag en flaskhals [...]". Detta exempel illustrerar sålunda huvudbetydelsen hos adjektivet. Om man klickar på de orange prickarna finner man korpusbelägg som "De tror att vi är enkelspåriga lantisar, de tror att vi är trångsynta, att vi är rasister och homofober." Här är således den bildliga betydelsen, som motsvarar underbetydelsen i SO, relevant. Sammanfattningsvis har DUREl i detta fall lyckats peka ut de två betydelserna hos ordet.

I ett specifikt experiment utgick [Sköldberg m. fl. \(2024\)](#) från de uppslagsord i SO som endast har en betydelse. Därefter studerade de hur många betydelsekluster av respektive ord som DUREl skapade baserat på användningar i SVT-korpusen. Två eller flera semantiska kluster kan tyda på att ordet har mer än en betydelse i texterna och att den semantiska beskrivningen i SO bör revideras. Sådana sökord kan prioriteras för manuell inspektion vid revideringsarbetet inför nästa upplaga. Ett exempel från detta experiment är substantivet *lydnad*, för vilket DUREl upptäckte två kluster. Verktuget indikerar med andra ord att substantivet *lydnad* har minst två olika betydelser vilket sedan bekräftades av lexikograferna som inspekterade klustren och korpusutdragen manuellt. De fann att det ena klustret inkluderade korpusexempel som "Nunnor följer de tre klosterlöftena: att leva i celibat, fattigdom och lydnad." Detta exempel illustrerar betydelsen som redan finns nedtecknad i SO. Det andra klustret innehåller däremot korpusexempel med ordkombinationer som "tävla i lydnad" och "nordiska mästerskap i lydnad". De representerar en betydelsespecialisering som är aktuell inom hund- och ridsport. Denna specifika betydelse av substantivet finns inte upptagen i den aktuella utgåvan av SO och kommer att läggas till i nästa uppdatering.

Genom studierna i anslutning till DUREl har flera insikter kunnat göras. Även om kvaliteten på SO generellt sett är hög, stödjer undersökningarna tidigare slutsatser av t.ex. [Blensenius m. fl. \(2021\)](#) som konstaterar att den semantiska sidan i ordboken kan vidareutvecklas. Experimenten är fortfarande småskaliga, men resultaten är ändå lovande. De datorstödda metoderna förväntas effektivisera och förbättra det lexikografiska arbetet på flera sätt. Hittills har arbetet bland annat uppmärksammat lexikograferna på underbetydelser som bör läggas till i ordboken. Även saknade huvudbetydelser har upptäckts. Arbetet har också visat att vissa betydelsebeskrivningar i SO är för generella och att de bör delas upp i enlighet med gällande principer för de semantiska beskrivningarna i ordboken.

6 Slutord

I detta kapitel har vi redogjort för den empiriska ordforskning som ligger till grund för det pågående lexikografiska arbetet inom ett ordboksprojekt vid Språkbanken Text, Göteborgs universitet. Arbetet inom projektet syftar bl.a. till att vidareutveckla samtidsordboken *Svensk ordbok utgiven av Svenska Akademien* (SO). Den senaste upplagan av verket publicerades år 2021 och arbete inför en uppdaterad version av ordboken pågår. I kapitlet har vi visat hur olika slags korpusar och verktyg, som tillhandahålls av Språkbanken Text, och andra slags textmaterial, vilka tillgängliggjorts av Kungliga biblioteket, används inom uppdateringsarbetet. Innehållet i olika textmaterial är centralt vid författande av nya ordboksartiklar, uppdatering av befintliga artiklar liksom vid utmönstring av artiklar som inte hör hemma i en mer deskriptiv samtidsordbok. Innehållet i olika slags korpusar är exempelvis avgörande för uppgiften att ge information om uppslagsordens böjning och semantik.

Arbetet med att välja ut lämpliga uppslagsord utgör en viktig uppgift för lexikografen. När orden har valts ut ska de, efter noggrann analys, förses med uppgifter om stavning, uttal, böjning, betydelse, kombinatoriska egenskaper, historik m.m. Dessutom ska språkexempel väljas ut, som visar hur uppslagsorden används. Men ordboksarbetet innefattar sålunda också revidering av befintliga artiklar och även det arbetet inbegriper en rad informationskategorier. Med andra ord aktualiseras flera olika språkvetenskapliga forskningsområden i samband med ordboksarbete. Men lexikografiskt arbete inbegriper också tvärvetenskapliga frågeställningar som vetter mot digitala humaniora. Exempelvis aktualiseras frågor kring behandling, analys och presentation av stora och komplexa datamängder.

