



GÖTEBORGS UNIVERSITET
Institutionen för svenska språket

Lexikaliska institutet
Institutet för svenska som
andraspråk
Språkbanken
Språkhandledningen

En webbaserad svensk akademisk ordlista

Carina Carlund, Sofie Johansson Kokkinakis,
Julia Prentice & Emma Sköldberg
Göteborg, november 2011

Det akademiska språket – en utmaning för många

Det svenska akademiska ordförrådet, dvs. ord som förekommer i universitetstexter oavsett vilket ämne som behandlas, orsakar problem för akademiska nybörjare – oberoende av vilket modersmål de har!

Innehåll i föredraget

- Carlund (2009): Undersökning av 11 verb i hemtentamina på Socialt arbete
- A. Coxheads *Academic Word List* (2000)
- Mot en svensk akademisk ordlista
- Nordiskt nätverk kring ämnet
Språkbruk i nordiska akademiska miljöer

Carlund (2009): 11 "svåra" ord i fokus

- Exempel på "svåra verb" som lyfts fram i samband med Språkhandledning (inom Göteborgs universitet):

analysera, argumentera, beskriva, diskutera, karakterisera, motivera, problematisera, redogöra, referera, reflektera, värdera

*(Miniordlista på den samhälls-
vetenskapliga fakulteten 2009)*

Undersökning

- Miniordlistan innehåller 11 verb vars betydelse hämtats från *Norstedts Svensk ordbok*, *Natur & Kulturs svenska ordbok* samt SAOL
- Material: 20 hemtentamina från de senaste 5 terminerna på grundkursen i Socialt arbete

Referensmaterial

- Ord i Läroböcker, OrdiL
 - läromedelstexter för högstadiet (SO, NO & matematik)
 - ca 1 miljon ord
- Stockholm – Umeå-Corpora, SUC
 - allmänsvenskt skriftspråk
 - drygt 1,1 miljoner ord

Resultat

- Sökning på ordens alla former
- Av ordlistans 11 ord förekommer 7 i hemtentorna.
- Orden *argumentera*, *redogöra* *karaktärisera/karakterisera* och *värdera* utesluts ur undersökningen.

Carlund (2009): slutsatser

- De flesta av ordlistans ord förekommer redan i de första hemtentorna
- De aktuella orden har lägre frekvens i OrdiL (högstadietexter; gymnasiematerial saknas)
- Orden *problematisera* och *reflektera* saknas helt i OrdiL-korpusen
- Ordlistans ord förekommer förvisso i SUC men frekvensen är lägre och betydelserna ofta en annan

Några engelska akademiska ordlistor

- *The University Word List* (UWL) (Xue & Nation 1984). Sammanställning av 4 listor från 1970-talet.
- *The Academic Word List* (AWL) (Coxhead 1998, se även 2000, 2002).
- Jfr *Longman Dictionary of Contemporary English* (2009) – fokus på bl.a. akademisk vokabulär

Svenska akademiska ordlistor?

- Ingen direkt svensk – eller nordisk – motsvarighet till de engelska akademiska ordlistorna.
- Jfr Högskoleverkets lista över ord som är kopplade till högskolestudier, t.ex.
antagningsbesked, avancerad nivå, avhandling
- Jfr OrdiL-projektet (Lindberg & Johansson Kokkinakis 2008). Fokus på återkommande ord i läromedel för grundskolans senare år

Coxheads *Academic Word List* (AWL 2000)

- Bakgrund: det behövdes en ny engelsk akademisk ordlista, baserad på viktiga principer inom korpuslingvistik och kunskaper om ordinläring och ordundervisning (Coxhead 2002:74)
- Coxheads metod:
 - 1) sätta samman en akademisk korpus
 - 2) utifrån korpusen ta fram en akademisk ordlista

AWL-korpusen

Innehåll: 3,5 miljoner ord. 414 akademiska texter (vetenskapliga tidskrifter, kursböcker m.m.).
Mer än 400 olika författare.

Består av 4 delkorpusar (7 ämnen inom varje del):

- 1) **Arts:** education, history, linguistics, philosophy, politics, psychology, sociology
- 2) **Commerce:** accounting, economics, finance, industrial, relations, management, marketing, public policy
- 3) **Law:** constitutional, criminal, family and medico-, legal, international, pure commercial, quasi-commercial, rights and remedies
- 4) **Science:** biology, chemistry, computer science, geography, geology, mathematics, physics.

