



UNIVERSITY OF GOTHENBURG

## Gothenburg University Publications

### **Språkhistoriskt korpusmaterial i lexikalisk forskning – möjligheter och utmaningar**

This is an author produced version of a paper published in:

**Studier i svensk språkhistoria 13. Red: Daniel Andersson, Lars-Erik Edlund, Susanne Haugen & Asbjørg Westum. Nordsvenska 25. Kungl Skytteanska Samfundets Handlingar 76 (ISSN: 0282-7182)**

Citation for the published paper:

Rogström, L. (2016) "Språkhistoriskt korpusmaterial i lexikalisk forskning – möjligheter och utmaningar". Studier i svensk språkhistoria 13. Red: Daniel Andersson, Lars-Erik Edlund, Susanne Haugen & Asbjørg Westum. Nordsvenska 25. Kungl Skytteanska Samfundets Handlingar 76, vol. 76 pp. 183-197.

Downloaded from: <http://gup.ub.gu.se/publication/237349>

Notice: This paper has been peer reviewed but does not include the final publisher proof-corrections or pagination. When citing this work, please refer to the original publication.

# Språkhistoriskt korpusmaterial i lexikalisk forskning – möjligheter och utmaningar

---

**ABSTRACT** The purpose of this article is to determine the extent to which it is possible to use historical text corpora in lexical research. In Språkbanken (The Swedish Language Bank) at the Department of Swedish, University of Gothenburg, the linguistic resource called Korp is being extended with historical text material of various kinds. A great deal of this material is made up of juridical texts, but there are also newspaper materials and different kinds of special purpose texts. In this paper I use two different genres of corpus material – juridical text and newspapers – from the 17th to the 20th century in order to see how the total of twelve different words belonging to three different semantic fields appear and vary according to the two genres. The results show that corpora made up of historical texts must be used with great care because the orthographical and morphological variation is very hard to handle. However, the results also show that the words are used and develop somewhat differently depending on genre. This is a good starting point for further analyses of the words' lexical development according to polysemy, genre, and the growth of languages for special purposes during the 18th century.

---

**KEYWORDS** historical text corpora, lexical development, polysemy

---

## 1. Inledning och syfte

På konferensen Svenska språkets historia 12 i Stockholm i april 2012 menade Lars-Gunnar Andersson att det kan förefalla märkligt att språket ändrar sig i den omfattning som det ändå gör, med tanke på att ingen verkar tycka att språkförändringar är särskilt positiva och att den dominerande åsikten är att "det var bättre förr" (Andersson 2014:11). För en språkhistoriker ger ju detta en extra dimension åt *varför*-frågan i forskningen; en fråga vars svar är underkastat ett i många fall ganska enahanda och litet material, produce-

rat av sedan länge döda språkbrukare. Utgångspunkterna för forskningen är därmed inte alltid optimala.

Denna artikel syftar därför inte till att svara på frågan *varför* utan är snarast ett försök att ta sig an frågan *hur*, med vissa utblickar åt *varför*. I föreliggande undersökning testas om elektroniska korpusar kan användas för att undersöka hur äldre ordmaterial utvecklas över tid. Resultaten skulle i sin tur kunna läggas till grund för en undersökning av *varför* materialet ser ut på ett visst sätt och *varför* det utvecklats i en viss riktning, men denna fråga kommer jag bara att beröra helt ytligt. Artikeln syftar främst till att diskutera fördelar och svårigheter med att använda elektroniska korpusar över historiska material och korpusarnas betydelse när man undersöker lexikaliska förändringar över tid.

Att visa på lexikaliska utvecklingsmönster genom att undersöka stora textmaterial är inget nytt. Vår absolut mest omfattande ordbok på och om svenska, SAOB, skildrar mycket ingående hur enskilda ord utvecklats genom att utgå från hur orden använts i text från olika tider. Detta ger naturligtvis en bra utgångspunkt till en beskrivning av ett enskilt ords utvecklingsprocess, men metoden ger ingen upplysning om hur flera ord med liknande betydelse förhåller sig till varandra och man kan heller inte se i vilken omfattning ett ord använts under vissa tider i förhållande till andra, liknande ord. Flera ordboksartiklar kan givetvis läggas till grund för jämförande undersökningar men det är ytterst svårt att säga något om utvecklingen mellan orden eftersom några frekvensberäkningar på ordens belägg under olika tidsperioder inte anges i SAOB. Bruklighetsangivelserna betecknar därtill ordets, eller den enskilda betydelsens, bruklighet vid den tidpunkt då artikeln skrevs, så dessa går heller inte att förlita sig på om man vill undersöka växelpelet mellan vissa ord. För en sådan undersökning behöver man stora textmaterial i vilka man kan göra kvantitativa undersökningar av de enskilda orden och följa dem över tid (se t.ex. Geeraerts *et al.* 1994). Sådana korpusar har hittills inte funnits tillgängliga för den äldre svenskans vidkommande, men man har nu börjat bygga korpusar också över äldre texter, vilket öppnar en del nya vägar för lexikalisk språkforskning.

