



GÖTEBORGS UNIVERSITET
INST FÖR SVENSKA SPRÅKET

GU-ISS-2014-02

Korpblickar

– om möjliga ingångar med sökverktyget Korp
för studier av orden *tänka* och *tanke* i korpusar

Inga-Lill Grahn



Forskningsrapporter från institutionen för svenska språket, Göteborgs universitet
Research Reports from the Department of Swedish

ISSN 1401-5919

www.svenska.gu.se/publikationer/GU-ISS

Korpblickar – om möjliga ingångar med sökverket Korp för studier av orden *tänka* och *tanke* i korpusar

Användning av orden *tänka* och *tanke* i tre samtal har undersökts i Grahn (2012). Ett syfte med undersökningen var att pröva metoder för hur människans komplexa meningsskapande med ord kan undersökas och beskrivas empiriskt. En utgångspunkt var den senare Wittgensteins språkfilosofiska perspektiv som lyfter frågor om betydelse och om meningskapande i dialog med omvärlden. ”Hur ett ord fungerar kan man inte gissa. Man måste *se* på dess användning och lära därav” (Wittgenstein 1992 [1953]:punkt 340). I den refererade undersökningen är materialet begränsat till tre samtal, varför vidare studier är befogade.

Jag har i Grahn (2012) urskilt användningar av orden *tänka* och *tanke* i tre samtal från programmet *Radiopsykologen* i Sveriges Radio. Resultat och tolkning har gjorts med hänsyn till den givna kontexten, och som en hjälp i tolkningsarbetet användes termen *tänkare* för att beteckna en semantisk roll, samt termen *kommunikativt projekt* för att tolka ordens användning i den aktuella institutionella kontexten. Båda dessa termer har lyfts fram med utgångspunkt i en systemisk-funktionell analys av processer. Dels sådana processer där verbet *tänka* är kärna, dvs. huvudverb i en fras, dels sådana där substantivet *tanke* är deltagare, dvs. antingen subjekt eller objekt.

Syftet med föreliggande uppsats är att med utgångspunkt i resultaten från Grahn (2012) undersöka möjligheterna att bredda materialet såväl kvantitativt (fler förekomster av de aktuella orden) som kvalitativt (t.ex. fler olika textsammanhang, och eventuellt ytterligare aspekter).

Jag vill i föreliggande uppsats kartlägga hur korpusundersökningen i allmänhet och sökverket Korp i synnerhet kan användas för att bygga vidare på den kvalitativa undersökning som redan föreligger. Det är då särskilt intressant att leta efter aspekter som en liten kvalitativ studie inte kommer åt, exempelvis frekvenser.

En processtypsanlys utgår från verbet eftersom detta är kärnan i varje process, men det rör sig inte om en kategorisering av verb. Teoretiskt kan nämligen verbet tillhöra olika processtyper beroende på vilka deltagare som också ingår i processen. Därför är konstruktionen av deltagare vid olika användningar avgörande för analysen. Och följaktligen är också varje enskild användning av stort intresse. Med traditionell grammatisk terminologi och en smula förenklat kan man säga att det är relationerna mellan verb, subjekt och objekt som är föremål för studier, och då specifikt i konstruktioner där substantivet *tanke* och verbet *tänka* ingår (se Grahn 2012, för närmare beskrivning av processtypsanlys och teoretisk bakgrund).

Disposition

I framställningen nedan inleder jag med att i stora drag presentera de funktioner i sökverket Korp, som jag alltså använder mig av. Eftersom min egen kunskap om Korp till stor del inhämtats genom att jag gjort olika sökningar och försökt dra slutsatser om resultaten¹, bygger också presentationen nedan till stor del på dessa sökningar. Efter den mer övergripande presentationen av Korp går jag över till tre specifika frågeställningar som har med mitt ämne att göra och försöker reda ut i vilken mån dessa skulle kunna besvaras, eller

¹ Och till lika stor del har jag tagit hjälp av Johan Roxendal, systemutvecklare på Språkbanken, för att förstå vilka förutsättningar som gäller för korpusarbete och för språkteknologiska verktyg. Varmt tack!

åtminstone belysas med korpusmaterial. En del av presentationen utgörs av analyser av hur den automatiska taggningen fallit ut i de sökningar jag gjort.

Om Korp

Språkbanken vid Göteborgs universitet har enligt hemsidan² till uppdrag ”[...] att samla in, förädla och tillgängliggöra språkresurser åt forskare och allmänheten, samt att bedriva forsknings- och utvecklingsarbete som främjar sån användning.” Bland de språkresurser som ingår i Språkbankens kompetens finns textkorpusar (det vill säga stora mängder text som samlas elektroniskt), parallellkorpusar (korpusar med samma texter på flera språk), elektroniska lexikon samt språkteknologiska verktyg vilka på olika vis gör korpusarna tillgängliga för användarna.

Ett av de gränssnitt som gör det möjligt för forskare och för allmänhet att ställa frågor till de samlade korpusarna går under benämningen *Korp*. Detta gränssnitt gör det möjligt att på olika vis söka information i korpusarna; man kan säga att varje sökning är en ingång till ett avgränsat rum i textmängden. Först i dessa avgränsade rum kan informationen bli mer överblickbar och därmed användbar. Sökningarna blir möjliga genom att orden i materialet har ”taggats”, det vill säga försetts med olika etiketter som skiljer dem från varandra och kategoriserar dem. I en sökning kan de olika taggningarna särskiljas eller kombineras på olika vis. Resultatet av en sökning ges i form av textexempel med beläggen (så kallade konkordanser), i form av vanligt förekommande ordkombinationer (så kallade ordbilder) och i form av statistik över träffarna.

I sin helhet omfattar det material som tillgängliggörs via Korp i dagsläget (12 aug. 2012) 85 korpusar med totalt 919 240 967 token (ordförekomster och skiljetecken) och 58 109 096 meningar. Texterna omfattar såväl sak- som bruksprosa och litterära texter. Arbetet med att lägga in nytt och mer material i Språkbanken pågår ständigt, vilket har den fördelen att resurserna är aktuella. Samtidigt är detta lite problematiskt ur forskningssynpunkt eftersom det material man arbetar med kan komma att ändras under pågående studier. Rent tekniskt finns dock möjlighet att ”frysa” ett material för forskningssyften, men den möjligheten är inte öppen för vanliga användare. För mer detaljerad information om Språkbanken och om Korp, se hemsidan.

Om lemgram

Innan jag går vidare till användning av Korp ser jag anledning att presentera termen *lemgram* som förekommer som ett sökalternativ och som jag själv fick ägna en del arbete till att förstå. Man kan nämligen antingen göra sökningar på *ord* eller på *lemgram*. För en riktig förståelse behöver termen avgränsas i förhållande till de termer som används inom projektet Lexikalisk databas, nämligen termerna *lemma* respektive *lexem*. Nedan redogörs därför för de olika termerna, för hur de förhåller sig till varandra och för vissa avgörande skillnader mellan ordboksarbete och korpusarbete.

Ett lemma motsvarar en grundform eller en uppslagsform, det vill säga den ordform som är sökbar i en alfabetisk ordbok (så till exempel *katt* men inte *kattens* eller *dyka* men inte *dök*). (Förhållanden som gäller sökbarhet i ordböcker har delvis ändrats eftersom de nu ofta finns i elektronisk form. Tekniskt sett är det möjligt att göra fler ordformer sökbara i en elektronisk ordbok, men den kategorisering som används är fortfarande en lemmalista som motsvarar

² <<http://spraakbanken.gu.se/swe/information>>, hämtat den 12/8 2012

grundformer.) Grunden för indelningen i lemman är stavning, böjning, uttal och ordklass, det vill säga helt formella kriterier. Vad som utgör ett lemma eller inte avgörs således på formella – inte semantiska – grunder.

I tillägg till denna formella kategorisering görs även en semantisk, eftersom ord som till formen är identiska kan ha olika betydelser. Grunden i den semantiska kategoriseringen är lexemen, som vart och ett representerar en huvudbetydelse. Ett lemma kan således omfatta flera lexem, huvudbetydelser, men bara under förutsättning att samtliga har samma böjningsmönster. Ordformen *glas* t.ex., förekommer därför i NEO som ett lemma (*glas* subst. -et, plur. -en, best. plur. -en) som har tre lexem (1) ämne 2) dryckeskärl 3) halvtimme). Ordformen *bok* däremot, utgör två lemman; grundformen *bok* är visserligen identisk i bägge men böjningsmönstret skiljer sig (*bok* subst. *boken*, *böcker* respektive *bok* subst. -en, -ar).

Sammanfattningsvis, ett lemma ”representerar klassen av samhöriga böjningsformer, stavningsformer och uttalsformer” (se NE uppslagsordet *lemma*). Med ”samhöriga” menas här att de kan härledas till en gemensam grundform (ett lemma).

Vad är då ett lemgram? Är lemgrammet att likställa vid ett lemma, det vill säga en formell kategorisering? I användarhandledningen till Korp förklaras termen såhär: ”Ett lemgram är ett ord eller ett flerordsuttrycks samtliga böjningar, och gör det möjligt att i en och samma sökning söka efter både "katt", "katter", "katterna" och så vidare.” Det vill säga att en sökning på lemgrammet *katt* resulterar i träffar även på samtliga böjningsformer. Det är också möjligt att söka på ordet (inte lemgrammet) *katt* och få ett resultat med träffar enbart på just formen *katt*.

När det gäller termerna lemma och lexem är de kopplade till varandra på så sätt att det framgår i ordboken hur många och vilka lexem som är kopplade till ett enskilt lemma. På det här sättet är pappersordboken exemplariskt överskådlig. Om lemmat skulle motsvaras av ett lemgram i Korp, skulle lexemen motsvaras av det som kallas för *betydelser*, och som är hämtade från den semantiska databasen SALDO. När man söker på ett ord direkt i SALDO får man mycket riktigt information om vilket eller vilka lemgram det tillhör och vilka betydelser som är kopplade till lemgrammen. Informationen är dock inte alldeles lättillgänglig. I exemplen nedan presenteras två sökningar i SALDO, den första på ordet *tank* och den andra på ordet *glas*.

tank

sammansättningsanalys

tank	behållare	tank (nn)	korpu
tank ²	stridsvagn	tank ² (nn)	korpu

glas

sammansättningsanalys

glas	material + genomskinlig	glas (nn)	korpus
glas ²	dricka + ur	glas (nn)	korpus
glas ³	halvtimme	glas (nn)	korpus

Vi ser att varje sökning ger en uppställning i fyra kolumner med rubriken sammansättningsanalys. I den första kolumnen anges de betydelser som är kopplade till sökordet på var sin rad, i fallet med *tank* två betydelser och i fallet med *glas* tre stycken. Betydelserna är numrerade och i kolumn nummer två är betydelserna utskrivna. I den tredje kolumnen skiljer sig de två uppställningarna genom att ordet *tank* har olika numrering på de två raderna, medan ordet *glas* saknar numrering här. Denna numrering avser lemgramstillhörighet, således är *tank* kategoriserat som två olika lemgram medan *glas* kategoriseras som ett lemgram. Detta kan man sluta sig till genom att klicka på varje ord i den tredje kolumnen. För ordet *tank* kommer

då olika böjningsformer upp beroende på vilken rad man klickar på, och för ordet *glas* kommer samma böjningsformer upp för de tre raderna i kolumn tre.