Praktiskt lexikografiskt arbete är traditionsbundet och alla olika upplagor av SO följer vissa principer som redovisas i t.ex. [Ralph m. fl. \(1977\)](#) och [Järborg \(1989\)](#). Men dessa principer nedtecknades för flera decennier sedan och mycket har förändrats sedan dess, inte minst tillgången till olika slags textmaterial och språkteknologiskt baserade verktyg. Därtill har SO gått från att vara en tryckt ordbok till att vara en heldigital produkt. Det senare har i sin tur medfört att ordboken når ut till andra slags användare (med andra behov). Faktorer som dessa påverkar hur nuvarande redaktion väljer att förhålla sig till de principbeslut som en gång fattades. Exempelvis kan det leda till att lexikografen väljer att avvika från de en gång fastlagda definitionsformaten för att betydelsebeskrivningen ska bli mer begriplig.

Språkbanken Texts korpusar, liksom de tidningsmaterial som är tillgängliga via Kungliga biblioteket, har många styrkor. Materialen utgör ett utmärkt stöd i anslutning till vissa lexikografiska arbetsuppgifter, inte minst när lexikografen vill studera hur ett ord stavas eller böjs. Det är förhållandevis enkelt

att ställa eventuella varianter mot varandra och därigenom avgöra vilka former som ska upptas i ordboken och i vilken ordning de ska presenteras. Textmaterialen är också ovärderliga när lexikografen söker information om hur ord kombineras med varandra eller när hen vill finna konkreta exempel på hur enskilda ord eller uttryck används.

Däremot är det inte lika lätt att, utifrån de textmaterial som står till buds idag, avgöra vad ett ord har för emotiv laddning. Därtill kan bristen på talspråkskorpusar leda till svårigheter att hitta information om hur ord uttalas och hur utpräglat talspråksartade ord, såsom många interjektioner, används.

En sak står dock klar. Tack vara nuvarande korpusar i Språkbanken Text, de texter som är publika via Kungliga biblioteket samt moderna språkteknologibaserade metoder, har det lexikografiska arbetet med bl.a. SO blivit mer vetenskapligt baserat och effektivt. SO-redaktionen arbetar också aktivt för tillgängliggörande av nya textmaterial, metodutveckling och ökat samarbete mellan olika kompetenser för att den vetenskapliga kvaliteten på ordboksarbetet ska bli ännu högre.

Referenser

- Agerbo, Heidi. 2018. Funktionsteoretisk diskussion af brugerreaktioner på ordbogsartikler – med fokus på varemærker, omdiskuterede ord og eksempelsætninger. *LexicoNordica* 25. 195–214.
- Atkins, B. T. Sue & Michael Rundell. 2008. *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.
- Bergenholtz, Henning & Inger Bergenholtz. 2011. A dictionary is a tool, a good dictionary is a monofunctional tool. I Pedro A. Fuertes-Olivera & Henning Bergenholtz (utg.), *e-Lexicography: The internet, digital initiatives and lexicography*, 187–207. London/New York: Continuum.
- Blensenius, Kristian. 2015. Jag tänker att. *Språktidningen* 1. 58–61.
- Blensenius, Kristian. 2019. Revision av konstruktionsuppgifter i Svensk ordbok utgiven av Svenska Akademien. *LexicoNordica* 26. 203–223.
- Blensenius, Kristian & Monica Martens. 2020. Analys av relativa ordformsfrekvenser för en bättre ordbok. *Svenskans beskrivning* 37. 56–69.
- Blensenius, Kristian, Emma Sköldberg & Erik Bäckerud. 2021. Finding gaps in semantic descriptions: Visualisation of the cross-reference network in a Swedish monolingual dictionary. *Proceedings of the Electronic lexicography in the 21st century (eLex) 2021 conference*. 247–258.