Språkbanken vid institutionen för svenska språket, Göteborgs universitet, har påbörjat ett arbete med att skapa en korpus för äldre språkmaterial liknande den som redan finns för modern svenska. Det är denna korpus som ligger till grund för min undersökning.

## 2. Lexikalisk förändring i gränslandet mellan allmänspråk och fackspråk

Sedan 2012 har jag ett projekt, *Lexisering bakom kulisserna*, finansierat av Riksbankens jubileumsfond, vars syfte är att undersöka principerna bakom

etableringen av ett svenskt ordförråd under 1700-talet. Det ordförråd som främst undersöks emanerar ur diskussioner om enskilda ord, dokumenterade i protokollen till arbetet med 1734 års lag. En stor del av diskussionerna i protokollen handlar om hur man ska skriva en lag på svenska med ett både högtstående och begripligt språk. Inte minst juridiska ord på latin förorsakar lagkommissionen bekymmer och dessa har jag lagt till grund för flera olika undersökningar i syfte att kartlägga förutsättningar för etablering av motsvarande svenska ord. Även om svenska lagar alltid har varit skrivna på svenska kom det juridiska fackspråket under åren att alltmer influeras av latin. Den latinska diskursen syns tydligt i de protokoll som upprättades under sammanträdena i Karl XI:s lagkommission som tillsattes i december 1686. Att bruket av svenska och latin varierar beroende på om kontexten är allmänspråklig eller fackspråklig går att se i flera domäner under 1700-talet och en språklig strävan under denna tid var att försöka utöka ordförrådet så mycket att svenska språket kunde användas inom alla områden, även vetenskapliga och fackspråkliga sådana (se t.ex. Teleman 2002).

Undersökningarna av hur lagkommissionen tog sig an sin språkliga uppgift har därför delvis kommit att kretsa kring problematiken kring allmänordförråd och fackordförråd i juridisk kontext. Det juridiska fackspråket är speciellt såtillvida att det i högre utsträckning än i många andra fackområden också används av lekmän, vilket får betydelse för utvecklingen av enskilda facktermer (se också Rogström 1998, kap. 8). I föreliggande undersökning används det elektroniska korpusmaterialet i ett försök att se om man kan urskilja olika förekomster och bruk av juridiska ord i olika typer av korpusar och koppla detta till utvecklingen av ett fackordförråd.

I några tidigare undersökningar visas att lagkommissionens arbete kan betraktas som en del i ett svenskt språkplanerande arbete (Rogström 2009) och att det verkar troligt att lagkommissionen i sitt språkliga arbete med de lexikaliska enheterna i stor utsträckning får understöd av lexikografiska arbeten från samma tid (Rogström 2010, 2013). Det visar sig dock att en del av det som förefaller vara lämpliga utbytesord till de latinska juridiska orden senare återtar sin latinska språkdräkt, för att idag åter igen användas som fullt gångbara juridiska ord med latinskt ursprung (Rogström 2014). Denna återgång till latinskt baserad vokabulär kan troligen spåras till utbytesordens utveckling av polysema betydelser där den juridiska betydelsen blir allt svagare till förmån för den mer allmänspråkliga betydelsen, samtidigt som den svenska språknormeringen blir allt fastare. Denna utveckling kan då underlätta en process där det latinska ordet åter kommer till större användning (som i *instans* > *domstol* > *instans*), eller där en viss juridisk betydelse av ett ord försvagas så att man i stället övergår till ett nytt, mer inhemskt ord (*appellera* > *vädja* > *överklaga*). Denna växelverkan mellan olika ord och deras

betydelser samt sambanden till framväxten av en svensk språknorm skulle vara tacksam att studera i ett större korpusmaterial där ordens utveckling kan kopplas både till epoker och frekvens – något som Korp skulle kunna bidra till. Den historiska Korp behöver först undersökas närmare, eftersom historiskt textmaterial är mer komplicerat att behandla digitalt än modernt material som följer tydliga språkliga normer.

### 3. Material

#### 3.1 Urval av ord

Undersökningen baseras på tre begrepp som diskuteras i lagkommissionens protokoll med målet att deras beteckningar bör bytas ut mot något lämpligare än det de hade i juridiskt språkbruk under 1600-talet. Orden är valda för att de tillhör den grupp ord som delvis återtar en latinsk språkdräkt och för att de betecknar tre centrala företeelser i den juridiska processen som går att följa från medeltida texter: att lämna in en ansökan till en domstol, dvs. *anhängiggöra* en sak; att begära överprövning av en dom eller ett beslut, dvs. *överklaga* samt platsen man gör detta på, både i fysisk och juridisk mening, dvs. *instansen*. Eftersom orden skiftar starkt i stavning i materialet använder jag här genomgående modern stavning. De latinska ord som användes för dessa begrepp var i lagkommissionens material *agera*, *appellera* och *instans*. Orden *latinsk* och *svensk* används i texten något förenklat, där framför allt *svenskt* egentligen ska läsas som *icke-latinskt ursprung*.