Från AWL-korpus till AWL

- Steg 1: Framtagning av AWL-kandidater, dvs. lexikala enheter med hög frekvens och stor spridning i AWL:s delkorpusar
- Steg 2: Jämförelse mellan AWL-kandidaterna och de ord som finns med i *A General Service list* (GSL, West 1953), dvs. en lista över frekventa (och för inlärare användbara) ord i allmän engelska.

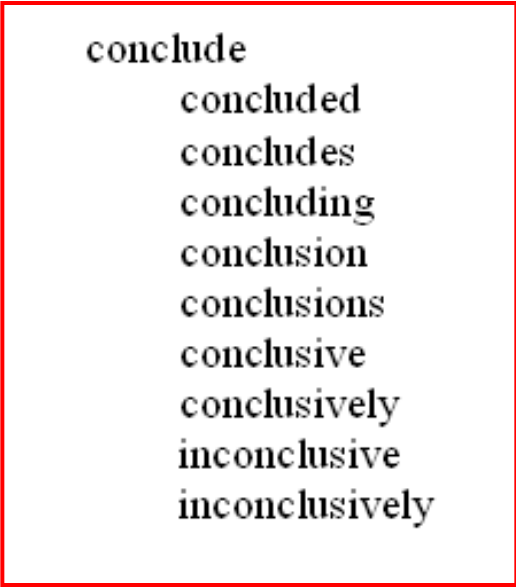
De 2000 ord som hade högst frekvens i allmänspråket (enligt GSL) togs bort bland AWL-kandidaterna, t.ex. *the, be, of, go, face, country, talk, work*.

- Resultat: 570 ordfamiljer(!).

Ordfamiljer

Stark tradition med bruk av ordfamiljer i ordlistor för engelskinlärare (se t.ex. West 1953, Xue & Nation 1984)

word family = a stem plus all closely related affixed forms
(Coxhead 2000:128)



conclude
concluded
concludes
concluding
conclusion
conclusions
conclusive
conclusively
inconclusive
inconclusively

AWL Sublist 1-10

- De 570 ordfamiljerna har delats upp i 10 undergrupper (sublists), beroende på ordfamiljernas frekvens och spridning i delkorpusarna.
- Sublist 1 innehåller de mest frekventa ordfamiljerna, sublist 2 de näst mest frekventa ordfamiljerna i korpusen, etc.
- Exempel på huvudord i Sublist 1:

analyse, approach, assume, benefit, concept, consist, constitute, contract, create, derive, distribute, establish, estimate, evident, factor, function, indicate, individual, interpret, involve, issue, legislate, major, method, occur, percent, period, policy, principle, proceed, process, require, research, respond, section, significant, similar, source, specific, structure, theory

Ur:
Academic
word list,
(Sublist 1)

analyse	reassessed	conceptually
analysed	reassessing	consist
analyser	reassessment	consisted
analysers	unassessed	consistency
analyses	<i>assume</i>	<i>consistent</i>
analysing	assumed	consistently
<i>analysis</i>	assumes	consisting
analyst	assuming	consists
analysts	assumption	inconsistencies
analytic	assumptions	inconsistency
analytical	<i>authority</i>	inconsistent
analytically	authoritative	constitute
analyze	authorities	constituencies
analyzed	<i>available</i>	constituency
analyzes	availability	constituent
analyzing	unavailable	constituents
<i>approach</i>	<i>benefit</i>	constituted
approachable	beneficial	constitutes
approached	beneficiary	constituting
approaches	beneficiaries	constitution
approaching	benefited	constitutions
unapproachable	benefiting	<i>constitutional</i>
<i>area</i>	benefits	constitutionally
areas	<i>concept</i>	constitutive
assess	conception	unconstitutional
assessable	concepts	<i>context</i>
assessed	conceptual	contexts
assesses	conceptualisation	contextual
assessing	conceptualise	contextualise
<i>assessment</i>	conceptualised	contextualised
assessments	conceptualises	contextualising
reassess	conceptualising	uncontextualised

Användning av AWL

Coxhead säger mycket litet om hur AWL är tänkt att användas.

"The AWL contains a large number of words of reasonable frequency and wide range and is an important tool for teaching and learning. The AWL is a useful example of corpus-based research leading directly to teaching and learning applications." (Coxhead 2002:79)

- AWL har haft stort genomslag inom andraspråksforskningen.
- Ordlistan har legat till grund för flera ordkunskapstest och många praktiska kursböcker med övningsuppgifter.
- Många handfasta råd till (in)lärare för hur man kan jobba med listan finns på AWL:s webbsida (School of Linguistics and Applied Language Studies, Victoria University, Wellington)

Några svagheter hos ordlistan

- Hyland & Tse (2007): AWL har olika grad av täckning inom olika ämnesområden.
- Hyland & Tse (2007): Betydelseangivelser saknas. Vissa ord har skilda betydelser inom olika ämnesområden.
- Paquot (2007): Vissa medlemmar i ordfamiljerna har mycket låg frekvens i autentiska texter.
- Nesi (2002): Endast skrivet akademiskt språk behandlas. Hur är det med det talade?