Om man jämför indelningen i lemgram och betydelser för orden *tank* och *glas* i SALDO, med indelningen i lemma och lexem för samma ord i NEO ser man följande:

ORD	NEO		SALDO	
	Antal lemman	Antal lexem	Antal lemgram	Antal betydelser
<i>glas</i>	1	3	1	3
<i>tank</i>	1	2	2	2
Summa	2	5	3	5

Vi ser att indelningen skiljer sig vilket beror på att de formella böjningsmöjligheterna för *tank* (behållare) och för *tank* (stridsvagn) tolkats olika. I SALDO har den möjliga pluralformen *tanks* för *tank* (stridsvagn) lett till att ordet blivit ett eget lemgram. I NEO anges pluralformen på -s som ett alternativ utan att ge upphov till ett eget lemma.

Ett lemgram är alltså, liksom lemmat, formellt avgränsat – det vill säga att det är stavning och ordböjning som i stora drag ligger till grund för indelningen. Och de så kallade betydelserna är liksom lexemen, semantiskt avgränsade.

Jag nämnde tidigare att pappersordboken med sina lemman och lexem är tydlig och överskådlig. En sådan tydlighet och överskådlighet är möjlig eftersom ordboken hanterar ett begränsat antal lemman (även om det inte är uttömmande och täcker hela språket är det vid varje givet tillfälle begränsat till det antal lemman som utgör den så kallade lemmalistan, ja helt enkelt uppslagsorden i ordboken). Korp däremot är ett korpusverktyg och materialet man söker i är inte en ordbok utan mycket stora textmängder. Den avgörande skillnaden är att systemet har att ta hänsyn till varje enskilt belägg. Det innebär att lemgram och betydelser inte bara ska kopplas till varandra (som lemma och lexem är kopplade till varandra i en ordbok) utan också till varje enskilt belägg av en viss ordform. Beroende på ordform kan det röra sig om hundratusentals belägg. Vart och ett av beläggen tillhör ett visst lemgram och har en viss betydelse beroende på kontext. Här uppstår problemen, som jag förstår det, ur användarens synvinkel.

Taggningen av varje enskilt belägg sker nämligen enbart utifrån formen (specifika bokstäver i en viss ordning), det vill säga lemgramstillhörighet, och om ordets form råkar återfinnas i flera lemgram blir ordet taggat som ingående i samtliga dessa lemgram. För ordformen *tankar* t.ex. ges informationen att det ingår i de tre lemgrammen *tanke*, *tank1* och *tank2*. Detta stämmer ju, som form betraktat ingår *tankar* i samtliga tre. Men om man tar hänsyn till betydelsen är det tvärtom så att varje enskilt belägg bara kan ha en. (Jag bortser här från fall där ett visst ord skulle kunna betyda till exempel två olika saker i samma kontext eftersom dessa får betraktas som markerade fall.)

Dessutom, även om inte själva ordformen i belägget ingår i ett visst lemgram (formen *tanke* ingår till exempel inte i lemgrammen *tank1* och *tank2*) blir det ändå taggat tillsammans med dessa lemgram, eftersom det *delar någon form i sitt böjningsmönster med någon form i böjningsmönstren hos de andra lemgrammen*. Försök har gjorts att automatiskt betydelseparera belägg, det vill säga ange vilket av lemgrammen som varje belägg tillhör, men resultatet är så pass otillförlitligt att man bestämt sig för att låta bli.

Man bör alltså ha i åtanke att de betydelser som anges för varje belägg i Korp genereras av lemgramstagningen och inte av den specifika kontexten. Det betyder att om ett belägg är taggat i tre lemgram kommer också samtliga betydelser för dessa lemgram upp. Sökverktyget

kan alltså inte tala om vilket lemgram och vilken betydelse som gäller för ett visst belägg ifall det finns flera att välja på.

Ett problem för användaren är att det inte går att utesluta ett visst lemgram från sökningen om ett ord skall råka tillhöra flera. I min fortsatta genomgång kommer jag in på vad sökningar på ord, lemgram och utökade sökningar ger för möjligheter och har för begränsningar.

Att använda Korp

Med hjälp av några inledande enkla sökningar på orden *tanke* och *tänka* ges nedan först en beskrivning av Korp som sökverktyg. Jag försöker presentera möjligheter och begränsningar mer övergripande först, för att sedan gå vidare mer specifikt och undersöka möjligheter och begränsningar i relation till mina forskningsfrågor.

Sökverktyget har tre söknivåer, en enkel, en utökad och en avancerad. Nedan presenteras först den enkla söknivån och därefter i någon mån den utökade. För att använda den avancerade nivån krävs som namnet antyder mer avancerade kunskaper i språkteknologi vilket jag, och förmodligen många användare med mig, inte har. Först och främst ska sägas att för att få tillgång till Korps alla funktioner bör man använda webbläsaren Firefox (inte Explorer).

På Korps enkla söknivå kan man söka på *ord* eller på *lemgram* i en eller flera korpusar. Söker jag på *tanke* som ord letas alla belägg av ordet *tanke* upp i ett valt material, alltså i obestämd form singular. Söker jag istället på *tanke* som lemgram letas belägg upp med samtliga böjningsformer (se användarhandledningen till Korp för information hur man söker på ord respektive lemgram). Det är lemgramssökningen jag är intresserad av för mina syften. För varje sökning anger man vilka korpusar som ska ingå i sökningen.

Resultatet av en sökning visas på tre stycken flikar, den översta är fliken KWIC (Keywords in Context), bakom den ligger flikarna Statistik och Ordbild. KWIC samlar konkordanser för beläggen direkt ur korpusarna, fliken Statistik visar siffror för beläggen och fliken Ordbild presenterar beläggen utifrån deras syntaktiska omgivning.

Jag gör en sökning på lemgram *tanke* i korpusarna Akademiska/Humaniora och Bloggmix. Jag valde dessa två för att de representerar olika texttyper. Humaniora innehåller 12 716 695 token och Bloggmix innehåller 266 899 558 token. En bra funktion med Bloggmix är att man kan klicka sig vidare från en konkordans till den blogg där den hämtats och alltså se den större kontexten. Sökningen på lemgram *tanke* i dessa båda korpusar ger 86 845 träffar. Denna siffra och de 25 första beläggen i konkordanser visas på fliken KWIC. Jag börjar dock här med att presentera vad fliken Statistik ger för information i det här sökläget, eftersom jag tänker mig att tekniken möjliggör tillgång till frekvensräkning som inte går att göra manuellt.

Statistik

Totalt antal träffar för lemgram *tanke* har jag alltså redan fått uppgift om på fliken KWIC, och när jag klickar fram statistikfliken får jag upp flera rader med dels totalt antal träffar för lemgrammets samtliga former (överst i figuren nedan) och per korpus, dels träffstatistik för lemgrammets olika former per korpus. (Skärmdumparna i framställningen är beskurna.)

Träff	Totalt	Bloggmix (januari 2012)	Humaniora
Σ	310,6 (86 845)	307,8 (82 140)	370,0 (4 705)
tanke	114,4 (31 983)	114,7 (30 613)	107,7 (1 370)
tankar	87,8 (24 554)	86,3 (23 044)	118,7 (1 510)
tanken	53,0 (14 829)	51,7 (13 799)	81,0 (1 030)
tankarna	20,6 (5 764)	20,6 (5 492)	21,4 (272)
Tanken	17,7 (4 956)	17,5 (4 684)	21,4 (272)
Tankar	7,9 (2 207)	7,9 (2 117)	7,1 (90)
TANKAR	4,1 (1 134)	4,2 (1 134)	
Tankarna	3,1 (855)	3,1 (821)	2,7 (34)
Tanke	0,6 (160)	0,5 (142)	1,4 (18)
tankens	0,5 (126)	0,3 (91)	2,8 (35)
tanke-	0,3 (93)	0,2 (45)	3,8 (48)
tankarna	0,2 (55)	0,1 (37)	1,4 (18)
Tankens	0,1 (35)	0,1 (32)	0,2 (3)
TANKE	0,1 (28)	0,1 (27)	0,1 (1)
tankars	0,1 (23)	0,1 (21)	0,2 (2)
TANKEN	0,1 (18)	0,1 (18)	
tankes	0,0 (10)	0,0 (10)	

Bild 1

Här visar det sig att sökningen är skiftlägesberoende, vilket får konsekvensen att träfflistan är uppdelad i t.ex. *tanke* och *Tanke* osv. Eftersom jag inte är specifikt intresserad av denna information för statistiken markerar jag alternativet ”skiftlägesoberoende” (fortfarande på fliken Enkel sökning) och gör en ny sökning men får ändå samma resultat. Man bör vara medveten om att Korp är under utveckling och just den här funktionen är inte tillgänglig. Via statistikfliken ovan får jag ändå information om antal belägg för lemgrammet *tanke* i alla böjningsformer (numerus, species, genitiv) i valda korpusar, även om jag här måste räkna ihop alla olika skiftlägen manuellt.

Träffarna anges i relativa tal (= träffar per miljon ord) med absoluta tal inom parentes. Det ger möjlighet att jämföra träfffrekvens korpusarna emellan. I fallet med de två valda korpusarna är således den relativa träfffrekvensen högre i Humaniora (370,0) än i Bloggmix (307,8).

Informationen om träffar per korpus i relativa och absoluta frekvenser ges även i diagramform, om man klickar på det lilla cirkeldiagrammet i den kolumn som heter Träff på statistikfliken. Eftersom detta cirkeldiagram åskådliggör hur träffarna är fördelade korpusarna emellan är den här funktionen bara tillgänglig när man sökt i fler än en korpus. Diagrammen ger inga nya siffror men är överskådliga, se bild 2 nedan som visar frekvenser per korpus för böjningsformen *tanke*.

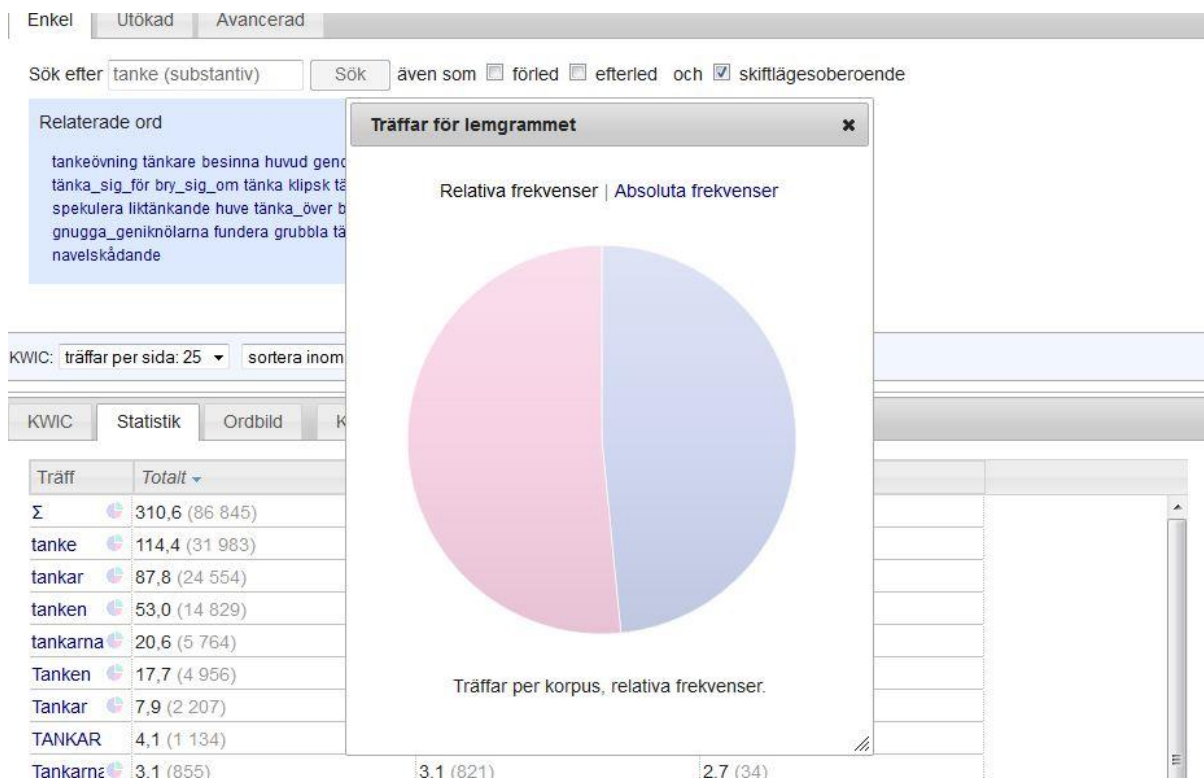


Bild 2

I diagrammet kan man klicka fram antingen relativa eller absoluta frekvenser. Dessutom kan man klicka i färgfälten i själva diagrammet för att få fram en liten extra ruta där antal träffar för varje korpus anges.