För att bilda mig en uppfattning om ordens allmänna förankring i juridisk diskurs vid denna tid har jag kontrollerat förekomster och användning gentemot en del andra juridiska material, framför allt de juridiska handböckerna av Rålamb (1674) och Kloot (1676), vilka båda är skrivna på svenska.

De tre begrepp som undersöks representeras redan i medeltidslagarna och man vet därför vilka fornsvenska ord som används i de aktuella lagarna som tillämpades fram till 1734. Vi vet också vilka ord som används för samma begrepp idag. Man kan dessutom anta att så pass centrala ord inte varierar särskilt mycket eftersom de har en fackordsprägel. Detta begränsar antalet sökningar som måste göras i materialen. Det är ändå omöjligt att veta att man undersökt alla tänkbara beteckningar för ett visst begrepp och denna undersökning är på intet sätt fullständig i detta avseende. Jag menar dock att de viktigaste orden som användes under 1600- och 1700-talen för respektive begrepp kommit med i undersökningen, och att man kan dra vissa slutsatser trots att ordmaterialet troligen hade kunnat utökas.

De juridiska begrepp som under 1600-talet i stor utsträckning hade fått latinskt grundade beteckningar ändras alltså på något sätt i 1734 års lag, och i

flera fall följer beteckningarna lagkommissionens rekommendation. Under årens lopp kan man dock se att orden genomgått fortsatta förändringar och ibland också återtagit sin latinska form. Idag används t.ex. *instans* som juridisk fackterm igen. *Vädja* har försvunnit till förmån för *överklaga* medan *appellera* inte är helt obrukbart som simplex och ganska frekvent i ordbildningar, t.ex. i sammansättningen *appellationsdomstol*. *Agera* har lämnat spår i ett ord som *akt*, och frasen *dra inför rätta* eller *stämman (inför rätta)* används ofta som en vardagligare variant till *anhängiggöra* eller *väcka talan*.

Följande ord har undersökts i Korps historiska korpusar:

Begrepp 'anhängiggöra'	Begrepp 'överklaga'	Begrepp 'instans'
Agera	Appellera	Domstol
Anhängiggöra/göra anhängig	Vädja	Domställe
Dra/gå/komma/svara till rätta	Överklaga	Instans
Stämman		Rätt
Väcka talan		

### 3.2 Korpusarna

Det sker alltså ett växelspel mellan de olika ord som använts för de tre juridiska begreppen och detta växelspel borde kunna säga en del om mekanismerna bakom språkförändringar och framväxten av fackterminologi. Men för att komma åt växelspelet behöver man studera hur orden använts genom tiderna, vilket i sin tur underlättas om man har tillgång till texter som gärna är sökbara elektroniskt, eftersom sökandet efter enskilda ord annars skulle bli orimligt tidskrävande.

Det finns emellertid ett antal försvårande begränsningar när det gäller korpusar för äldre material. En första begränsning är att en korpus aldrig kan innehålla fler ord än de material som finns bevarade. Ju äldre material man vill undersöka, desto färre texter har man tillgång till.

En andra begränsning är att stora delar av det material som finns bevarat inte är tillgängligt digitalt. Materialen är svåra och kostsamma att digitalisera och de behöver ofta skrivas in för hand eller bearbetas kraftigt manuellt. De material som digitaliserats maskinellt har ibland många läsfel (dvs. datorn har inte tolkat tecknen i texten på rätt sätt vid inscanningen), vilket också minskar användbarheten.

I modern korpshantering vill man gärna använda väl avvägda korpusar som återspeglar den typ av språk man vill undersöka för att resultaten ska vara tillförlitliga (se t.ex. Geeartes *et.al* 1994). Detta krav går inte att ställa på historiska material. Språkbankens historiska korpusar består för tillfället av alla möjliga material som har råkat vara tillgängliga, samt några material som har digitaliserats speciellt för ändamålet. Detta gör att enskilda ord kan

slå igenom väldigt starkt frekvensmässigt om de förekommer i stor omfattning i någon enstaka text.

En tredje begränsning är att man ännu inte löst problemet med att återföra all stavnings- och böjningsvariation i materialet till en enhetlig form (se Adesam *et al.* 2016). För det moderna materialet i Korp finns långt fler sökmöjligheter än för den historiska korpusen (se Grahn 2014 för en utförlig beskrivning och utvärdering av sökningar i Korp). I den historiska Korp går det således inte att samla alla former av det sökta ordet under en post. Detta innebär att användaren än så länge måste utföra varje sökning flera gånger om söksträngen i fråga kan stavas på flera sätt. Alla stavningsvarianter kan ibland vara svåra att hitta och är kanske inte alltid värda att lägga extra tid på, men de vanligaste varianterna är viktiga att täcka in.

### 3.3 Urval av korpusmaterial

Materialet i den historiska korpusen spänner över en lång tidsperiod, från medeltid fram till modern tid, och genrerna är blandade. Eftersom en stor del av det fornsvenska materialet utgörs av medeltidslagar som skrivits in i andra projekt, framför allt Fornsvenska textbanken finns det juridiskt material från flera epoker. Särskilt värdefulla för denna undersökning är de digitaliserade protokollen från Lagkommissionens arbete (Sjögren 1900–1909) samt den för hand inskrivna 1734 års lag. Utöver dessa material har moderna lagar från 1800-talet använts samt en del material från internet i form av pdf-filer av inskrivna domböcker. Dessa har alltså inte funnits tillgängliga i form av korpuser utan har endast hanterats i pdf-format.