Mot en svensk *Akademisk Ordlista* (AO)

- Vårt mål: en fritt tillgänglig elektronisk akademisk ordlista som kan främja akademiska studier (men även användas vid gymnasiala studier)
- AO:s främsta målgrupper:
 - Studenter med annan språklig bakgrund
 - Studerande med mindre erfarenhet av akademiska studier
 - Lärare (med annat modersmål än svenska)

Översättning av AWL?

- AWL innehåller många ord vars svenska motsvarigheter tillhör akademisk prosa, t.ex. *analys, indikera, struktur*.
- Översättning kan gå snabbt → kostnadseffektivt
- Test med automatisk översättning (genom Lexins Engelsk-svenska ordbok, som är fritt tillgänglig). Resultat:

27% av uppslagsorden har bara en ekvivalent i ordboken, t.ex.

consist – innefatta
previous – föregående
evaluate – utvärdera
exclude – utesluta

Dessa ord är intressanta kandidater för en svensk akademisk ordlista

Översättning av AWL? (2)

Många problem p.g.a. homonymi och polysemi

approach → **verb** (intransitivt och transitivt)

1. närma sig, nalkas, rycka närmare, stunda, förestå
2. närma sig, nalkas
3. komma nära, närma sig, nå upp till, gå upp mot, likna
4. göra vissa trevare hos, ta kontakt med, inleda förhandlingar med
5. ta itu med, ge sig i kast med
6. anfälla

substantiv

1. närmande, annalkande, förestående ankomst
2. tillträde
3. tillvägagångssätt, taktik
4. inställning, syn, betraktelsesätt
5. första steg (försök), ansats

(förkortat ur Norstedts ord, <http://www.ord.se/>)

Rätt svenskt urval?

**Framtagning av en svensk motsvarighet till AWL
utifrån en svensk akademisk korpus**

En akademisk korpus

- Korpusen ska innehålla många, moderna, svenska akademiska texter (avhandlingar, tidskriftsartiklar, kurslitteratur etc.). Texterna ska vara skrivna av så många olika författare som möjligt
- Många olika universitetsämnen ska behandlas i texterna.
- Begränsning:
engelska – starkt undervisnings-/forskningspråk.
87 % av alla avhandlingar i Sverige skrivs på engelska
 - 94 % inom naturvetenskap
 - 65 % inom samhällsvetenskap
 - 37 % inom humaniora (Salö 2010)

Just nu: 3 akademiska delkorpusar (test)

1) Avhandlingskorpus

Metod: gräv där du står!

2) SweAc-korpus

Metod: med verktyg som mer slumpartat söker fram akademiska texter på Internet

3) SwePub-korpus

Metod: aktivt sökande i nationell databas över akademiska texter

Avhandlingskorpusen

- 9 språkvetenskapliga doktorsavhandlingar från 2000-talet. Utarbetade vid Inst. för svenska språket, GU.
- 9 författare, alla modersmålstalare (jfr Granger & Paquot 2009 som visat på skillnader mellan L1-talare och L2-talares akademiska språkbruk)
- Rensade texter: vi har tagit bort litteraturlistor, längre exempel, citat etc. (ca 15% av vissa texter)
- Storlek: ca 800 000 tokens
- Format: språkteknologiskt bearbetad. Finns i SketchEngine.

SweAc-korpusen

- Framtagen av Håkan Jansson genom WebBootCaT (som ingår i SketchEngine).
- Storlek: ca 20 miljoner ord
- Innehåll: 912 texter, antal författare? I huvudsak moderna, akademiska texter. Hela texter.
- Tidsödande manuell klassificering av texterna. Ämnen: främst ekonomi, utbildningsvetenskap och informatik.