Träfffördelning för hela lemgrammet mellan de sökta korpusarna kan man också nå snabbt via den liggande stapeln ovanför konkordanserna på fliken KWIC. När man håller muspekaren ovanför stapelns olika delar visas antal träffar per korpus och man kan också klicka sig vidare här till konkordanserna. Eftersom stapeln visar träfffrekvens per korpus visas den bara när man sökt i fler än en korpus.

Standardinställningen på statistikfliken är att informationen är sorterad utifrån *ord*. Man kan också välja att sortera utifrån olika ordattribut eller olika textattribut, men den information man får upp är svårtillgänglig. Som sagt, Korp är under utveckling.

Sammanfattningsvis har jag via statistikfliken fått veta absoluta och relativa träfffrekvenser för lemgrammet *tanke* totalt och fördelat på de genomsökta korpusarna. Jag kan också få reda på absoluta och relativa träfffrekvenser för lemgrammets olika böjningsformer, men måste då räkna lite manuellt på grund av skiftlägesproblematiken. En betydande felkälla när det gäller frekvenserna är det faktum att lemgrammet *tanke* råkar ha flera böjningsformer gemensamma med andra lemgram. Formen *tankar* t.ex. är pluralform till de tre substantiven *tanke*, *tank* (behållare) och *tank* (stridsvagn) och är därtill en finit böjningsform av verbet *tanka*.

KWIC

Så vidare till fliken KWIC där konkordanserna med de olika beläggen presenteras. Dessa kan visas och sorteras på flera sätt, och som nämnts är standardvisningen av konkordanser den som kommer upp på fliken KWIC vid en enkel sökning. Här kan man gå in och hämta

specifik information om varje belägg. Varje belägg visas i en kontext som är en (oftast grafisk) mening. När man klickar på själva belägget i en konkordansrad markeras det (som ordet *tankar* på sjätte raden i bilden nedan), och en sidopanel visas med information sorterad under rubrikerna Korpus, Textattribut och Ordattribut (längst till höger i bilden nedan).

Jag har här begränsat sökningen till enbart korpusen Akademiska/Humaniora och för sökningen på [lemgram *tanke* [substantiv]] ser den första konkordanssidan ut som nedan.

KWIC Statistik Ordbild

Antal träffar: 4 705

Föregående 1 2 3 4 5 6 7 8 9 10 11 ... 188 189 Nästa

HUMANIORA

Ekelunds och Lundgrens bok väcker **tankar** om medias makt, forskningsetik och journalistis
 epticism angående möjligheten att veta vilka **tankar** vi har.
 tt det är svårt att observera innehållet i våra **tankar** direkt.
 Om man går med på att det är möjligt att en **tanke** logiskt följer av en annan, att två tankar är oför
 tt en tanke logiskt följer av en annan, att två **tankar** är oförenliga, eller att tankar handlar om fakta i
 i annan, att två tankar är oförenliga, eller att **tankar** handlar om fakta i världen, så kan vi lika gärna
 n primärt tänker sig att teorier är innehållet i **tankar**, vilket skall förstås som att teorier därmed inte
 der förändringar är ju en klassisk metafysisk **tanke**.
 t vi ju normalt beskriver som uttryck för våra **tankar**.
 och fritt att lägga till att utsagor är uttryck för **tankar**; det spelar mindre roll, så länge som samma ty
 samma typer av relationer föreligger mellan **tankar** som mellan satser.
 Den uppgiften blir med **tanke** på 38 förberedelsearbetet ett mått på församlin
 Tro & **Tanke** 1993: 9 (Religion och Samhälle nr 88).
 Tro & **Tanke** 1995: 6.
 idan lösning kunde också vara attraktiv med **tanke** på att svärordets bokstavliga betydelse ibland k
 mentera detta ordbruk kort, framför allt med **tanke** på att jag inte utgår från den s.k. konstruktions
 Med **tanke** på att mina belägg är hämtade från talspråksim
Tanken med riktvärdena var att de skulle skapa förutsä

Korpus
 Humaniora

Textattribut
 typ: Article
 ämne: Philosophy

Ordattribut
 ordklass: substantiv
 msd: NN,UTR,PLU,IND,NOM ⓘ
 grundform:
 tanke
 tank
 lemgram:
 tank (substantiv)
 tank² (substantiv)
 tanke (substantiv)
 betydelse:
 tanke
 tank
 tank²
 dependensrelation: Subjekt (övrigt subjekt)
 förled: [tom]
 efterled: [tom]

Bild 3

Under rubriken Korpus anges vilken korpus belägget är hämtat från. (I två korpusar, nämligen Bloggmix och Wikipedia, kan man här via länkar komma vidare till de hemsidor träffarna i konkordanserna är hämtade. När det gäller Wikipedia kan dock innehållet ha ändrats.) Konkordanserna kan sorteras på *förekomst*, *träff*, *vänsterkontext* eller *högerkontext*. Sortering på förekomst är standardvisningen och den visar beläggen korpus för korpus med beläggen osorterade. Sortering på träff innebär att belägg med samma böjningsformer visas tillsammans. Sortering på vänsterkontext visar konkordanserna i alfabetisk ordning utifrån första bokstaven i ordet direkt till vänster om beläggen, och slutligen visar sortering på högerkontext konkordanserna i alfabetisk ordning utifrån första bokstaven i ordet direkt till höger om beläggen.

Under rubriken Textattribut anges vilken typ av text det rör sig om (så som t.ex. titel, datum, ämnesområde och den här informationen ser olika ut för olika korpusar). För sökningen ovan får man alltså veta att det markerade belägget är hämtat från en vetenskaplig artikel i filosofi.

Vidare under rubriken Ordattribut finns taggningsinformation för belägget. Samtliga attribut här utom ett beskriver ordets form kontextlöst. Det vill säga att attributen *ordklass*, *msd*, *grundform*, *lemgram*, *betydelse*, *förled* respektive *efterled* gäller för just den här ordformen oavsett kontext. Attributet *dependensrelation* däremot, gäller för just det här belägget i dess kontext. Som synes i sidopanelen i bilden ovan är den angivna dependensrelationen här *Subjekt (övrigt subjekt)*. Relationen som beskrivs är den mellan det i konkordansen markerade ordet *tankar* och det rosamarkerade ordet *handlar*. När det gäller dependensrelationerna finns anledning att undersöka tillförlitligheten, vilket jag återkommer till.

Ordbild

Den tredje fliken i en enkel sökning heter Ordbild. Här visas det sökta lemgrammet i en statistisk uppställning över de syntaktiska relationer där det ingår. I bild 4 nedan visas ordbilden för en enkel sökning på [lemgram *tanke* (substantiv)] i korpusen Akademiska texter/Humaniora.

Preposition		Attribut		tanke		Efterställt attribut		tanke		Verb		Verb		tanke																																																																																																																																					
1. för	128	1. central	26	1. som pröva	2	1. vara	225	1. pröva	47	2. med	108	2. ny	29	2. som presentera	4	2. leda	39	3. utifrån	21	3. ursprunglig	7	3. som finnas	7	3. utvecklas	13	3. uttrycka	30	4. åt	14	4. filosofisk	6	4. som finna	7	4. finna	29	4. avvisa	18	5. mot	19	5. skeptisk	3	5. som ge sig	2	5. utveckla	13	5. föra	27	6. kring	10	6. statsrättslig	3	6. som genomsyra	2	6. gå	19	6. utveckla	31	7. inför	7	7. esoterisk	4	7. som slå rot	1	7. förmå	6	7. väcka	17	8. emot	3	8. vanlig	8	8. som stamma	1	8. ligga	15	8. tänka	21	9. i livet	3	9. rimlig	4	9. som stamma ²	1	9. finnas	25	9. formulera	18	10. över	9	10. nationalistisk	3	10. som framföra	2	10. formulera	8	10. släppa	9	11. mitt i	3	11. hälsosam	2	11. som väcka	2	11. föra	8	11. stödja	10	12. på samma gång	1	12. sån	4	12. som formulera	2	12. uttrycka	9	12. framkasta	5	13. nära	2	13. pedagogisk	4	13. som så	1	13. genomsyra	4	13. föra fram	7	14. nära	2	14. strukturalistisk	2	14. vilken forma	1	14. kretsa	4	14. förmedla	9	15. nära	2	15. orimlig	2	15. som mogna	1	15. dyka upp	4	15. presentera	11

Bild 4

Det är dependensrelationerna för beläggen som ligger till grund för ordbilderna på den här fliken. För den här sökningen visas träffar inom fem olika dependensrelationer, markerade med olika färger.

För varje dependensrelation anges de 15 ord som har högst värde beräknat enligt *Lexicographer's Mutual Information Index* (LMI) (se t.ex. Church & Hanks 1990). Syftet med detta index var enligt angiven källa att utveckla en metod för att finna regler för ordassociation och för samförekomst (eng. *Word Association Norms*). Enligt Church & Hanks hade ordassociation hittills undersökts endast psykolingvistiskt och med hjälp av informanter som fick ange vilka ord de associerade fram till exempelvis ordet *doctor*. Det index som arbetades fram bygger istället på en formel med vilken vissa associationsregler kan spåras utifrån ordfrekvenser i stora textmängder.

En fullskalig redogörelse för LMI-beräkning ryms inte inom den här studien, men indexet bygger på så kallad ”mutual information”. Syftet är att skilja på samförekomster av ord som uppträder slumpvis i en korpus och på samförekomster av samma ord som uppträder beroende på att orden ingår exempelvis i en kollokation, vilket då betraktas som en ordassociationsregel. Mutual information jämför sannolikheten (P) för att två ord (x och y) ska förekomma tillsammans i en viss ordning [$P(x, y)$] (*joint probability*) med sannolikheten för att de slumpvis ska förekomma bredvid varandra [$P(x)P(y)$] (*probability by chance*). Vid beräkning anges ett antal ord inom vilket samförekomsterna ska räknas, ett så kallat *window of words*. I artikeln anger författarna sitt *window of words* till 5 ord. Om den gemensamma sannolikheten (*joint probability*) är större än den slumpvisa (*probability by chance*) (således $P(x, y) > P(x)P(y)$) finns anledning att anta att det beror på att någon typ av regel associerar de två orden till varandra (för en utförligare redogörelse för LMI, se Church & Hanks 1990:23). Man försöker alltså sortera bort samförekomster som inte kan förklaras med någon typ av regel.

Vid en första anblick på ordbildsfliken är det lite oklart vad det är för information man egentligen får, men jag tror att man bör ha samma inställning till uppgifterna här som Church & Hanks (1990:29) menar att man bör ha till LMI, nämligen att "[...] the association ratio could be an important tool to aid the lexicographer, rather like an index to the concordances. It can help us decide what to look for; it provides a quick summary of what company our words do keep".