Det juridiska materialet i Korp återfinns digitalt under [www.spraakbanken.gu.se/korp](http://www.spraakbanken.gu.se/korp). Protokollen ur domböckerna finns elektroniskt utgivna på internet av Jämtlands läns fornskriftsällskap <http://fornskrift.se/Avskrifter.htm>. Omfattningen av materialet framgår av uppställningen nedan.

#### Lagmaterial från Korp:

Stockholms stads tänkeböcker 1626	121 366 token
1734 års lag	980 120 token
Föarbetena till 1734 års lag	1 603 126 token
Lagar från 1800-talet (Regeringsformen 1809, 1974) (Författningssamling Låssa kyrkas arkiv 1800)	44 604 38 token
Moderna lagar och rättsfall (Svensk författningssamling) (Domar)	40 983 094 token

Protokoll ur domböcker (endast pdf):		
<i>Underåkers tingslags domboksprotokoll</i>		1649–1690
Tingsprotokoll 183 sidor		
<i>Revsunds tingslags domboksprotokoll</i>		1649–1700
Tingsprotokoll 355 sidor		
<i>Hammerdals tingslags domboksprotokoll</i>		1649–1690
Tingsprotokoll 183 sidor		
<i>Härjedalsbrev</i>		1531–1645
Ca 170 brev (diplom)		

Lagmaterialet är således ganska olika i omfattning med en klar övervikt för det modernare materialet från 1800-talet.

Tillgången på allmänspråkligt material är begränsat i den historiska Korp. Kungliga biblioteket har tillgängliggjort ett tidningsmaterial som kallas Kubhist, och detta har legat till grund för de allmänspråkliga sökningarna. Kubhist är emellertid av ganska ojämn kvalitet genom att det innehåller väldigt många läsfel och materialet är därmed inte helt tillförlitligt.

Omfattningen av korpusarna framgår i uppställningen nedan:

Historiskt tidningsmaterial från Korp (20 tidningar):

Kubhist	1749–1925	877,5 milj token
---------	-----------	------------------

Tidningsmaterialet är således mycket större än lagmaterialet. Det tidiga materialet är minst, och de fyra tidningar som kan sägas representera 1700-talet utgör en tiondel av det totala materialet. De fyra tidningarna är också mycket olika stora, vilket syns i uppställningen nedan.

Utsnitt ur Kubhist (4 tidningar från 1700-talet)		85,6 milj. token
Fahlu weckoblad	1786–1821	(3 milj. token)
Göteborgs weckolista	1749–1757	(607 000 token)
Post- och inrikes tidn.	1775–1860	(41 milj. token)
Stockholmsposten	1778–1833	(41 milj. token)

### 3.4 Söksträngarna

I stora korpusar kan man snabbt få fram siffror rörande olika ords förekomst och användning utan att behöva läsa materialen i sig. För att arbetet ska bli så tillförlitligt som möjligt är det viktigt att täcka in alla stavnings- och böjningsvarianter och i vissa fall ordbildningsformer. Det är också en stor fördel om orden kan homografsepareras, så att olika ordklasser hos samma ordform kan föras till olika lemman. Ordet *stämma* kan ju vara antingen verb eller substantiv och måste därför fördelas på två olika lemman. Om



ordklassfördelningen kan göras automatiskt underlättas arbetet betydligt. Ett lemma kan dessutom vara polysemt till sin betydelse (t.ex. *stämma* (vb) i musikalisk eller juridisk betydelse), och denna uppdelning är betydligt svårare att komma åt på automatisk väg och kräver för historiska material en del manuell analys.

För att slippa att göra ett stort antal sökningar på varje enskilt ord bör söksträngarna därför helst utformas på så sätt att de täcker in så många varianter som möjligt. Detta innebär i sin tur att man får träff på en del ordformer som är ointressanta i sammanhanget och man måste därför avväga om fördelen med att få många träffar uppväger nackdelen med att en del träffar är fel.

Varianterna av *instans* kan tjäna som exempel på resonemanget. Ordet har i mitt material ett flertal stavningar varav de vanligaste är *instance*, *instantz*, *instantie*. För att slippa göra olika sökningar på varje stavningsvariant används då söksträngen *instan.\** där de olika stavningsvarianterna täcks in av jokertecknet *[.]* som kan upprepas valfritt antal gånger *[\*]*. Resultatet av denna sökning ger 41 olika ordformer (types) där de flesta är stavnings- och böjningsvarianter av *instans* såsom *instans*, *instance*, *instantier*, *instantz*, *instantierne* etc.