SwePub-korpusen

- Insamling av texter som finns med i SwePub – en nationell och ämnesindeldad akademisk databas vid Kungliga biblioteket.
- Fokus på humaniora: **etnologi, historia, litteraturvetenskap, språkvetenskap, filosofi, religion**
- Ca 11 miljoner ord. 220 dokument, mer än 140 olika författare
- Finns i SketchEngine

Sammanfattning av korpusläget

- **Avhandlingskorpusen:** Mindre och ämnesmässigt begränsad (språkvetenskap). "Rensad" från icke-akademiskt språkbruk. För oss välkänt ämne och välkända författare.
- **SweAc-korpusen:** Relativt stor, gick mycket snabbt att sammanställa men har tagit lång tid att kartlägga. Vissa ämnen mycket överrepresenterade, andra saknas helt
- **SwePub-korpusen:** Mellanstor. Bara humaniora (etnologi, historia, litteraturvetenskap, språkvetenskap, filosofi, religion). Ungefär lika stora delkorpusar → de blir jämförbara med varandra. Förhållandevis lätt att göra kompletteringar med fler texter – om det finns sådana.

Olika korpusar kan komplettera varandra!

Jämförelse: akademiska ord – romankorpus

- Vi söker alltså efter ord som
 - har hög frekvens och god spridning i våra delkorpusar.
 - inte är högfrekventa i allmänspråket.
- Jämförelse (med hjälp av Keywords-funktion i Sketch Engine) mellan de akademiska orden och orden i en romankorpus

Keywords							
Corpus: user/GUX00/Academic							
Reference corpus: preloaded/swedishparole							
Page 1 Go Next >							
word	user/GUX00/Academic			preloaded/swedishparole			Score
	Freq	ARF	ARF/mill	Freq	ARF	ARF/mill	
-	410	150.2	1547.1	0	0.0	0.0	1548.1
emotiv	424	133.6	1375.6	0	0.0	0.0	1376.6
evaluering	380	121.1	1247.3	9	1.0	0.1	1164.8
informanterna	196	56.4	581.1	0	0.0	0.0	582.1
jfr	129	65.6	675.4	8	2.6	0.2	568.8
evaluerande	170	53.9	555.4	0	0.0	0.0	556.4
emotiva	175	54.8	564.3	1	1.0	0.1	527.5
evaluativa	171	46.0	474.0	0	0.0	0.0	475.0
semantiska	137	51.6	530.9	12	6.7	0.5	359.1
informanter	101	33.7	346.8	0	0.0	0.0	347.8
evaluativ	66	23.8	245.4	0	0.0	0.0	246.4
flerspråkiga	99	31.7	326.1	6	4.7	0.3	244.9

Några frekventa ord i materialen

angreppssätt, avgränsningar, figur, intervju, reliabilitet, metod, studie, innehållsförteckning, tillvägagångssätt, datainsamling, undersökningen, förförståelse, tillförlitlighet, implementering, generaliserbarhet, referensram, strukturering, frågeställning, poängtera, problemlösning

Många sammansättningar och sammansatta avledningar. Ord med germansk bakgrund. Hur fungerar det med ordfamiljer?

*tillförlitlighet (tillförlitlig, tillförlitligare, tillförlitligast, tillförlitligt, otillförlitlig, jfr *tillförlita)*

avgränsning (avgränsa, avgränsa sig, avgränsad, avgränsande, avgränsning, avgränsbar, oavgränsbar, oavgränsbarhet...)

- Många svårbegripliga och polysema ord.
- Uppgifter om ordens fraseologi saknas.

AO-korpusens slutgiltiga innehåll

- Hur stor?
- Hur många ämnesområden vill/kan vi ha?

Ordlistans slutgiltiga form och innehåll

- Anpassas till svenska språkets strukturella egenskaper (Ordfamiljernas roll?)
- Det elektroniska formatet bör utnyttjas!
- Ordlistan ska kunna brukas i fler användningssituationer → fler informationskategorier krävs.

Viktigare informationstyper i inlärningsordböcker

Tarp (2008) mfl.: ordböcker är bruksföremål som måste anpassas till

1) tänkta användare

2) tänkta användningssituationer

- Reception (förståelse) (av tal/text): betydelse
- Produktion (av tal/text):
 - ortografi
 - böjning
 - ordbildning
 - uttryck (kollokationer m.m.)
 - synonymer, antonymer etc.
- Ordlistan kan testas i användarundersökningar, t.ex. inom Språkhandledningens verksamhet

Koppling korpus - ordlista

- Länk mellan uppslagsord och korpus?
 - användarna kan se hur orden brukar användas i olika kontexter
 - exempel fungerar som komplement till betydelseangivelserna