Alltså, ordbildsfliken skulle kunna ses som en innehållsförteckning till konkordanserna utifrån vilken man sedan kan gå vidare till konkordansanalys. Eller, fritt översatt, en sammanfattning över hur orden sällskapar med varandra. Ord nr 1 i varje färgfält i bild 4 ovan har således högst LMI-värde inom den aktuella relationen, ord nr 2 näst högst LMI-värde osv. (Det är i gränssnittet inte möjligt att få fram fler än dessa 15 ord.) För varje ord anges också antal träffar i absoluta tal efter ordet. LMI-värde och absolut träfffrekvens korrelerar alltså inte, vilket man bör ha i åtanke. Så långt vad gäller numreringen 1–15 på ordbildsfliken.

För att tyda uppgifterna bör man också känna till att när flera dependensrelationer anges till det sökta lemgrammet i samma ordbild (som i fallet Preposition> Attribut> *tanke*> Efterställt attribut – markerade gult, grönt och aprikos ovan), har dessa ingen inbördes relation sinsemellan. Varje angivet ord ingår alltså i en dependensrelation enbart till lemgrammet [*tanke* (substantiv)]. Det är alltså inte ordföljd som ligger till grund för ordbildsinformationen utan taggningen av dependensrelationer. Så som det presenteras här kan man dock förledas att tro att det är ordföljden som avgör ordbilden. Det faktum att rubrikerna i ordbilderna ibland utgör ordklasser (*preposition, verb*) och ibland syntaktiska relationer (*attribut, efterställt attribut*) gör också informationen svårtydd. Huruvida ordföljden i någon mån ligger till grund för automattaggningen av dependensrelationer är ytterligare en fråga. Nedan visas de tre första dependensrelationerna i ordbilden ovan, med konkordans exempel för varje.

PREPOSITION **tanke** [...] vilket vi ju normalt beskriver som uttryck **för** våra tankar.

(I sidopanelen till detta belägg anges att *tankar* är taggat som prepositionskomplement (rektion) till *för*, medan *för* är taggat som efterställd bestämning till *uttryck*.)

ATTRIBUT **tanke** [...] hänger ihop med en **central** tanke hos Taylor.

(I sidopanelen till detta belägg anges att *tanke* är taggat som prepositionskomplement till *med*, och *central* är taggat som framförställt attribut till *tanke*.)

EFTERSTÄLLT ATTR. **tanke** En svindlande tanke, **som prövade** fantasin till det yttersta – men var kanske det han såg just nu varken mer eller mindre än en vandring från stjärna till stjärna på en bro av eld?

(I sidopanelen till detta belägg anges att *som* är taggat som subjekt till *prövade*, att *prövade* är taggat som efterställd bestämning till *tanke* och att *tanke* är taggat som makrosyntagm i förhållande till *men*.)

Den korta exemplifieringen ovan visar en generell svårighet när det gäller taggningen av dependensrelationer, nämligen att taggningen inte är ömsesidig i varje relation. I det första exemplet ovan hade en ömsesidig taggning inneburit att inte bara *tankar* var taggat som prepositionskomplement till *för*, utan också att *för* var taggat som huvudord till *tankar*. Istället är de två orden taggade i relation till åter andra ord.

Ordbilden i den fjärde kolumnen ovan visar dependensrelationen där lemgrammet *tanke* är taggat som subjekt till ett verb. Man kan även här av uppställningen förledas att tro att det rör sig om belägg där ordet *tanke* helt enkelt är placerat före ett verb, men så är alltså inte fallet.

Detsamma gäller för den femte kolumnen; här listas *tanke* med taggningen objekt till ett verb oavsett om det är placerat före eller efter detsamma.

Till höger om varje ord i kolumnerna på ordbildsfliken visas en dokumentikon. Klickar man på den kommer man vidare till konkordanserna för belägget. I konkordanserna visar sig eventuella feltaggningar och beroende på forskningsfråga måste man här gå in och bilda sig en uppfattning om vad man kan använda i den information som ges. (För mina syften, se vidare nedan.)

Om vi dröjer lite vid själva ordbildsfliken, och bortser från dess begränsningar, kan man använda den för att jämföra sökningar i olika korpusar. I bild 5 nedan visas fliken Ordbild för en sökning på [lemgram *tanke* (substantiv)] i korpusen Bloggmix.

KWIC		Statistik		Ordbild					
Preposition	Attribut	tanke	Efterställt attribut	tanke	Verb	Verb	tanke		
1. av	1686	1. negativ	321	1. som räkna	219	1. vara	5047	1. tänka	1726
2. vid	502	2. positiv	286	2. som snurra	126	2. gå	1035	2. ha	2262
3. med	1606	3. blott	171	3. som slå sig	92	3. snurra	236	3. skänka	300
4. utan	194	4. hemsk	142	4. som slå	104	4. slå sig	163	4. ägna	259
5. bakom	110	5. klok	118	5. som dyka upp	46	5. slå	226	5. rensa	230
6. om	514	6. djup	108	6. som dyka	46	6. fara	149	6. släppa	294
7. vad för	86	7. glad	113	7. som fara	26	7. vandra	81	7. väcka	200
8. utifrån	24	8. kritisk	59	8. som flyga	26	8. tänka	165	8. sända	169
9. kring	36	9. intressant	112	9. som tanka	11	9. svindla	53	9. skingra	106
10. ur	60	10. dum	69	10. som röra sig	21	10. svindla ²	53	10. samla ²	166
11. utav	27	11. automatisk	44	11. som röra	23	11. flyga	103	11. samla	166
12. i dag	23	12. hög	155	12. som slå om	9	12. finnas	302	12. läsa	350
13. komma på fråga	4	13. konstig	82	13. som cirkulera	10	13. finna	302	13. slå	205
14. för sig	30	14. god	378	14. som komma	72	14. dyka upp	66	14. slå sig	117
15. inom kort	6	15. jobbig	66	15. som hinna	10	15. dyka	67	15. skriva ner	97

Bild 5

Med tillgång till ordbilden för samma sökning i korpusen Akademiska texter/Humaniora (bild 4) kan vi göra vissa jämförelser. Exempelvis går att utläsa att de tre prepositionerna *av*, *vid* och *med* ligger i topp i relation till ordet *tanke* (enligt LMI) i båda korpusarna, dock med olika placering. Vi kan se att de tre översta verben till vilka *tanke* är subjekt är *vara*, *gå* och *snurra* i Bloggmix och *vara*, *väcka* och *utvecklas* i Humaniora. Vidare ser vi att när *tanke* är objekt ligger *tänka* högst i Bloggmix och *pröva* högst i Humaniora. (Inom parentes sagt: som jag förstår saken är det faktum att verbet *tänka* ligger högst när *tanke* är objekt, medan verbet *ha* (nr 2 ovan) är det mest frekventa när *tanke* är objekt, ett tydligt exempel på hur LMI-beräkningen faller ut. Nämligen på så vis att 1 (*tänka en tanke* – i olika former) förekommer visserligen färre gånger totalt men verbet *tänka* har också färre andra objekt i korpusen totalt. Om man jämför med verbet *ha* förekommer det fler gånger med objektet *tanke* (*ha en tanke* – i olika former) men detta verb förekommer även med en mängd andra objekt – till skillnad från verbet *tänka*. Därmed framstår relationen mellan *tänka* och *tanke* som mer stabil, det vill säga sannolikheten för att de två orden ska förekomma tillsammans är större än för att de två orden *ha* och *tanke* ska förekomma tillsammans (eftersom konkurrerande objekt till *ha* är betydligt fler.) En annan snabb iakttagelse är att uttrycket ”det är tanken som räknas” tycks vara väl använd i Bloggmix; *som räkna* ligger högst upp som efterställt attribut till *tanke*. I Humaniora är det, även här, *pröva* som ligger högst.

För att vara säker på vad ordbilderna egentligen säger måste man dock analysera konkordanserna, och redan vid en första anblick är vissa resultat förbryllande. Exempelvis är

nummer 12 i listan över prepositioner i relation till *tanke* i Bloggmix ovan *i dag* (samma ord ligger som nr 14 i Akademiska, se bild 4). Eftersom *i dag* inte är en preposition kan man undra vad som hänt, eller har taggningen skett utifrån *i*? Klickar man på dokumentikonen till höger om *i dag* får man fram konkordanserna och då visar det sig att de innehåller fall där *tanke* visserligen är prepositionskomplement till *i*, men där ordet *dag* återfinns någonstans i samma mening, som t.ex. i exemplet nedan (mina kursiveringar).

Men jag tänker inte drunkna *i* snurriga *tankar* en alldeles lagom *dag* som denna.

Det är oklart varför ordet *dag* kommit med *i* taggningen här. Ett annat exempel är verbet *svindla* i den ordbild där *tanke* är subjekt ovan i bild 5. Här anges *svindla* som nr 9 och *svindla*² som nr 10 och de har båda frekvensen 53 i absoluta tal. Detta är svårtolkat utifrån ordbildsfliken så man får gå vidare till konkordanserna. Det visar sig att träffarna för *svindla* och *svindla*² är desamma och anledningen till att verbet kommer upp som två poster är att ordformen tillhör två olika lemgram. Det vill säga att även om det i konkordanserna är betydelsen hos lemgram *svindla* som gäller, är ordet även taggat som *svindla*² eftersom ordformen ingår i bägge dessa lemgram. Det faktum att lemgramstaggningen inte är betydelseseparerad får alltså konsekvenser för träffstatistiken på flera plan.

Ett annat exempel är nr 13 och 14 i samma ordbild (bild 5). Här är det verben *finna* och *finnas* som båda återges med träfffrekvensen 302 i absoluta tal. Ser man bara till den här tabellen kan man alltså förledas att tro att *tanke* återfinns som subjekt till *finna* 302 gånger i korpusen Bloggmix och även som subjekt till *finnas* 302 gånger i samma korpus. Även i det här fallet är det lemgramstaggningen som spökar. Varje belägg i konkordanserna tillhör två olika lemgram (*finna* respektive *finnas*) till formen, men i varje konkordansrad tillhör de naturligtvis bara ett lemgram vilket dock inte framkommer. För att ta reda på mer exakt vilket verb som alltså används här med subjektet *tanke* måste man gå igenom samtliga träffar. Det finns väl anledning att tro att *tanke* är vanligt som subjekt till *finnas* (t.ex. *tanken finns*), men mer ovanligt som subjekt till *finna* (t.ex. *tanken finner sin väg*), men detta framgår alltså inte av ordbilden utan kräver en genomgång av konkordansraderna.

Sammanfattningsvis, den information som ordbilderna ger kan ses som ingångar till konkordanserna, och beroende på vad man ställer för frågor till materialet måste man undersöka ordentligt hur informationen ska tolkas. Genomgången ovan visar några generella möjligheter och begränsningar utifrån den enkla sökningen i Korp. Nedan övergår jag till att undersöka hur Korp skulle kunna användas för mina specifika forskningsfrågor och i samband med det går jag också vidare till att använda den utökade sökfunktionen.

Några forskningsfrågor

I mitt arbete vill jag alltså försöka kartlägga användningar av orden *tänka* och *tanke*. Som nämnts ovan beskrivs och tolkas dessa ord i min tidigare undersökning i relation till de samtalskontexter där de ingår (Grahn 2012). Ett sådant perspektiv är dock inte möjligt i korpusundersökningar där kontexterna endast utgör grafiska meningar (och ibland inte ens det på grund av bristande interpunktion i texterna). Forskningsfrågorna i en korpusundersökning måste avgränsas i förhållande till vad som är möjligt. I ett initialt skede vill jag med Korpusökningarna undersöka följande frågor:

1. Vilka verb används i konstruktioner där *tanke* är subjekt? (Och, vilka är objekten i dessa konstruktioner?)
2. Vilka verb används i konstruktioner där *tanke* är objekt? (Och, vilka är subjekten i dessa konstruktioner?)
3. Vilka subjekt används i konstruktioner med verbet *tänka*? (Och, vilka är objekten i dessa konstruktioner?)