Man får också fram sammansättningar med förledet *instans*: *instansordning*, *instansordningsprincip*, *instansordningsgrundsatsen*, *instansvägar* etc. Dessa kan i förstone verka ointressanta för ändamålet, men de visar ändå att ordet *instans* blivit så etablerat att det bildat sammansättningar. (Om förhållandet mellan simplex och ordbildningsformers förekomst i stora material, se Geeraerts *et al.* 1994:32ff.) Därutöver får man fram helt obesläktade ord: *instannande*, *instansning*, *instansats* etc. Dessa är dock så få att jag bedömer att de inte får någon betydelse för analysen.

Materialets beskaffenhet är således en begränsning för undersökningen jämfört med hur man kan använda korpusar med modernt material. I just denna undersökning är inte antalet ord större än att man hade kunnat göra enskilda sökningar på alla vanliga varianter var för sig, men eftersom tanken är att testa Korps användbarhet i språkhistoriska undersökningar har jag ändå valt att använda de söksträngar som kan tänkas ge bäst resultat i en enda sökning. Genom att använda t.ex. söksträngen *appell.\** i stället för ordet *appellera* kan man få träffar på alla ord som innehåller teckenföljden *appell*, t.ex. närbesläktade ord som *appellation*, *appellant* etc. Om ordet ifråga inte hittas i den exakta form man velat undersöka, men kan hittas i besläktade ordformer i form av avledningar, kan detta samtidigt vara ett tecken på att ordet trots allt användes, men att det undersökta materialet av någon anledning, exempelvis storlek, inte innehåller just den efterfrågade ordformen. I material som inte är så stora är detta en fördel.

## 4. Resultatredovisning

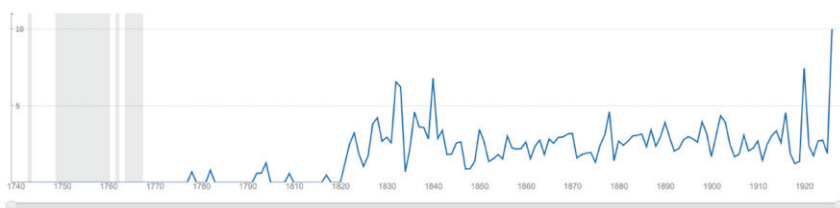
Ett vanligt sätt att redovisa kvantitativa undersökningar på är att ange relativ frekvens för de enheter man undersökt och jämfört. Detta är av flera skäl något svårt att göra i föreliggande undersökning. En första anledning är att materialen till viss del ser olika ut och att de pdf-filer som undersökts inte är möjliga att räkna frekvens på eftersom de endast föreligger som rena textfiler och inte införlivats i Korp. Ett andra skäl är att några tidningar i Kubhist är så små att också relativa frekvensberäkningar kan bli otillförlitliga genom att ganska få belägg på ett visst ord ändå kan få ett stort frekventiellt genomslag om den totala ordmängden i materialet är liten. Geeraerts *et al.* (1994:32ff.) diskuterar "lexikalisk mättnadsgrad" där storleksökningen på en korpus inte längre får betydelse för frekvensen för enstaka ord. Denna mättnadsgrad har alltså inte uppnåtts i alla de historiska materialen i Korp, vilket kan ge väldigt skeva siffror också vid beräkning av relativ frekvens. De ganska allmänt hållna söksträngar som använts fångar också upp en hel del ord som är helt irrelevanta för undersökningarna, något som t.ex. gäller ett antal sammansättningar som inleds med ordledet *rätt*. Detta ger också utslag i frekvensen. Några ord är dessutom starkt homonyma, såsom *stämma* och *rätt*, vilket innebär att deras frekvens bara av den anledningen blir mycket högre. Jag har därför valt att inte presentera frekvenssiffror över materialet utan i stället diskutera ordens utveckling i mer generella termer och relatera resultaten till materialen över tid. Detta gör naturligtvis inte felkällorna mindre, men man kommer ifrån den känsla av, i detta fall, falsk objektivitet som en konkret sifferpresentation kan ge. På detta vis är det också möjligt att införliva pdf-materialet i undersökningen på samma villkor som det övriga materialet. Genom att betrakta ordens förhållande till varandra över tid får man ändå en överblick över deras växelspel utan att behöva ta hänsyn till exakta siffror. Denna överblick kan sedan läggas till grund för en mer noggrann genomgång i SAOB där ordens betydelseutveckling kan analyseras.

Lagmaterialet redovisas därför med markeringar i form av plustecken för varje ord, där fler plustecken betyder stor förekomst. Plustecken inom parentes markerar att ordet förekommer i ytterst få fall, och ett streck markerar att ordet inte förekommer alls i materialet.