Ex. *metod:*

kvalitativ metod

kvantitativ metod

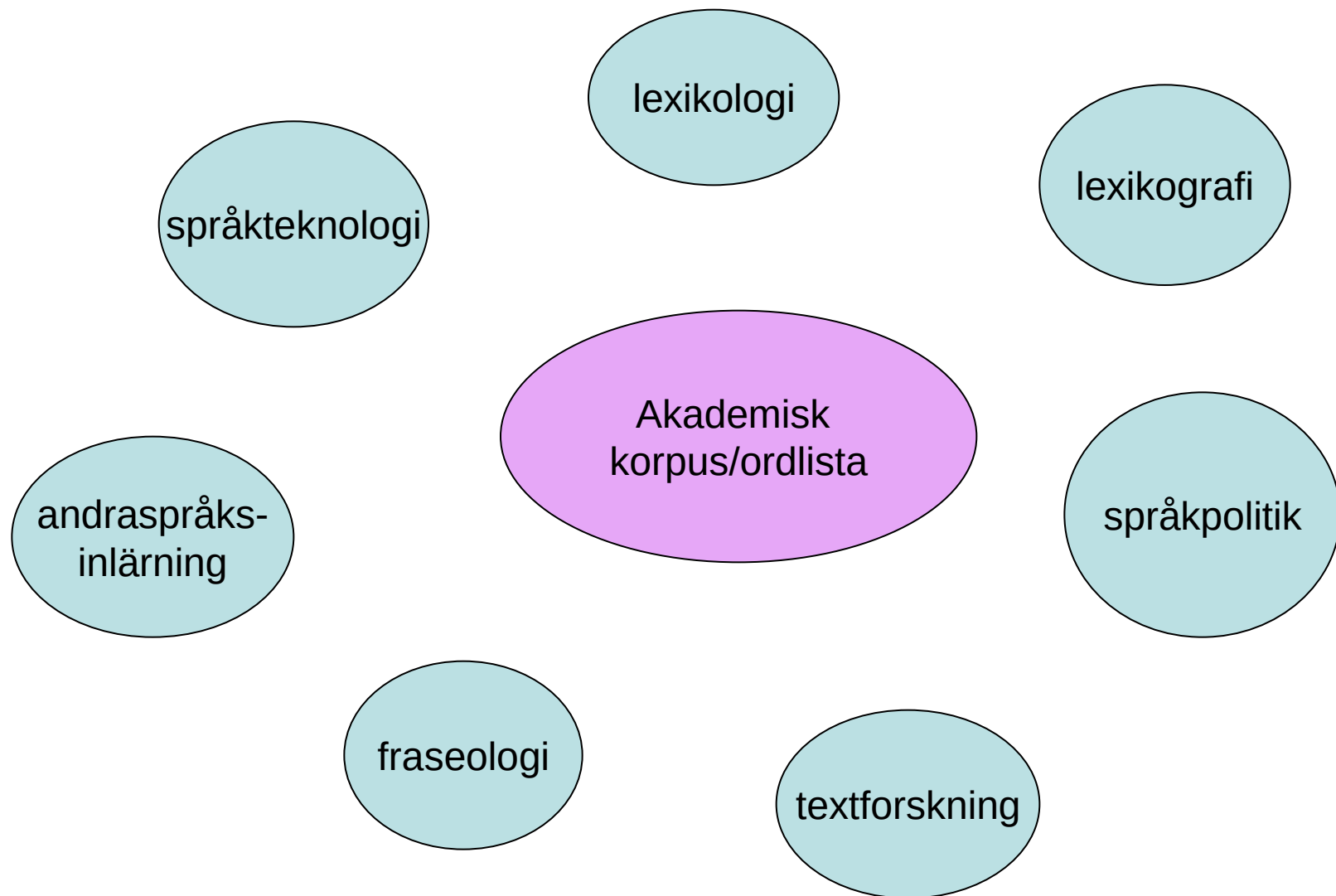
statistiska metoder

forskare har utvecklat en ny metod för att kunna omvandla giftigt avfall till dricksvatten

- Frekvensangivelser?

Många frågor återstår!

AO - ett tvärvetenskapligt arbete



Nordiskt samarbete

- Nätverk för framtida akademiska ordlistor för flera nordiska språk
- Huvudpartners (hitintills):
 - Institutionen för svenska språket, Göteborgs universitet
 - Textlaboratoriet, Universitetet i Oslo
 - Center for Internationalisering og Parallelsproglighed (CIP), Københavns universitet
 - Institutionen för språkdidaktik, Stockholms universitet
 - Språk- och Litteraturcentrum, Lunds universitet
- Workshop, Göteborg, maj 2011; gemensamma konferensbidrag
- Det svenska projektet delfinansieras av Göteborgs universitets språkår.