Nedan går jag igenom hur man skulle kunna gå tillväga för att komma vidare med dessa frågor.

1. Vilka verb används i konstruktioner där *tanke* är subjekt?

En möjlighet för att närma sig frågan är att använda de enkla sökningar jag beskrivit ovan, i det här fallet en enkel sökning på [lemgram *tanke* (substantiv)]. Sökningen ger 247 639 träffar i hela korpusamlingen. Man får också statistik över träfffrekvenser i respektive korpusar. Som nämnts ovan är statistiken inte tillförlitlig eftersom lemgrammet *tanke* har gemensamma böjningsformer med lemgrammen *tank* (behållare) och *tank 2* (stridsvagn).

Den statistik som ges kan kanske för mitt vidkommande ändå ge en fingervisning om vilka korpusar som kan finnas anledning att fiska vidare i. Intuitivt skulle jag vilja påstå att ordet *tanke* är överlag mer använt än ordet *tank* (i bägge betydelserna). Det vill säga att vi skriver oftare om tänkta tankar än om behållare och stridsvagnar. Man kan därför anta att de träffar som används i betydelsen *tank* och *tank*² är mindre frekventa. Det skulle i så fall betyda att i korpusar där de relativa träfffrekvenserna är höga för [lemgram *tanke* (substantiv)] finns det anledning att gå vidare och undersöka konkordanserna manuellt.

Jag väljer korpusen Humaniora och får 4 705 träffar med samma enkla sökning på [lemgram *tanke* subst)]. Om man nu inte vill gå igenom samtliga träffar manuellt för att komma åt vilka verb som ordet *tanke* är subjekt till kan man få viss hjälp via fliken ordbild, som jag beskrivit ovan. Här finns dock begränsningarna som nämnts ovan i avsnittet om ordbildsfliken, och dessutom presenteras bara 15 verb. En annan möjlighet är att istället använda den utökade sökfunktionen i Korp. Man klickar på fliken utökad sökning och där kan man sedan välja ett token och lägga till en dependensrelation. Jag väljer [lemgram *tanke* (substantiv)], klickar på ett plustecken längst ner i sökrutans vänstra hörn och väljer [dependensrelation (Subjekt)] som i bild 6 nedan.

Enkel Utökad Avancerad

lemgram är
tanke (substantiv)

eller

och

dependensrelation är
Subjekt (övrigt subjekt)

eller

Sök

Bild 6

Det jag då söker efter är belägg på olika former som tillhör lemgrammet *tanke* och som är taggade med dependensrelationen subjekt. Sökningen ger 789 träffar och den första träffsidan visas nedan.

KWIC Statistik Ordbild

Antal träffar: 789

Föregående 1 2 3 4 5 6 7 8 9 10 11 .. 31 32 Nästa

HUMANIORA

man går med på att det är möjligt att en	tanke	logiskt följer av en annan, att två tankar är ofö-
tanke logiskt följer av en annan, att två	tankar	är oförenliga, eller att tankar handlar om fakta i v
tan, att två tankar är oförenliga, eller att	tankar	handlar om fakta i världen, så kan vi lika gärna ta
	Tanken	med riktvärdena var att de skulle skapa förutsätt
	Tankarna	var få, osammansatta, nakna som skelett.
	Tanken	utvecklas närmast till ett dekonstruktionsprogram
velser: Det gick till så när hon tänkte, att	tankarna	lät om sig som hörbart tal.
Men när deras hjärna vägrade göra	tankarna	klart flytande som vatten, så kunde de ändå inte
	Tankar	kring det bedrägliga och omöjliga i att mäta och b
Inte minst viktig är	tanken	om tidens åtskiljande från rummet som en markö
Hans	tankar	var emellertid opersonliga som om han endast tä
etsats, 177 inte vara säker på att hennes	tankar	är " riktiga " förrän de godkänns av en manlig auk
garna av den lilla flickan figurerar också	tanken	på det alternativa språket, den som Aronson så g
stadian, men detta samtidigt som " hans	tankar	gick åt annat håll också, som att sonen kunde va
1 det som efterlyses i tidigare texter: hur	tanken	förmår finna andra vägar än via orden.
ntifikationen med, Mira: " Hennes yrsliga	tankar	följde kvinnan som irrade därute i mörkret " (168
Dessa	tankar	plågar Jenny och driver henne att börja söka sin
På ett annat ställe utvecklas	tanken	om den nödvändiga fördröjningen: Detta var när
ännens ----- 235 Hur denna	tanke	genomsyrar hela författarskapet återkommer jag
nniskor förlorar frihet, rätt att tala, deras	tankar	räknas inte, deras protester tystas ned - De betr
	Tanken	utvecklas dock inte vidare av Ingebrand.
ändå ihop eftersom han förstod att hans	tankar	skulle betraktas som gammalmodiga och förlöjlig

Korpus

Humaniora

Textattribut

typ: PhD
ämne: Literature

Ordattribut

ordklass: substantiv
msd: NN.UTR.PLU.DEF.NOM ⓘ
grundform:
tanke
tank
lemgram:
tank (substantiv)
tank² (substantiv)
tanke (substantiv)

betydelse:
tanke
tank
tank²

dependensrelation: Subjekt (övrigt subjekt)
förlöd: [tom]
efterled: [tom]

Bild 7

Samtliga träffar på [lemgram *tanke* (substantiv)] är alltså taggade som subjekt här, men feltaggingar förekommer. För att bilda mig en uppfattning om tillförlitligheten har jag gått igenom de 100 första träffarna. Bland dessa återfinns 30 vilka jag bedömer vara felaktiga taggningar. Dessa har jag grupperat i främst fyra felkategorier vilka presenteras nedan under varsin rubrik följd av de aktuella konkordansträffarna.

Konkordanserna har klippts ut direkt från Korp och nedan anges från vilken träffsida varje belägg är hämtat (varje träffsida omfattar 25 konkordansrader). Numreringen har jag lagt till för att underlätta hänvisning. De två ord i varje konkordansrad som jag har markerat med fetstil är de ord mellan vilka dependensrelationen är angiven i Korp, utifrån beläggen *tanke*. Ord som inte är taggade i relationen, men som enligt min bedömning borde vara det, har jag kursiverat.

Beläggen är objekt till det markerade verbet (3 fall):

Den här kategorin omfattar träffar där beläggen alltså inte är subjekt till det verb som är markerat, utan istället objekt.

/träffsida 1–25/

1. Men när deras hjärna vägrade **göra tankarna** klart flytande som vatten, så kunde de ändå inte äga dem " (182).
6. Men den **tanken** har vi aldrig **haft**.

/träffsida 76–100/

14. Genom användandet av begreppet lesbisk litteratur hoppas jag kunna **väcka tankar** om att lesbisk litteratur är en del av det allmängiltiga samtidigt som det allmängiltiga i sig inte är neutralt utan består av till exempel heterosexuell litteratur.

Beläggen är objekt till ett annat än det markerade verbet/ordet (3 fall):

Den här kategorin omfattar träffar där beläggen liksom ovan inte är subjekt i satsen, utan istället objekt men här dessutom till ett annat verb än det angivna huvudordet. Det vill säga att de vare sig har subjekts- eller objektsrelation till det verb som anges men däremot objektsrelation till ett annat verb.

I exempel 8 nedan är *tanken* objekt till verbet *tangeras*, men det är hjälp verbet *har* som står angivet som huvudord i dependensrelationen. I exemplen 22 respektive 10 är det verben *ansluta* respektive *fördjupa och förverkliga* som ordet *tanke* är objekt till; det är alltså inte subjekt till *mena* respektive *kunde* som anges i taggningen. I exempel 36 är *tankar* enligt taggningen subjekt till ordet *dels*, vilket är ett intressant fel eftersom detta ord inte ens är ett verb. Rätt taggning skulle vara att *tankar* är objekt till *har tagit fasta på*.

/träffsida 26–50/

8. **Tanken** att vara ursprung till värden **har** vi *tangerat* tidigare, när det konstaterades att självtranscenderande subjekt förutsätts i Lonergans moralteori.

/träffsida 76–100/

10. För att *fördjupa* och *förverkliga* **tanken** på vad lärjungaskap **kunde** vara och för att få kontinuitet i seminariet uppkom tanken på ett broderhus.
36. Här har jag *tagit fasta på* **dels** sexologiska **tankar** om lesbiskhet, *dels bilden av " den nya kvinnan "*.

Beläggen är subjekt till ett annat huvudverb än det markerade verbet (4 fall):

Vidare, i de träffar som visas nedan är beläggen visserligen subjekt, men till ett annat verb än det markerade. Jag ser ingen riktig systematik i de här felen. De verb som taggats som huvudord verkar närmast slumpmässigt angivna.

I exempel 15 anges att *tankar* är subjekt till *benämna*, en infinit verbform som ingår i bestämningen till *tankar*. Här får man alltså inte reda på att *tankar* i själva verket är subjekt till verbet *vara*. I exempel 16 skulle den riktiga taggningen vara att *tanken* är subjekt till *figurera*, men istället anges verbet *kan* som huvudord i relationen. Vidare i exempel 13 skulle man kunna misstänka att det är ett fel i uppdelningen i grafiska meningar som försvårat taggningen. Det verb, *höja*, som *tankar* anges vara subjekt till återfinns inte ens i samma mening utan istället borde verbet *föra* vara markerat som huvudord.

Exempel 34 är lite speciellt. Här har verbet *tvingas* taggats som huvudord till subjektet *tankarna*, medan det väl är den finita formen *behövde* som är den riktiga taggningen. Samtidigt kan man säga att subjektet i relativsatsen, *som*, refererar till *tankarna* som också här *tvingas* så på något sätt blir taggningen halvt rätt här.

/träffsida 1–25/

15. **Tankar** kring det bedrägliga och omöjliga i att mäta och **benämna** är starkt närvarande.
16. I just beskrivningarna av den lilla flickan *figurerar* också **tanken** på det alternativa språket, den som Aronson så gärna återkommer till: " Claudette **kan** ett litet lustigt språk av gester – axlar, pekfingrar, nickningar, energiska huvudskakningar är orden – och jag kan ett tyst språk av hemligt instämmande " (19).

/träffsida 76–100/

13. **Tankar** om homosexualitet som medfödd, som smittsam och i kvinnors fall som en brist på män fördes fram. 232 Positiva röster **höjdes** också vil 70 4 lesbisk litteratur under 1930-talet ket kanske läkaren Anton Nyström är det främsta exemplet på.
34. **Tankarna** , som annars alltid måste tvingas till uppmärksamhet – de behövde inte **tvingas** här, de var där och stannade utan att kunna slita sig därifrån.

Beläggen är subjekt men det är hjälpverbet som är markerat (18 fall):

Jag har i 18 fall bedömt att det verb som beläggen anges stå i subjektsrelation till är hjälpverb snarare än huvudverb. En gränsdragning mellan hjälp- respektive huvudverb är inte alltid självklar men i de följande exemplen tyder det på att det rör sig om hjälpverb (se SAG Verb>Hjälpverb §29–33).

/träffsida 1–25/

5. " Medelsvensson " tyckte att pornografiskyltningen var skamlig, vidrig och motbjudande, men bet ändå ihop eftersom han förstod att hans **tankar skulle betraktas** som gammalmodiga och förlöjligas, skrev tidningen.