- Borin, Lars, Markus Forsberg, Martin Hammarstedt, Louise Holmer & Arild Matsson. 2025. Korp: Språkbanken's word research platform. I *Sixty years of Swedish computational lexicography*. Dana Dannélls, Kristian Blensenius & Lars Borin (utg.). De Gruyter. 175–194. DOI: [10.1515/9783111577234-010](https://doi.org/10.1515/9783111577234-010).
- Borin, Lars & Louise Holmer. 2024. Tradita innovare, innovata tradere: The Gothenburg approach to computational lexicography. *Proceedings of the Huminfra Conference (HiC 2024)*. 41–50.
- Borin, Lars, Emma Sköldberg, Ann Lillieström, Nick Smallbone, Maria Öhrman, Jonatan Uppström & Louise Holmer. 2025. Karp: Språkbanken's data editing platform. I *Sixty years of Swedish computational lexicography*. Dana Dannélls, Kristian Blensenius & Lars Borin (utg.). De Gruyter. 195–210. DOI: [10.1515/9783111577234-011](https://doi.org/10.1515/9783111577234-011).
- Bouma, Gerlof, Markus Forsberg, Justyna Sikora & Emma Sköldberg. 2024. Konsten att bedriva svensk ordforskning utan att kränka upphovsrätten. *Proceedings of the Huminfra Conference (HiC 2024)*. 161–167.
- Bäckerud, Erik, Pär Nilsson & Emma Sköldberg. 2020. Så används Svenska Akademiens ordböcker på nätet: Implicit och explicit feedback från användarna. *Nordiska studier i lexikografi* 15. 91–101.
- Durkin, Philip (utg.). 2015. *The Oxford handbook of lexicography*. Oxford: Oxford University Press.
- Forsberg, Markus & Louise Holmer. 2024. Datatillgång, metodutveckling och lexikografiskt arbete vid Språkbanken Text. *LexicoNordica* 31. 61–79.
- Forsberg, Markus & Emma Sköldberg. 2022. Ordvektorer i lexikografiskt arbete. I Elena Volodina, Dana Dannélls, Aleksandrs Berdicevskis, Markus Forsberg & Shafqat Virk (utg.), *Live and learn: Festschrift in honor of Lars Borin*, 37–41. Gothenburg: Department of Swedish, Multilingualism, Language Technology, University of Gothenburg.
- Holmer, Louise. 2022. *Neutrala substantiv på -ande i text och ordbok*. Gothenburg: Meijerbergs institut för svensk etymologisk forskning.
- Holmer, Louise, Ann Lillieström, Emma Sköldberg & Jonatan Uppström. 2024. Time to say goodbye revisited: On the exclusion of headwords from the Swedish Academy Glossary (SAOL). *Proceedings of the European Association for Lexicography (EURALEX) 2024*. 443–452.
- Holmer, Louise, Monica von Martens & Emma Sköldberg. 2015. Making a dictionary app from a lexical database: The case of the Contemporary Dictionary of the Swedish Academy. *Proceedings of the Electronic lexicography in the 21st century (eLex 2015) conference*. 32–50.
- Hult, Ann-Kristin. 2016. *Ordboksanvändning på nätet: En undersökning av användningen av Lexins svenska lexikon*. Gothenburg: University of Gothenburg. (diss.).

- Jackson, Howard (utg.). 2022. *The Bloomsbury handbook of lexicography*. 2. utg. London: Bloomsbury Academic.
- Josephson, Olle. 2022. *Språkpolitik*. 2. utg. Stockholm: Morfem.
- Järborg, Jerker. 1989. *Betydelseanalys och betydelsebeskrivning i Lexikalisk databas: Preliminär version*. (Research Reports from the Department of Swedish No. GU-ISS-89-01) Gothenburg: Department of Swedish, University of Gothenburg.
- Karlsson, Ola. 2021. *Lesserwisser, lårskaw och läppstiftseffekt: Presentation och problematisering av urvalskriterierna för den svenska nyordslistan*. *LexicoNordica* 28. 101–120.
- Landqvist, Hans & Emma Sköldberg. 2024. Interjektioner som lexikografisk utmaning: En fallstudie av interjektioner med engelskt ursprung utifrån Svensk ordbok utgiven av Svenska Akademien. *ASLA:s skriftserie/ASLA Studies in Applied Linguistics* 31. 26–55.
- Lew, Robert. 2013. Identifying, ordering and defining senses. I Howard Jackson (utg.), *The Bloomsbury companion to lexicography*, 284–302. London: Bloomsbury Publishing.
- Lönnroth, Harry. 2018. Portalen svenska.se: En ny digital samlingsplats för språkresurser från Svenska Akademien. *LexicoNordica* 25. 281–292.
- Malmgren, Sven-Göran. 1992. From Svensk ordbok ('A dictionary of Swedish') to Nationalencyklopediens ordbok ('The Dictionary of the National Encyclopedia'). *Proceedings of the European Association for Lexicography (EURALEX) 1992*. 485–491.
- Malmgren, Sven-Göran. 2009a. Från Nationalencyklopedins ordbok (1995–96) till Svensk ordbok utgiven av Svenska Akademien (2009). *LEDA-Nyt* (47). 14–20.
- Malmgren, Sven-Göran. 2009b. On production-oriented information in Swedish monolingual defining dictionaries. I Sandro Nielsen & Sven Tarp (utg.), *Lexicography in the 21st century: In honour of Henning Bergenholtz*, 93–102. Amsterdam: John Benjamins.
- Malmgren, Sven-Göran & Emma Sköldberg. 2013. The lexicography of Swedish and other Scandinavian languages. *International Journal of Lexicography* 26(2). 117–134. DOI: [10.1093/ijl/ect008](https://doi.org/10.1093/ijl/ect008).
- Petersson, Stellan & Emma Sköldberg. 2020a. Beskriva utan att diskriminera. Representation av könsidentitet och sexuell läggning i Svensk ordbok. *LexicoNordica* 27. 97–115.
- Petersson, Stellan & Emma Sköldberg. 2020b. To discriminate between discrimination and inclusion: A lexicographer's dilemma. *Proceedings of the European Association for Lexicography (EURALEX) 2021*. 382–386.