För Kubhistmaterialet, som ju finns i konkordansform helt och hållet, är det speciellt vanskligt att jämföra frekvenser mellan materialen eftersom framför allt delar av 1700-talsmaterialet består av väldigt få token (se uppställningen ovan). Jag har därför utnyttjat Korps funktion att kunna rita s.k. trenddiagram. Dessa bygger också på relativ frekvens men noterar varje enskilt belägg i ett diagram utifrån det år som belägget publicerades, varvid man kan följa ett ords utveckling via dess relativa frekvens över varje år. På så sätt får man en överblick över hur orden utvecklas generellt över tid

och om de förekommer i högre utsträckning under någon särskild period. Genom att föra muspekaren över diagrammet får man siffror på både absoluta träffar och relativ frekvens för varje år. Enskilda "spikar" i diagrammet kan därför lätt undersökas och analyseras. Ofta beror dessa på att ett litet antal belägg av det undersökta ordet slår igenom i små tidningsmaterial. Diagrammet har den fördelen att det ger en samlad bild över utvecklingen av ett viss ord, och att de enskilda topparna går att förklara ganska enkelt.

Figur 1 visar utvecklingen av *instans*\*. Där kan man se att ordets användning ökar över tid och att den stora ökningen sker på 1800-talet. Det totala antalet belägg för ordet är dock ganska litet, oftast under fem belägg per år.



Figur 1. Skärmdump av trenddiagram över sökningen *instans*\* (Hämtad 2014-09-02)

#### 4.1 Resultat: 'anhängiggöra'

##### Juridiskt material

	agera	anhängiggöra	vb + till rätta	stämma	väcka talan
1600-tal	+	(+)	+	++	-
1734	-	-	(+)	+++	-
1800-tal	(+)	-	(+)	+++	-
1900-tal	(+; ej jur)	+++	+	++	++

Begreppet 'anhängiggöra' betecknas i det juridiska materialet främst av orden *stämma* och just *anhängiggöra*. Det sista ordet finns i princip enbart i 1900-talsmaterialet medan *stämma* förekommer under hela perioden. *Agera* finns inte i så många belägg över huvud taget, men de förekomster som hittats finns framför allt i 1600-talsmaterialen. Här kan man troligen se att det latinska ordet *agera* övergår till att användas i en annan betydelse som är mer allmänspråklig och därför har låg förekomst i det juridiska materialet. Beläggen från 1900-talet betecknar vanligen den betydelse som kopplas till att agera på en scen. *Väcka talan* finns endast i 1900-talsmaterialen. Som fackord används framför allt *stämma* och *anhängiggöra*.

I tidningsmaterialet kan man se att *agera* har flest förekomster under 1700-talet och att trenden därefter sjunker. *Anhängiggöra* kommer starkt

runt 1835 för att därefter ha en hög förekomst. *Stämman* ökar tydligt från 1790-talet medan frasen *väcka talan* endast har 24 träffar totalt. *Vb+till rätta* har en ganska jämn förekomst, men har vanligen ingen juridisk betydelse utan företräds framför allt av fraserna *komma/ställa till rätta*.

Tidningsmaterialet, som borde vara mer allmänspråkligt till sin karaktär, har ganska få belägg av rena juridiska fraser såsom *väcka talan*. Att *agera* inte minskar lika mycket som i den juridiska korpusen kan förklaras med att det är "teaterbetydelsen" som har flest belägg. *Stämman* är det ord som ökar mest, och detta kan förklaras utifrån ordets höga grad av såväl homonymi som polysemi samt att ordets juridiska betydelse är ganska allmänspråkligt företrädd.

Det går således att se att de juridiska orden (*anhängiggöra, stämman, väcka talan*) blir vanligare i juridiska material medan de ord som har både allmän och juridisk betydelse (*agera, stämman*) används mer i tidningsmaterialet.

#### 4.2 Resultat: 'instans'

##### Juridiskt material

	domstol	domställe	instans	rätt(en)
1600-tal	(+)	+	+	++
1734	+	-	-	++
1800-tal	++	-	(+)	+
1900-tal	++	-	+	++

I det juridiska materialet är det orden *rätt(en)* samt *domstol* som är dominerande under hela perioden. *Instans* har vissa förekomster redan under 1600-talet men försvinner under 1700-talet, för att sedan återkomma under 1900-talet. Ordet har då en tydligt juridisk fackordsprägel, och har övertagit en del av de betydelser som *domstol* hade under äldre tid. *Rätt*, som är ett ord som känns igen från medeltida lagspråk användes mycket frekvent under hela den undersökta perioden (liksom idag.)

I tidningsmaterialet är det *domstol* och *rätt* som förekommer i ungefär lika stor utsträckning medan *instans* har en dipp under 1700-talet. *Domställe* har endast 4 belägg, och var därmed inget framgångsrikt förslag till ersättningsord från Lagkommissionens sida.

Utvecklingen av betydelsen hos *domstol* och *instans* väver egentligen en mer utförlig beskrivning relaterad till omdaning av svenskt rättsväsende under 1800-talet (se t.ex. Modéer 2010), vilken naturligtvis också återspeglas i språket. I denna undersökning får jag dock nöja mig med att koppla ordet *domstol* och dess omfattande förekomster till den generella användningen utan att redogöra för den mer fackspråkliga utveckling som skett hos orden *domstol* och *instans*.