/träffsida 26–50/

17. **Tanken hade** ju varit, innan frågan om integrering i EEC kom upp, att samfundsbildningen och överlåtelseakten skulle ske vid en konferens i samband med ett besök från ÖM-ledningen i Sverige.
18. **Tanken** att vända på saken, att se missionärerna som medarbetare till de inhemska kristna i deras församlingsbyggande, **hade** ännu inte uppstått.
19. **Tanken** på en begränsning av missionärsantalet **hade** inte varit uppe tidigare.
20. Sjutton år senare, 1938, **kunde** denna **tanke förverkligas**.

21. Ofta **brukar** denna **tanke sammankopplas** med Sokrates och tidigt grekiskt tänkande. (??)
23. Istället kommer materialet vara hämtat från tänkare som kan anses som dominerande i, och representativa för en modern tanke- och föreställningstradition, som jag och många andra är starkt präglad av; och i vilken **tankar** om självförverkligande, föränderlighet, det universella och det individuella **har varit** synnerligen centrala teman.

/träffsida 51–75/

24. Meister Eckhardts **tanke kan** därmed **sammanfattas** som att vårt ”avbildningsvara” aldrig kan nås genom att avbilda det vi avbildar, istället måste vi av-bildas för att kunna nå hyperbildning.
25. Denna **tanke har levt** vidare som insikten att det är med hjälp av rationaliteten vi kommer närmast och bäst förstår världen, Gud och oss själva.
26. Sanningen, slutligen, är den högsta punkten som den rationella **tanken kan nå** och därmed en gränspunkt mellan rationalitet och det onåbara.
27. Från flera håll har det ifrågasatts hur adekvat **tanken** om människans kallelse **kan vara** i vår tid.
28. **Tanken** om människans kall **har** traditionellt även **gällt** t.ex. sjuksköterskor, lärare, poliser, domare och officerare.
29. **Tanken** om plikten att stå till förfogande för sin nästa, **har uppfattas** vara ställd i en motsättning till professionalisering. [felstavning i original]
30. En **tanke kan vara** att en historia som handlar om förhållanden på Mauritius inte talar till lika många indier som en där spänningen mellan moderlandet Indien och den forna kolonialmakten Storbritannien står i centrum.

/träffsida 76–100/

11. **Tanken** att samhället bör grundas på (a-historiska) vetenskapliga principer och inte genom ideologi eller historisk legitimitet **kan** i sig ses som en sådan avhistorisering.
31. När Andra vatikankonciliet avslutades 1965 **hade** många **tankar** om förnyelse **tagit form**.
32. **Tanken** om en rigid ordning i samhället där varje företeelse och person hade en fixerad roll **hade funnits** med sedan åtminstone 1600-talet i Sverige.
33. Den **tanken kan fungera** som ett avstamp in i Agdas könsöverskridanden, genom vilka hon, som Paqualén säger, ” intar en kvinnlig maskulinitetsposition ” .

Tittar man på de verb som ovan alltså taggats med en böjningsform av lemmen *tanke* som subjekt är det verben *ha* (9 st.) och *kunna* (7 st.) (och i ett fall *skola*, i ett annat *bruka*). De har typiska hjälpverbsegenskaper (se SAG 2:527) vilket bland annat innebär att subjektreferenten i dessa fall inte har någon semantisk roll i den aktion som anges med hjälpverbet. *Tankens* roll i exemplen ovan är således inte snarast att *ha* och att *kunna*, vilket taggningen gör sken av. *Tankens* semantiska roll ingår istället i huvudverbens aktioner, som till exempel *uppstå*, *förverkligas*, *nå*, *gälla*, *ta form*, *fungera* med flera.

Beläggen är relaterade till ett obegripligt/obefintligt ord (1 fall) eller är vare sig objekt eller subjekt (1 fall):

Ytterligare en typ av feltaggning är när belägget angivits som subjekt till ett ord som i kontexten inte ingår i syntaktiska relationer. Bland de 100 genomgångna träffarna finns ett exempel på detta. Här rör det sig om en taggning där siffran 52 angetts som huvudord i den relation där *tanke* ska vara subjekt. Som synes saknas verb överhuvudtaget i den här konkordansraden.

/träffsida 26–50/

35. **52 » Tankar** i konstitutionsfrågan«, Huss till ÖM-styrelsen, bilaga till Huss t Boström 3.10.1960.

Vidare finns ett fall där ordet *tanke* är taggat som subjekt men snarast utgör en rektion till prepositionen *till* (eller prepositionskomplement, som är den term som används i taggningen).

22. Och denna **tanke** om att det primärt är det universella, allmänmänskliga samt det individuellt unika, som ska realiseras eller uppnås i självförverkligandet, vill jag **mena** att även Niebuhr och Gewirth i grunden ansluter sig *till*.

Bland de 100 första träffarna finner jag, som redovisats ovan, alltså 30 feltaggningar, det vill säga 30 %. I samtliga fall är beläggen taggade som subjekt, men ibland är denna taggning felaktig (belägget är inte ett subjekt) och ibland ligger felet i det andra ordet som anges i relationen (exempelvis när ett hjälpverb anges istället för ett huvudverb). För 22 av de felaktiga fallen kan belägget betraktas som subjekt, men det verb som ordet anges stå i dependensrelation till är fel (till hjälpverb i 18 fall, till annat verb i 4 fall). För 7 av fallen gäller att beläggen är taggade som subjekt trots att det rör sig om ett objekt, i 3 fall till det angivna verbet och i 3 till ett annat, och i 1 fall en rektion. Slutligen för 1 av de 30 fallen gäller att en böjningsform av *tanke* har taggats som subjekt till en siffra.

Om 30 % feltaggningar gäller överlag för den här sökningen är det en så pass stor felkälla att det finns anledning att vara försiktig om man vill använda sig av frekvensuppgifter. Det visar istället att sökningarna får betraktas som *ingångar* till konkordansanalyser. Dels är taggningen gjord automatiskt, dels kan syntaktisk taggning knappast betraktas som en absolut vetenskap, det vill säga att det ibland finns alternativa taggningsmöjligheter av vilka ingen direkt är fel. För att tydligare förstå den bakomliggande taggningen är manuell analys av konkordanserna den enda möjligheten. I min analys av de 100 första konkordansraderna finner jag exempelvis följande 6 taggningar som inte kan betraktas som felaktiga, men som är diskutabla i förhållande till mina frågor. Det rör sig om passiver där *tanke*, i någon form, är taggat som subjekt.

/träffsida 1–25/

2. På ett annat ställe **utvecklas tanken** om den nödvändiga fördröjningen: Detta var närmandets mystiska akt.
3. Människor förlorar frihet, rätt att tala, deras **tankar räknas** inte, deras protester tystas ned - De betraktas som ägodelar.
4. **Tanken utvecklas** dock inte vidare av Ingebrand.

/träffsida 26–50/

7. **Tankarna inriktades** återigen på Ouesso, men även andra områden diskuterades.

/träffsida 51–75/

9. **Tanken** om människans kallelse **användes** i detta sammanhang sparsamt och närmast liktydigt med krävande arbete, engagerande och viktig uppgift.

/träffsida 76–100/

12. Dessa **tankar togs** något senare upp i 230 det äldre migrationshistoriska forskningsparadigmet.

De med fetstil markerade dependensrelationerna ovan är relationer där huvudordet i förhållande till ett verb i passiv form betraktas som subjekt, vilket överensstämmer med hur man traditionellt ser på den här typen av relation (se SAG 4:366ff.). Det kan således inte betraktas som feltaggningar. För mina forskningsfrågor är de dock lite svårhanterliga eftersom de inte är givna förstadedeltagare inom SFG som de flesta andra subjekt är. Snarare liknar rollen andradeltagarens (se GMB s.76 i not, s. 81). Här krävs alltså konkordansanalys.

I den här utökade sökningen har jag alltså sökt efter ordformer av substantivet *tanke* vilka är taggade som subjekt i en dependensrelation. Resultatet av sökningen är träffsidor med konkordanser på fliken KWIC. Samma dependensrelationer ligger också till grund för informationen på ordbildsfliken som man når via den enkla sökningen på lemgrammet, men där presenteras relationerna med hänsyn till LMI. Om man, som jag, är ute efter just vilka olika verb som olika former av *tanke* är subjekt till får man alltså tillgång till fler verb via den utökade sökningen än med ordbildsfliken som bara ger 15 verb.

(Ett annat sätt att ta reda på *hur många* belägg för lemgrammet *tanke* som är taggade som subjekt är att i resultatet av den enkla sökningen gå till statistikfliken och välja ”sortera på dependensrelation”. Då får man fram antal träffar för varje relation, i det här fallet 789 vilket är samma antal som jag får fram i den utökade sökningen. Härifrån kommer man dock inte vidare till konkordanserna för just den aktuella dependensrelationen.)

Genomgången ovan ger en fingervisning om olika typer av feltaggingar som förekommer och i vilken omfattning man kan förvänta sig att de dyker upp. Dock måste man ha i åtanke att man inte kommer åt alla feltaggingar i varje sökning. I min utökade sökning, t.ex., där jag ville få fram ordformer av *tanke* som är taggade som subjekt, kunde jag konstatera att somliga belägg inte alls var subjekt. Man kan då misstänka att feltaggingar förekommer även åt ”andra hållet”, det vill säga att det finns ordformer av *tanke* som är taggade i andra dependensrelationer men som borde vara taggade som subjekt. Möjligen kan man finna former som borde vara taggade som subjekt men inte är det inom samtliga andra kategorier av dependensrelationer. En utökad sökning på lemgram *tanke* med dependensrelation ”oklassificerbar satsfunktion”, visas i bild 8 nedan, bekräftar misstanken.

The screenshot shows the KWIC search interface. At the top, there are tabs for 'Statistik' and 'Ordbild'. Below the tabs, it says 'Antal träffar: 14'. The main content area displays search results for the word 'tänkar'. The results are highlighted in blue and show various contexts where 'tänkar' is used as a subject. For example, 'Härmed kan också tänken om Aronsons förmenta osjälvständighet avvisas.' and 'Men tänken om medborgarskap i detta rike menar jag ger en öppning...'. On the right side, there is a sidebar with metadata for the search results, including 'Korpus: Humaniora', 'Textattribut: typ: PhD, ämne: Literature', 'Ordattribut: ordklass: substantiv, msd: NN,UTR,SIN,DEF,NOM', 'grundform: tanke, tank', 'lemgram: tank (substantiv), tank² (substantiv), tanke (substantiv)', 'betydelse: tanke, tank, tank²', and 'dependensrelation: Oklassificerbar satsfunktion, förled: [tom], efterled: [tom]'.

Bild 8

Här ges 14 belägg av vilka 4 borde vara taggade som subjekt, men alltså istället fått märkningen oklassificerbar satsfunktion. Man kan därför tänka sig att även i de andra kategorierna finns belägg som är subjekt men som taggats som någonting annat.

Ovan har jag undersökt frågan vilka verb som används i konstruktioner där *tanke* är subjekt. Jag sluter mig till att man kan få svar på frågan men först egentligen efter noggranna analyser av konkordanserna. Frekvensuppgifter drar jag mig för att använda på grund av feltaggingar och på grund av problematiken med överlappande lemgram, men de kan ge en vink om vilka sökningar som är värda att gå vidare med i form av konkordansanalys.

2. Vilka verb används i konstruktioner där tanke är objekt?

Min nästa forskningsfråga lyder som ovan. Ett sätt att finna svar är återigen att göra en enkel sökning på [lemgram *tanke* (substantiv)] (det vill säga samma sökning som till frågan ovan) och klicka sig vidare till ordbildsfliken. Nedan återges bild från denna sökning igen och nu är det kolumnen längst till höger som ska tillhandahålla resultat, det vill säga där ordet *tanke* taggats med dependensrelationen objekt i förhållande till ett verb.

