



GÖTEBORGS UNIVERSITET
Institutionen för svenska språket

Lexikaliska institutet
Institutet för svenska som
andraspråk
Språkbanken

A och O om akademiska ord

Om framtagning av en svensk akademisk ordlista

Emma Sköldberg & Sofie Johansson Kokkinakis
11:e konferensen om lexikografi i Norden,
Lund, maj 2011

Innehåll i föredraget

- Bakgrund
- Kort om A. Coxheads *Academic Word List* (1998)
- Mot en svensk akademisk ordlista
- Nordiskt nätverk kring ämnet
Språkbruk i nordiska akademiska miljöer

Det akademiska språket – en utmaning för många

- Det akademiska ordförrådet, dvs. ord som förekommer i universitetstexter oavsett vilket ämne som behandlas, orsakar problem för akademiska nybörjare – oavsett modersmål!
- Exempel på svåra verb som lyfts fram i samband med Språkhandledning (inom Göteborgs universitet):

analysera, argumentera, beskriva, diskutera, referera, reflektera, karakterisera, motivera, problematisera, redogöra, värdera
(Carlund & Winlund 2011)

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

Kort om Coxheads *Academic Word List* (AWL, 1998)

- 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 (tokens). 414 akademiska texter (vetenskapliga tidskrifter, kursböcker m.m.). Mer än 400 olika författare.

Består av 4 delkorpusar (å 875 000 tokens). 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.

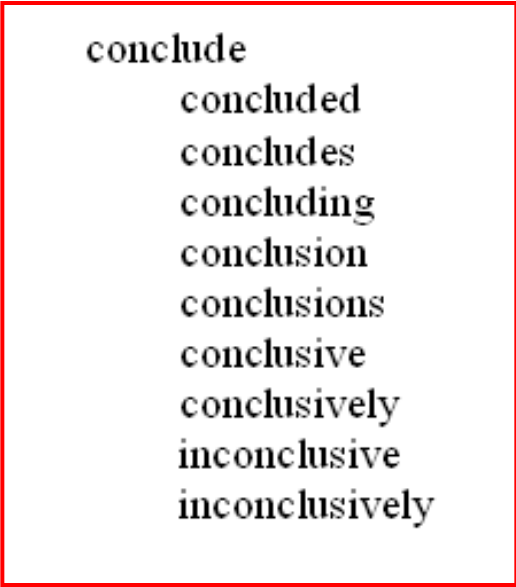
De 2000 ord som hade högst frekvens i allmänspråket (enligt GSL) togs bort bland AWL-kandidaterna, t.ex. *the, be, of, go, face, country, talk, work*.

- 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.
- Undergrupp 1 innehåller de mest frekventa ordfamiljerna, undergrupp 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 stor 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 på AWL:s webbsida (School of Linguistics and Applied Language Studies, Victoria University, Wellington)

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
Problem p.g.a. homonymi och polysemi!
- Ordfamiljer?
- Rätt svenskt urval?

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/>)

**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.)
- Många olika ämnen ska behandlas i texterna.
Jämför Göteborgs universitets fakulteter:
Handelshögskolan, Humaniora, IT-fakulteten, Konst, Naturvetenskap, Sahlgrenska akademien, Samhällsvetenskap, Utbildningsvetenskap
- Begränsning:
Engelska – starkt undervisnings-/forskningsspråk

Just nu: två akademiska delkorpusar (test)

- **Avhandlingskorpusen**

Innehåll: 9 språkvetenskapliga doktorsavhandlingar från 2000-talet. Rensade texter: vi har tagit bort litteraturlistor, längre exempel, citat etc.

Storlek: ca 800 000 tokens

Format: tokeniserad, ordklasstaggad, lemmatiserad. Finns i SketchEngine.

- **SweAc-korpusen**

Framtagen av Håkan Jansson genom WebBootCaT.

Innehåll: olika typer av i huvudsak moderna, akademiska texter. Olika ämnen. Hela texter.

Storlek: ca 20 miljoner tokens

Format: i SketchEngine. Ej lemmatiserad.

Olika delkorpusar kan komplettera varandra!

Jämförelse: akademiska ord – Parole

- Jämförelse (med hjälp av Keywords-funktion i Sketch Engine) mellan AO-orden och Parole-korpusen (ca 19 miljoner tokens, ordklasstaggad)

Keywords							
Corpus: user/GUX00/Academic							
Reference corpus: preloaded/swedishparole							
Page 1 Go Next >							
	<i>user/GUX00/Academic</i>			<i>preloaded/swedishparole</i>			
word	Freq	ARF	ARF/mill	Freq	ARF	ARF/mill	Score
-	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

Jämförelse: akademiska ord – Parole

Keywords

Corpus: **user/GUX00/Academic**

Reference corpus: **preloaded/swedishparole**

Page [Next >](#)

	<i>user/GUX00/Academic</i>			<i>preloaded/swedishparole</i>			
word	Freq	ARF	ARF/mill	Freq	ARF	ARF/mill	Score
-	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(former) i SweAc-korpusen

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...)