### 4.3 Resultat: 'överklaga'

#### Juridiskt material

	appellera	vädja	överklaga
1600-tal	++	++	(+)
1734	-	++	(+)
1800-tal	-	+	++
1900-tal	(+)	+	++

*Appellera* förekommer mest i det äldsta materialet men har några enstaka belägg i 1900-talsmaterialet, vilket också framgick i resultaten från den lexicografiska undersökningen i Rogström (2013). Ordet förekommer flitigare i sammansättningar, vilket visar att det blivit befäst i sin latinska form. Det vanligaste ordet på 1600- och 1700-talen är annars *vädja*, som under 1800-talet övertas av *överklaga*.

I tidningarna är *vädja* vanligast från början och ökar mest, därefter kommer *överklaga* som är störst under 1900-talet. *Appellera* används minst och ökar minst men har flest belägg under 1770-talet fram till 1820. I relationen mellan dessa ord ser man i tidningsmaterialet att de ord som har mest koppling till allmänordförrådet också har flest belägg.

## 5. Sammanfattning

Ordmaterialen som undersökts har haft en spridning från rena fackord (*appellera, instans*) till ord som också har mer allmänspråklig karaktär samt är mer homonyma (*rätt, stämna, vädja*). Den generella utvecklingen som syns i de olika materialen visar att de ord som har mer fackspråklig betydelse är vanligare i det fackspråkliga materialet. Detta märks tydligt för orden *anhängiggöra, väcka talan, instans* och *överklaga*. De juridiska ord som också kan användas i allmänspråklig betydelse är vanligare i tidningsmaterialet (*rätt, stämna, vädja*). Man kan också se att de latinska orden är vanligare i fackspråksbetydelsen och latinska ord som har många träffar i tidningsmaterialet då har det i sin allmänspråkliga betydelse, såsom *agera*. Polysema ord där betydelserna är dels juridiska, dels allmänspråkliga är generellt sett fördelade så att den juridiska betydelsen är vanligare i de juridiska materialen och den allmänspråkliga i tidningsmaterialet. Detta skulle i så fall också stämma med resultaten i Rogström (2013), som visar att ordens polysemi i samtida ordböcker utvecklas så att den fackspråkliga betydelsen ofta får stå tillbaka för den allmänspråkliga (*agera, appellera*), och att den allmänspråkliga betydelsen under 1800-talet skildras som den vanligare betydelsen gentemot den juridiska.

Juridiskt fackspråk är intressant såtillvida att det i högre grad än många andra fackspråk hamnar i något slags gränsland mellan allmänspråk och

fackspråk. Just för att den vanlige medborgaren berörs av juridiken i sin vardag är det nödvändigt att hen både förstår och kan använda sig av juridiskt fackspråk, åtminstone till viss del. Denna tanke var ju central för Lagkommissionen genom riktlinjen att lagen skulle vara begriplig och användbar för alla. Rimligen utvecklar sig då det gemensamma juridiska fackspråket något olika bland experter och allmänhet vilket också syns i denna undersökning.

## 6. Går det att använda elektroniska korpusar för språkhistoriska undersökningar?

Det huvudsakliga syftet för denna undersökning var att se om man kunde använda Korps historiska resurser till språkhistoriska undersökningar rörande förändringar i ordförrådet. Slutsatsen är ett mycket försiktigt ja. Undersökningen har visat att man ser skillnad mellan de olika materialen i fördelningen mellan fackord och allmänord och att man kan dra vissa slutsatser om utvecklingen av polysema ord som är värda att gå vidare med vad gäller kartläggningen av olika ords etablering över tid. Men undersökningen har också visat på vissa svårigheter i användandet av materialen.

Det är för det första svårt att jämföra material som är så olika stora och har så olika bakgrund. Man bör ha bra kännedom om materialen för att kunna använda dem rätt.

Sökningarna hade underlättats betydligt om orden varit homografseparerade. Detta är emellertid ett problem som har sin grund i den stora form- och stavningsvariation som finns i äldre material och som man arbetar hårt på för att lösa. Dock löser inte homografseparering alla problem eftersom den semantiska disambigueringen trots det återstår. Till detta ska också läggas problem som kan uppstå beroende på att den automatiska taggningen inte alltid analyserar alla belägg rätt. Också i de stora korpusarna över modernt material i Korp måste kvantitativa sökningar kompletteras med korpusanalys av beläggen, inte minst beroende på mängden feltagningar (se Grahn 2014).

Den tveksamma kvaliteten på Kubhistmaterialet är en klart negativ faktor. Just allmänspråkligt material är extra värdefullt när man undersöker språk ur ett historiskt perspektiv, eftersom det ofta domineras av fackspråk, inte sällan religiöst och juridiskt. Sammantaget betyder detta att man får lägga ner ett ganska stort arbete på manuella kontroller av materialet om man vill ha någorlunda säkra resultat, i alla fall för lexikologiska undersökningar.