KWIC		Statistik		Ordbild					
Preposition	Attribut	tanke	Efterställt attribut	tanke	Verb	Verb	tanke		
1. för	128	1. central	26	1. som pröva	2	1. vara	225	1. pröva	47
2. med	108	2. ny	29	2. som presentera	4	2. väcka	11	2. leda	39
3. utifrån	21	3. ursprunglig	7	3. som finnas	7	3. utvecklas	13	3. uttrycka	30
4. åt	14	4. filosofisk	6	4. som finna	7	4. finna	29	4. avvisa	18
5. mot	19	5. skeptisk	3	5. som ge sig	2	5. utveckla	13	5. föra	27
6. kring	10	6. statsrättslig	3	6. som genomsyra	2	6. gå	19	6. utveckla	31
7. inför	7	7. esoterisk	4	7. som slå rot	1	7. förmå	6	7. väcka	17
8. emot	3	8. vanlig	8	8. som stamma	1	8. ligga	15	8. tänka	21
9. i livet	3	9. rimlig	4	9. som stamma ²	1	9. finnas	25	9. formulera	18
10. över	9	10. nationalistisk	3	10. som framföra	2	10. formulera	8	10. släppa	9
11. mitt i	3	11. hälsosam	2	11. som väcka	2	11. föra	8	11. stödja	10
12. på samma gång	1	12. sån	4	12. som formulera	2	12. uttrycka	9	12. framkasta	5
13. nära	2	13. pedagogisk	4	13. som så	1	13. genomsyra	4	13. föra fram	7
14. nära	2	14. strukturalistisk	2	14. vilken forma	1	14. kretsa	4	14. förmedla	9
15. nära	2	15. orimlig	2	15. som mogna	1	15. dyka upp	4	15. presentera	11

Bild 4 (upprepad)

Här får man återigen ta hänsyn till att de 15 verb som presenteras är viktade enligt LMI, vilket i praktiken ska betyda att de verb som ligger högst på listan med högre sannolikhet än slumpens återfinns i dependensrelation med objektet *tanke*. För att bilda mig en uppfattning om taggningen gör jag en utökad sökning även här, enligt nedan.

Enkel Utökad Avancerad

+

lemgram
är

tanke (substantiv)

-

eller

och

Matcha någon av följande regler:

+

dependensrelation
är

Indirekt objekt (dativobjekt)

-

dependensrelation
är

Direkt objekt (ackusativobjekt)

-

eller

+
⚙

Bild 9

De taggade objektsrelationerna är uppdelade i indirekta och direkta objekt. Det är troligt att de flesta träffarna när det gäller *tanke* som objekt kommer att vara direkta objekt men för att inte utesluta något väljer jag att ange bägge i sökkriterierna, enligt ovan. För att få fram denna sökning fyller jag först i [lemgram är *tanke* (substantiv)] och klickar därefter på plustecknet i sökrutans nedre vänstra hörn för att få fram ytterligare en sökruta. Där fyller jag i att [dependensrelation är (indirekt objekt)]. Därefter klickar jag på ”eller” i det nedre högra hörnet på denna ruta för att få fram ytterligare en sökruta där jag kan fylla i [dependensrelation är (direkt objekt)]. Således: lemgrammet ska vara *tanke substantiv* OCH beläggen ska vara taggade med dependensrelationen *indirekt objekt* ELLER *direkt objekt*. (Generellt kan man tänka om sökfunktionerna *och/plus* samt *eller* att *och* begränsar sökningen genom att göra den mer specifik medan *eller* gör sökningen bredare eftersom man specificerar fler alternativa kriterier som kan gälla.)

Till en början kan konstateras att sökningen ger 843 träffar, det vill säga något fler än vad sökningen på [lemgram *tanke* deprel subjekt] gav i samma korpus. Som nämnts hade jag misstanken att *tanke* som direkt objekt skulle vara vanligare än *tanke* som indirekt objekt. Detta kan jag kontrollera genom att gå in på statistikfliken och ange [sammanställ på dependensrelation] och sedan uppdatera sidan. Då kommer standardfliken KWIC upp igen men om man återigen klickar på statistikfliken visas den sammanställning man begärt, i mitt fall följande:

KWIC Statistik Ordbild		
Träff	Totalt	Humaniora
Σ	66,3 (843)	66,3 (843)
OO	66,1 (841)	66,1 (841)
IO	0,2 (2)	0,2 (2)

Bild 10

Här ser vi det totala antalet träffar efter sigmatecknet, och därunder träffarna uppdelade per dependensrelation. Dependensrelationen OO står för direkt objekt och det är mycket riktigt dessa som överväger stort, 841 träffar mot 2 stycken indirekta objekt. (En lista över de engelska förkortningar för dependensrelationer som används hittar man genom att klicka på kugghjulet högst upp till höger på varje Korp-sida, välja *Användarhandledning* och sedan klicka på länken *Annotering av korpusar i Korp* i menyn till vänster.)

Om man nu, som jag, blir nyfiken på kontexterna för dessa två indirekta objekt kan man tyvärr inte hitta dem direkt från denna sida. Jag försöker genom att klicka på IO men kommer då istället till en konkordanssida med träffar där just bokstäverna "IO" ingår i konkordanserna. Siffran för antal träffar är inte klickbar. Det enda sättet att snabbt få reda på vilka de två träffarna med indirekt objekt är, är att göra samma sökning som ovan men utesluta dependensrelationen direkt objekt. Jag finner att bägge träffarna är feltaggade. Det rör sig i det första fallet istället om ett direkt objekt, i det andra om ett subjekt:

Kulturradikalen, skönhetsaposteln och kvinnoaksasideologen Ellen Key (1849 – 1926), som i flera bemärkelser kan beskrivas som Kjelléns raka motpol, hade **uttalat** samma holistiska **tanke** ett år tidigare, 1897.

Tillämpad på teammedlemmarnas samtal under teamkonferenserna **tydliggör** denna **tanke** det självklara i behovet att till exempel diskutera de skilda tolkningar av en händelse eller iakttagelse som de olika teammedlemmarna gör.

Att döma av sökningen ovan finns alltså inga belägg där *tanke* är indirekt objekt, men teoretiskt skulle ändå sådana kunna dölja sig bland exempelvis träffarna på direkt objekt. För att bilda mig en uppfattning om taggningen för sökningen på lemgram *tanke* substantiv i direkt eller indirekt objektsrelation går jag igenom de 100 första träffarna.

Sammanlagt finner jag 24 fall där beläggen inte är objekt trots att de taggats som sådana. Jag har grupperat dem i fyra felkategorier vilka går igenom nedan tillsammans med de aktuella träffarna. De två ord i varje konkordansrad som jag har markerat med fetstil är de ord mellan vilka dependensrelationen är angiven i Korp. Ord som inte är taggade i relationen, men som enligt min bedömning borde vara det, har jag kursiverat.

Beläggen är subjekt till det markerade verbet (15 fall):

Den här kategorin omfattar träffar där alltså *tanke* i någon form är taggat som objekt i förhållande till ett verb men egentligen är subjekt. Det rör sig om 15 fall bland de 100 konkordanser som analyserats.

/träffsida 1-25/

1. Pronomenet " vi " förekommer frekvent och vissa gånger övergår berättandet till tredje person: " i takt med regnskurarna **dansade** en ring av **tankar** som besatta i huvudet på syster Vilmas övergivna kamrat " (79)

2. Vid upprepade tillfällen **återkommer tankar** kring detta att hitta svaret i sig själv, att allt som sker har en innebörd för den som förmår se det.

3. Mira har vid detta tillfälle bestämt sig för att ägna sina tankar åt den familj hon just lämnat, men istället liksom **invaderas** hennes **tankar** av de främmande människor hon mött, men inte haft någon kontakt med, under dagens bönemöte.

5. I " Testamente " talas om " hjärtat som hör det förstummade spela " (57) och i " Till Bengt " **finns** åter **tankarna** om det ordlösa språket: Minns du vi sam genom namnlösa vågor?

/träffsida 26-50/

7. Som framgick av det ursprungliga arbetsprogrammet **fanns tanken** att starta skolor med redan från början.
8. Inom SMF: s Kongomission **väcktes tanken** på att ge ut en handbok rörande församlingsvården redan 1928.
9. **Var tanken** att MBS skulle behålla dem eller ansågs de vara en del av församlingarnas » evangeliska arbete« som skulle överlämnas?
10. I slutet av Man ' s Nature and His Communities **återkommer** den numera bekanta **tanken** om att självförverkligande inte nås genom att försöka realisera ett egocentriskt, atomistiskt förstått själv.
12. Hos Taylor och Lonergan **återfinns** inte samma **tanke** om en grundläggande motsättning mellan individualitet och kontext, som den som Maslow och Niebuhr, men som till viss del även Gewirth, tänker sig.
13. I Holms påpekande **bär tanken** om kallelse en uppfattning om omsorgsförmågor som något som inte är självklart, utan något som måste odlas och tas tillvara genom att studeras.
14. I arbetets organisation, där kvinnor underordnades män, **förverkligades** däremot inte **tanken** om kvinnors lika ställning med män eller deras självständighet.
15. För det andra **avvisas tanken** på karriär.

/träffsida 51-75/

16. Eftersom hon inte vågar leva öppet **kretsar** hennes **tankar** nästan ständigt kring detta.
17. Anledningen var framförallt farhågan att dessa och liknande faktorer skulle påverka församlingens villighet att ställa upp på forskningsprojektet, dessutom **fanns tanken** att det skulle vara svårt att utföra forskningsprojektet, i min utformning, i en församling som t.ex. är på väg att omgruppera sig eller som har uttalade konflikter.

/träffsida 76-100/

21. Som en kontrast mot detta och den lesbiska kärlekens helhet **väcks tankar** om hennes heterosexuella relation.

Beläggen är subjekt och fel verb eller fel ord är markerat (4 fall):

Vidare finner jag 3 fall där taggningen brister dels genom att beläggen är subjekt, trots objekttaggningen, dels genom att hjälp verbet istället för huvud verbet taggats i relationen. I det fjärde fallet här är det andra ordet i dependensrelationen inte ens ett verb (se nr 24 nedan ”mord”)

/träffsida 1-25/

4. Men hos dem alla **hade** en gemensam **tanke fått fäste** i sinnet: att deras gamla liv inte kunde fortsätta [---]
6. I ÖM-styrelsen **hade** däremot **tanken** *uppkommit* att församlingarna i norra Republiken Kongo skulle anslutas till Evangeliska kyrkan i Kongo (Eglise Evangélique du Congo, EEC), det samfund som var på väg att bildas 1961 utifrån SMF: s mission i Mellankongo.

/träffsida 76-100/

22. ”Förställning var henne främmande, men på sista tiden **hade tanken** på andra människors omdöme om henne *förföljt* henne som en envis närgångenhet, och för att undkomma hade hon omedvetet försökt pressa in sig själv i en mera konventionell form. ”
24. Att träningen gett resultat visas i att Puck omöjligt kan föreställa sig att Viveka skulle kunna hemlighålla **mord**: ” **Tanken** att hon skulle kunna ha något med de båda morden att skaffa *var* direkt

vanvettig; jag kunde i varje fall inte föreställa mig att en normal och hygglig genomsnittsmänniska som hon så effektivt skulle kunna dölja en dylik hemlighet.”

Beläggen är egentligt subjekt (3 fall):

I den här kategorin rör det sig om fall där beläggen utgör egentliga subjekt, men alltså taggats som objekt.

/träffsida 51-75/

18. **Kommer** det fram olika **tankar** hos informanter rör de istället, framför allt, på vilka sätt Arken kan utvecklas till att bli något mer, något större och viktigare

/träffsida 76-100/

19. Förhållandet sattes på prov och eftersom det aldrig **funnits** någon allvarlig **tanke** på ett stabilt förhållande bakom förälskelsen bröt hon med pojkvännen.