- Ralph, Bo, Jerker Järborg & Sture Allén. 1977. *Svensk ordbok och lexikalisk databas: Förstudierapport*. Gothenburg: Department of Computational Linguistics, University of Gothenburg.
- Rundell, Michael. 2015. From print to digital: Implications for dictionary policy and lexicographic conventions. *Lexikos* 25. 301–322.
- SAOB. 1898–2023. *Svenska Akademiens ordbok*. Lund: Gleerups. <https://www.saob.se/%20and%20https://svenska.se/>.
- SAOL 14. 2015. *Svenska Akademiens ordlista*. 14. utg. Stockholm: Norstedts. svenska.se.
- Schlechtweg, Dominik, Sabine Schulte im Walde & Stefanie Eckmann. 2018. Diachronic usage relatedness (DUrel): A framework for the annotation of lexical semantic change. *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics / Human Language Technologies (NAACL/HLT) 2018: Volume 2 (Short Papers)*. 169–174. DOI: [10.18653/v1/N18-2027](https://doi.org/10.18653/v1/N18-2027). <https://aclanthology.org/N18-2027>.
- Schlechtweg, Dominik, Shafqat Virk, Pauline Sander, Emma Sköldbberg, Lukas Theuer Linke, Tuo Zhang, Nina Tahmasebi & Sabine Schulte im Walde. 2024. The DUrel annotation tool: Human and computational measurement of semantic proximity, sense clusters and semantic change. *Annual Meeting of the European Chapter of the Association for Computational Linguistics (EACL) 2024: System Demonstrations*. 137–149.
- Sköldbberg, Emma. 2017. Innehållet i Svensk ordbok utgiven av Svenska Akademien – eller kampen mellan norm och bruk. I Saga Bendegard, Ulla Melander Marttala & Maria Westman (utg.), *Språk och norm. Rapport från ASLA:s symposium*, 123–129. Uppsala: ASLA: Svenska föreningen för tillämpad språkvetenskap.
- Sköldbberg, Emma. 2022. Phraseological theory, evidence in corpora and lexicographical practice: On collocations in a monolingual dictionary of Swedish. I Kristian Blenselius (utg.), *Valency and constructions: Perspectives on combining words*, 155–182. Gothenburg: Meijerbergs institut för svensk etymologisk forskning.
- Sköldbberg, Emma. 2024. Andra upplagan av Svensk ordbok (SO): Förut-sättningar, teoretiska överväganden, insatser och mottagande. I Denny Jansson, Ida Melander, Gustav Westberg & Daroon Yassin Falk (utg.), *Svenskans beskrivning 38: Förhandlingar vid trettioåttonde sammankomsten. Örebro 4–6 maj 2022, Del I*, 165–180. Örebro universitet.
- Sköldbberg, Emma, Shafqat Virk, Pauline Sander, Simon Hengchen & Dominik Schlechtweg. 2024. Revealing semantic variation in Swedish using computational models of semantic proximity: Results from lexicographical experiments. *Proceedings of the European Association for Lexicography (EURALEX) 2024*. 169–182.

- SO. 2009. *Svensk ordbok utgiven av Svenska Akademien*. Stockholm: Svenska Akademien.
- SO. 2021. *Svensk ordbok utgiven av Svenska Akademien*. 2. utg. Stockholm: Svenska Akademien.
- Språkbanken Text. U. å. SVT nyheter. Dataset. DOI: [10.23695/vvd5-j848](https://doi.org/10.23695/vvd5-j848).
- Svensén, Bo. 2009. *A handbook of lexicography: The theory and practice of dictionary-making*. Cambridge: Cambridge University Press.
- Tarp, Sven. 2008. *Lexicography in the borderland between knowledge and non-knowledge: General lexicographical theory with particular focus on learner's lexicography*. Tübingen: Max Niemeyer Verlag.
- Trap-Jensen, Lars. 2002. Normering og deskription i Den Danske Ordbog – mere eller mindre? *LexicoNordica* 9. 63–78.
- Trap-Jensen, Lars. 2022. Svensk ordbok – anden og reviderede udgave. *LexicoNordica* 29. 197–213.

Korresponderande författare

Emma Sköldberg
Institutionen för svenska,
flerspråkighet och
språkteknologi
Göteborgs universitet
emma.skoldberg@svenska.gu.se

Kristian Blensenius
Institutionen för svenska,
flerspråkighet och
språkteknologi
Göteborgs universitet
kristian.blensenius@gu.se