Att kunna utnyttja elektroniska material är en bra utgångspunkt i språkliga undersökningar men oftast behöver de kompletteras med manuella kontroller. Grahn (2014) visar hur pass stora felkällor det kan vara i ett mo-

dernt material som taggats syntaktiskt, och att använda ett historiskt material till lexikaliska sökningar bjuder troligen på ännu fler felkällor. Detta beror dels på korpusarnas sammansättning av tillgängligt material, dels på svårigheterna att göra korrekta automatiska analyser av ett språk som inte är ortografiskt och morfologiskt normerat till en och samma norm. Tolkning av äldre språkmaterial erbjuder dessutom ett flertal andra utmaningar vad gäller korpusanalys. Trots detta vill jag påstå att Korp kan utgöra ett bra incitament för lexikaliska undersökningar. Det går att vaska fram material som man på det viset slipper att excerpera för hand, men materialet måste fortfarande genomgå en omfattande manuell granskning och analys.

## KÄLL- OCH LITTERATURFÖRTECKNING

---

### KÄLLOR

Fornsvenska textbanken. <<http://project2.sol.lu.se/fornsvenska/>>;

Jämtlands läns fornskriftsällskap. <<http://fornskrift.se/Avskrifter.htm>>; hämtad 6 januari 2014.

Språkbanken, Korp. <<http://spraakbanken.gu.se/korp>>;

### LITTERATUR

Adesam, Yvonne, Ahlberg, Malin & Bouma, Gerlof 2012. Processing spelling variation in historical text. *Proceedings of the Fourth Swedish Language Technology Conference (SLTC)*. Lund: Lunds universitet.

Adesam, Yvonne, Ahlberg, Malin, Andersson, Peter, Borin, Lars, Bouma, Gerlof & Forsberg, Markus 2016. Språkteknologi för svenska språket genom tiderna. I denna volym s. 65–87.

Andersson, Lars-Gunnar 2014. Svensken om svenskan: om synen på variation och förändring I: M. Bylin, C. Falk & T. Riad (red.), *Studier i svensk språkhistoria 12. Variation och förändring*. (Acta universitatis Stockholmiensis. Stockholm studies in Scandinavian Philology. New series 60.) Stockholm: Stockholms Universitet. S. 11–26.

Geeraerts, Dirk, Grondelaers, Stefan & Bakema, Peter 1994. *The structure of lexical variation. Meaning, naming, and context*. (Cognitive Linguistics Research 5.) Berlin: Mouton de Gruyter.

Grahn, Inga-Lill 2014. Korpblickar – om möjliga ingångar med sökverket Korp för studier av orden *tänka* och *tanke* i korpusar. GU-ISS. Forskningsrapporter från Institutionen för svenska språket. <<http://www.svenska.gu.se/publikationer>>

Kloot, Claudius 1676. *Then svenska lagfarenheets spegel*. Götheborg.

Modéer, Kjell Å. 2010. *Historiska rättskällor i konflikt. En introduktion i rättshistoria*. Stockholm: Santérus.

Rogström, Lena 1998. *Jacob Serenius lexikografiska insats*. (Meijerbergs arkiv för svensk ordforskning 22.) Göteborg: Institutionen för svenska språket, Göteborgs universitet.

Rogström, Lena 2009. Språkplanering i praktiken? En lexikalisk studie av förarbetena till 1734 års lag. *Språk och stil*. 19. S. 5–34.

Rogström, Lena 2010. I akt och mening. En jämförelse av ordförrådet i 1734 års lag med samtida ordböcker. I: M. Reinhammar & L. Elmevik (red.), *Studier i svenska språkets*

- historia*. II. (Acta Academiae Regiae Gustavi Adolphi CXIII.) Stockholm: Kungl. Gustav Adolfs Akademien för svensk folkkultur. S. 189–199.
- Rogström, Lena 2013. Lagspråk, lexikon och lexisering: en studie av språklig samverkan under 1700-talet. *LexicoNordica*. 20. Oslo: Nordisk forening for leksikografi i samarbeid med nordisk språksekretariat. S. 167–184.
- Rogström, Lena 2014. Från dilation till uppskov. Var ordböcker viktiga i etableringen av ett juridiskt ordförråd? *Rapport fra Konferanse om leksikografi i Norden*. Oslo 13–16 august 2013. (Nordiske studier i leksikografi 12.) Oslo: Nordisk forening for leksikografi. S. 385–401.
- Rålamb, Clas 1674. *Observationes juris practicæ*. Stockholm.
- SAOB = *Ordbok över svenska språket, utgiven av Svenska Akademien*. 1898–. Lund: Gleerups.
- 1734 års lag. Se *Sveriges rikes lag*.
- Sjögren Wilhelm 1900–1909. *Förarbetena till Sveriges rikes lag: 1686–1736: efter offentligt uppdrag utgifna af Wilhelm Sjögren*. Uppsala.
- Sveriges rikes lag. Gillad och antagen på Riksdagen åhr 1734*. 1780. [1984]. (Skrifter utgivna av Institutet för rättshistorisk forskning grundat av Gustav och Carin Olin. ) Lund.
- Teleman, Ulf 2002. *Ära, rikedom och reda. Svensk språkvård och språkpolitik under äldre nyare tid*. Stockholm: Norstedts.

Lena Rogström  
Institutionen för svenska språket  
Göteborgs universitet  
lena.rogstrom@svenska.gu.se