20. Det **fanns** enligt Ingrid Åberg en **tanke** att på så sätt mildra den sociala turbulensen i samhället.

Beläggen är rektioner (2 fall):

Till sist har jag bland de 100 första träffarna funnit 2 fall där beläggen är rektioner (prepositions-komplement), inte objekt.

/träffsida 26-50/

11. Denna **tanke** hos Lonergan **ska** jag återkomma *till* längre fram, när den självförverkligande individens förhållande till omgivningen ska undersökas.

/Träffsida 76-100/

23. Synsättet **inbjuder till tankar** om glidande skalor vad gäller både kön och sexualitet.

Sammanfattningsvis, bland de 100 första träffarna på sökningen [lemgram *tanke* (substantiv)] OCH [dependensrelation (indirekt objekt)] ELLER [dependensrelation (direkt objekt)] kan 24 av taggningarna betraktas som felaktiga, det vill säga en felprocent på 24 %. Den största felkällan är när det ord som taggats som objekt egentligen är subjekt till det verb som angetts (15 fall). I 4 fall är ordet subjekt och dessutom är fel verb angivet. Dessa sammanlagt 19 fall blir då också ett slags felkälla till den sökning jag gjort tidigare där jag ville få fram ordformer av *tanke* som subjekt, eftersom dessa borde ha ingått bland träffarna där. (Omvänt gäller också för de fall bland de taggade subjekten som visat sig vara objekt.) Utöver dessa finns 3 träffar där beläggen är egentliga subjekt, och 2 träffar där de är rektioner.

Att söka efter träffar där ordet *tanke* är objekt tycks erbjuda ungefär samma möjligheter och svårigheter som att söka efter träffar där *tanke* är subjekt.

3. Vilka subjekt används i konstruktioner med verbet tänka?

Min tredje forskningsfråga rör verbet *tänka*. Jag vill främst undersöka vilka subjekten är i konstruktioner med verbet *tänka*, men i förlängningen även objekten. Även i denna förstudie använder jag mig av korpusen Akademiska texter/Humaniora och jag börjar med en enkel

sökning. Min avsikt är att söka på [lemgram *tänka* (verb)] men det visar sig att jag får många alternativa lemgram att välja mellan som innehåller *tänka*, nämligen följande:

- tänka (verb)
- tänka om (verb)
- tänka ut (verb)
- tänka sig (verb)
- tänka över (verb)
- tänka till (verb)
- tänka efter (verb)
- tänka sig för (verb)
- tänka på refrängen (verb)

Att döma av listan ovan är partikelverb, reflexiva verb och åtminstone ett idiom åtskilda som egna lemgram. Jag väljer att bara söka på [lemgram *tänka* (verb)] och får 4 373 träffar. För att få svar på min fråga att leta fram subjekt till verbet *tänka* går jag vidare till ordbildsfliken, se bild 10 nedan.

Humaniora vald — 12 716 695 token

Enkel Utökad Avancerad

Sök efter tänka (verb) Sök även som förled efterled och skiftlägesberoende

KWIC: träffar per sida: 25 sortera inom korpus på: förekomst Statistik: sammanställ på: lemgram

KWIC Statistik Ordbild KWIC KWIC KWIC KWIC KWIC

KWIC KWIC

Subjekt	tänka	Objekt	Adverbial
1. människa	28	1. tanke	21
2. tycka	3	2. tank	16
3. lagstiftare	3	3. tank ²	16
4. Y	2	4. skapare	5
5. y	2	5. forskningsarbete	3
6. missionär	4	6. arbetarklass	3
7. elev	7	7. självförverkligande	3
8. läsare	5	8. ja då	2
9. karaktär	4	9. antropolog	2
10. jag	4	10. ja ja	2
11. sammanhang	3	11. människa	8
12. sommar	2	12. överföring	2
13. same	3	13. existens	3
14. panter	2	14. process	4
15. bukt	1	15. sak	6
		1. då ²	30
		2. då	30
		3. nu	28
		4. i ungdomstid	5
		5. i termer	9
		6. i term	9
		7. kring död	4
		8. först	9
		9. i bana	3
		10. ja då	4
		11. borta	3
		12. sen	5
		13. fri	2
		14. idag	7
		15. vanligen	2

Bild 10

De ord som återfinns i den vänstra kolumnen under rubriken subjekt ska vara de ord som på grundval av taggningen och med hänsyn till LMI är vanligast förekommande som subjekt till verbet *tänka*. Här blir man misstänksam: nr 2 är ett verb, nr 3 och 4 är en bokstav (i versal respektive gemen form), nr 10 är pronomenet *jag* som enligt listan ska vara subjekt till *tänka* endast 4 gånger i en korpus på över 12 miljoner ord. Det verkar orimligt. Vidare, bland de 15 vanligaste subjekten till verbet *tänka* finns också några i sammanhanget oväntade ord, åtminstone för mig. Till exempel *sommar* (nr 12), *panter* (nr 14) och *bukt* (nr 15). Däremot finns inte vare sig pronomenet *hon* eller *han*, vilket min språkkänsla säger mig att man borde kunna förvänta. Nedanstående utökade sökning ger en lite annan bild:

Bild 12

Här har jag sökt efter subjekt som föregår verbet *tänka* och där upp till fyra andra ord kan förekomma emellan. Sökningen ger 2 941 träffar och statistikfliken visar följande:

Träff	Totalt	Humaniora
Σ	236,5 (3 007)	236,5 (3 007)
man tänka	5,7 (72)	5,7 (72)
Jag tänker	4,9 (62)	4,9 (62)
jag tänker	3,9 (49)	3,9 (49)
man tänker	3,5 (45)	3,5 (45)
vi tänker	3,4 (43)	3,4 (43)
att tänka	3,1 (40)	3,1 (40)
jag tänkte	2,6 (33)	2,6 (33)
som kan tänkas	2,0 (26)	2,0 (26)
Man kan tänka	2,0 (25)	2,0 (25)
han tänker	1,9 (24)	1,9 (24)
man kan tänka	1,7 (22)	1,7 (22)
man tänkte	1,7 (21)	1,7 (21)
Jag tänkte	1,5 (19)	1,5 (19)
Vi tänker	1,4 (18)	1,4 (18)
sätt att tänka	1,3 (17)	1,3 (17)
Man tänker	1,3 (16)	1,3 (16)
Att tänka	1,2 (15)	1,2 (15)
hon tänker	1,2 (15)	1,2 (15)

Bild 13

Statistiken visar att det är andra subjekt som är mest frekvent använda än de som visas i ordbilden. Nämligen pronomen, vilket verkar överensstämma med åtminstone min

språkkänsla. Bilden ovan är beskuren och hela listan med redogörelse för samtliga subjekt bland nästan 3 000 träffar är diger att gå igenom. En kraftig begränsning här är också att jag tvingats ange ordföljd i sökningen och inte bara relation. Återstår alltså här subjekt som följer efter sitt huvudverb, eller sådana som ligger längre från sitt huvudord än det antal ord jag angivit ovan. (Denna siffra kan dock anges till oändlig om rutan lämnas tom.)

Flera saker gör att det är svårt att på ett enkelt sätt komma åt samtliga subjekt till verbet *tänka*. Inte minst det faktum att dependensrelationerna inte är ömsesidigt taggade, vilket jag varit inne på tidigare. Optimalt hade varit om jag kunde ange att jag vill söka en relation där ett valfritt ord är subjekt till verbet *tänka* och där verbet *tänka* i sin tur är taggat som finit predikat till just detta subjekt. Med språkteknologens kunskaper kan man i viss mån kringgå sökverktygets begränsningar och göra mer förfinade sökningar, men som vanlig användare har man inte dessa möjligheter.

Jag är också intresserad av vilka ord som förekommer som objekt till verbet *tänka*. I ordbilden ovan i bild 10 är det flera av orden som faller på sin egen orimlighet. Att ordet *tanke* återfinns som nummer 1 är inte så tokigt enligt min intuition. Däremot att nummer 2 och 3 som objekt till verbet *tänka* skulle vara *tank* (behållare) och *tank* (stridsvagn) verkar otroligt och mycket riktigt rör det sig om belägg för *tanke*. Flera av de andra orden verkar kanske inte omöjliga som objekt men förvånande att finna på den här listan. Tittar man i konkordanserna för t.ex. *skapare* rör det sig i samtliga fem fall om samma citat: "*Tänk på din skapare...*" Och så vidare, på det här viset får man alltså analysera sökresultatet för att förstå dess betydelse.

Det är alltså svårt att på ett enkelt sätt leta upp subjekt (och objekt) till ett specifikt verb, eller åtminstone att få en översiktlig bild av resultatet. Ordbildsfliken har sina begränsningar men kan kanske ändå användas som ett slags ingång till analys.

Slutsatser

Korpussamlingarna är fantastiska materialsamlingar med många möjligheter till ingångar för forskning. Med Korp kan man komma över mängder av exempel på de fenomen man är intresserad av även om taggning och frekvenser inte är helt tillförlitliga. Arbetar man kvalitativt bör man också kunna dra slutsatser av de träffar man just har för händerna, även om det finns andra och fler träffar bland de miljoner och åter miljoner korpusorden. Att jämföra träfftyper korpusar emellan bör också kunna vara en möjlighet om man ställer frågorna på rätt sätt.

Jag måste konstatera att de tillgängliga korpusarna utan tvekan tillhandahåller spännande information om användning av orden *tänka* och *tanke*, men att det skulle krävas stora manuella insatser i form av konkordansanalys för att belysa just mina frågor. Det faktum att taggningen är automatisk och därmed en stor felkälla (24–30 % i de träffar jag gått igenom), i kombination med att beläggen inte är betydelseseparerade, gör att sökresultaten är svårtolkade utan manuell konkordansanalys.

Nu var det ju olyckligt att just mitt intresseområde – ordet *tanke* – skulle råka ha böjningsformer gemensamt med ytterligare två andra lemmar (*tank* och *tank2*). Tack vare det har jag ändå fått anledning att fundera ordentligt över både språk och språkteknologi. Jag kan också se att med väl valda frågor om väl valda fenomen kan Korp verkligen vara användbart – både för forskning och i undervisning.

Litteratur

- Church, Kenneth Ward & Patrick Hanks 1990. *Word Association Norms, Mutual Information, and Lexicography*. *Computational Linguistics* 16:1, p. 22–29.
- Grahn, Inga-Lill 2012. *Tala om att tänka. Om processer och projekt vid användning av orden tänka och tanke i tre samtal*. (Göteborgsstudier i nordisk språkvetenskap 18.) Göteborg: Göteborgs universitet, Institutionen för svenska språket.
- SAG = Teleman, Ulf, Staffan Hellberg & Erik Andersson 1999. *Svenska Akademiens grammatik*. Stockholm: Norstedts Ordbok.
- Svensén, Bo 2004. *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- Wittgenstein, Ludvig 1992 [1953]. *Filosofiska undersökningar*. Stockholm: Thales.

GU-ISS, Forskningsrapporter från Institutionen för svenska språket, är en oregelbundet utkommande serie, som i enkel form möjliggör spridning av institutionens skriftliga produktion. Det främsta syftet med serien är att fungera som en kanal för preliminära texter som kan bearbetas vidare för en slutgiltig publicering. Varje enskild författare ansvarar för sitt bidrag.

GU-ISS, Research reports from the Department of Swedish, is an irregular report series intended as a rapid preliminary publication forum for research results which may later be published in fuller form elsewhere. The sole responsibility for the content and form of each text rests with its author.

Forskningsrapporter från institutionen för svenska språket, Göteborgs universitet
Research Reports from the Department of Swedish

ISSN 1401-5919

www.svenska.gu.se/publikationer/GU-ISS