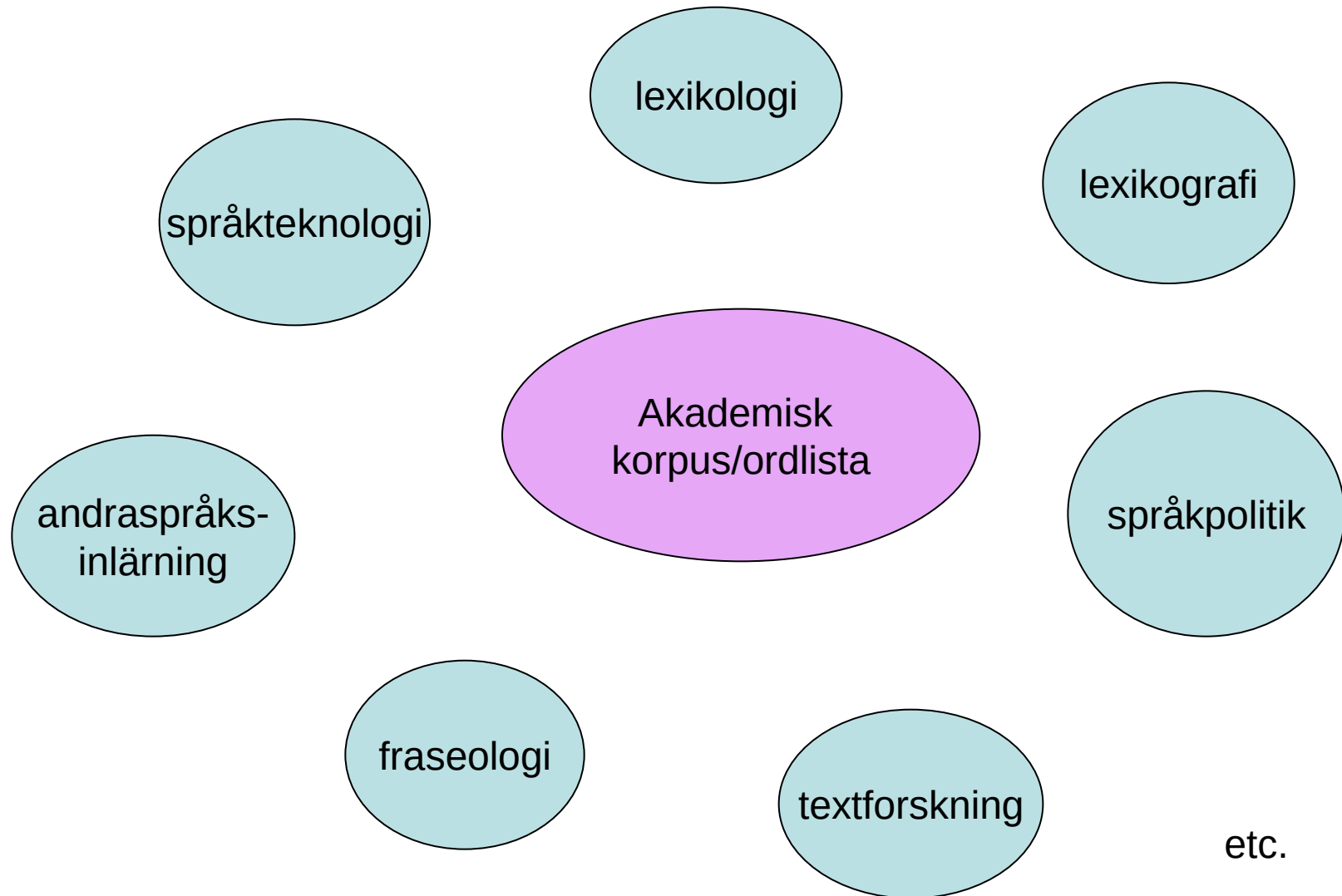
- Många svårbegripliga och polysema ord.
- Uppgifter om ordens fraseologi saknas.

AO-korpusens/ordlistans innehåll & utformning

- Korpus:
 - Hur många tokens?
 - Hur många ämnesområden?
 - Hur rensade ska texterna vara?
 - Vilken allmänspråklig jämförelsekorpus ska användas?
- Ordlistan:
 - Anpassning till svenska språkets strukturella egenskaper. (Ordfamiljernas roll?)
 - Ska kunna brukas i fler användningssituationer → fler informationskategorier krävs.
 - Ska testas fram i användarundersökningar
- Korpus/ordlista:
 - Länk mellan uppslagsord och korpus.
 - även frekvensangivelser

Många frågor återstår!

Avslutningsvis: AO - ett tvärvetenskapligt arbete



Nordiskt samarbete

- Det har bildats ett nätverk för att det i framtiden ska utarbetas akademiska ordlistor för flera nordiska språk
- Huvudpartners (hitintills):
 - Institutionen för svenska språket, Göteborgs universitet
 - Textlaboratoriet, Universitetet i Oslo
 - Center for Internationalisering og Parallelsproglighed (CIP), Köpenhamns universitet
 - Institutionen för språkdidaktik, Stockholms universitet
 - Språk- och Litteraturcentrum, Lunds universitet
- Workshop, Göteborg, maj 2011.