

# Language resources and tools

Markus Forsberg

Språkbanken  
University of Gothenburg

GF Summer School 2015

# Today's talk

- Language resources and tools at Språkbanken (the Swedish language bank).
- A quick introduction to Corpus Workbench.
- Demonstration of some of Språkbanken's tools.

## A couple of years ago: Legacy systems at Språkbanken

- Språkbanken has been around since 1975.
- Service unit for linguists  $\Rightarrow$  LT research unit.
- The old way to fly: a language resource = database + interface
- The structure of the LR was largely irrelevant (as long as everything looked nice in the interface).
- Made linguists (somewhat) happy, and LT researchers unhappy.

## Legacy systems: konk and ORDAT, ...

[Språk Pedagoger](#)
[Historik](#)
[Bokrecensioner](#)
[Ärkebiskop](#)
[Hjälp](#)

[http://sprakbanken.gu.se/kon/](#)

[Mest besökta](#)
[Language Log](#)
[Lingvistbloggen](#)
[Tg3b bmp \(BMP fil...\)...](#)
[The LINGUIST U...](#)
[Institutionen för s...](#)

**Konkordanser**

Årsmåls början: **GP 01**    Konsekvens i tecken: **120 tecken**    Konsekvensklass: **50%-50%**    Antal träffar: **20 träffar**

**Språk BANKEN**

Grund på format: **standard**    Söktaxa: **svenska**

Sök i:
 

- ☐ lexikonbas
- ☐ frekvens
- ☐ frekvens tabell

 Sök: **Sök**    Återställ

[hitta](#)
[validera](#)
[förändra](#)
[spara för svenska språkbanken](#)
[SÄKER](#)
[in förändra](#)
[spara språk](#)

**Nästa (21-40)**    [Teknisk: 120993 (visar 1/20)] [Söktaxa: svenska] [Material: p01]

som i fjol belönades med Whitbread-priset. Nysa utkommen på **svenska**.  
 var utset till särskilt utnämnd och dessutom utset till **svenska**.  
 i Sverige. I dag finns det 73 000 norska bönder och 80 000 **svenska**.  
 er, shoppar och pratar. Är aktuell med skivan Folkmusik på **svenska**.  
 n frägar mig vilka språk jag kan, säger jag snaringsenligt **svenska** !  
 8) Ett land i mörker (1973) Årslänet arvar den boken kom på **svenska**  
 test 180 miljoner. An så länge finns bara fyra spel klara. **svenska** A och B samt matematik A och B. En rad kommuner, stu  
 Gordon SW Vindkraft A Bernth Alexandersson Västasvand **Svenska** AB  
 i Norge öka försäljningen av DBS-cykeln i Skandinavien. DBS **Svenska** AB bildades och Terasse gick nu DBS-cykeln till S  
 i AB Dayton Bridgestone Firestone Sws. AB Delica Goodyear **Svenska** AB Delta Goodyear Svenska AB Dunlop Goodyear Svens  
 resterna Sws. AB Delica Goodyear Svenska AB Delta Goodyear Svenska AB Dunlop Goodyear Svenska AB Eira Goodyear Svenska  
 Goodyear Svenska AB Delta Goodyear Svenska AB Dunlop Goodyear Svenska AB Eira Goodyear Svenska AB Europa Bridgestone Fire  
 odyear Svenska AB Dunlop Goodyear Svenska AB Eira Goodyear Svenska AB Europa Bridgestone Firestone Sws. AB Falken Fra  
 one Sws. AB Formula Pirelli Scandinavia AB Palla Goodyear **Svenska** AB General Continental Dack Sverige AB Gislaved Co  
 AB Gislaved Continental Dack Sverige AB Goodyear Goodyear Svenska AB Hank  
 year Goodyear Svenska AB Hankook Amring AB India Goodyear **Svenska** AB India Goodyear Svenska AB Kally  
 ankook Amring AB India Goodyear Svenska AB Kally Goodyear Svenska AB Kleber  
 importen AB Kormoran Michelin Gummitrigger AB Lw Goodyear Svenska AB Mabal  
 cormorans Radioanstalt. Han avgick 1952 till ett känd tjänst i **Svenska** AB Philip  
 irelli Scandinavia AB Pirelli Vartax AB Pneuamant Goodyear **Svenska** AB Riken

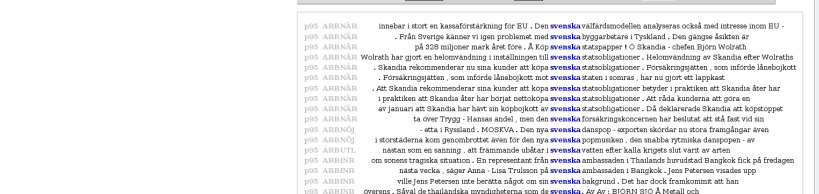
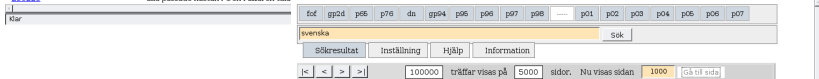
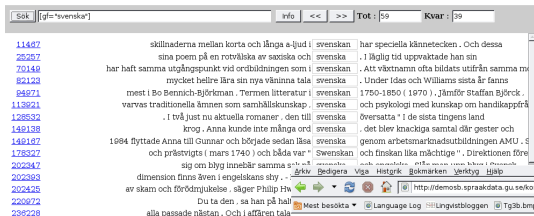
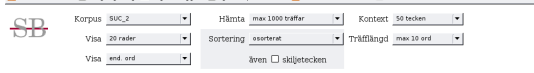
[http://sprakbanken.gu.se/ORDAT/search.php?m71232453145447](#)

153    3814    sidor    **Inställning**    **Anteckn**    **SÖK**

	1	2	3	4	5	6	7	8
När	18	19	20	21	22	23	24	25
	26	27	28	29	30	31	32	33

1923 rav bättre anpassade bestämmelser. 1923 års **svenska** motorkrönika rymmer visserligen utöver den nä  
1923 r på Edsvikens främst, och de därvid samt **svenska** hastighetrekorden för såväl bil som motorcycl  
1923 r sig svårt att ge ett allmänt omdöme om den **svenska** penningmarknaden år 1923. Från viss synpunkt  
1923 g, i samband härmed bör kanske nämnas, att **svenska** folkets sammanlagda tillgodohavande å sparkä  
1923 ller mindre delar av sina fonder, så nämnas **Svenska** Handelsbanken (105, 0 mill. kr. avskr. tå  
1923 lda bank) (19, 0 resp. 17, 7 mill. ) samt **Svenska** lantmännens bank (18, 7 resp. 16, 1 mill.  
1923 För att rädda banken inträdde då som känt år **svenska** staten såsom huvuddelägare i banken. Karaktä  
1923 r riksbankens räntepolitik betingats av den **svenska** kronans ställning till dollarn och därmed till  
1923 utländska valutorerna i allmänhet har dock den **svenska** kronan ej blivit hävdats en ställning, utan n  
1923 lt under året icke borta i förhållande till **svenska** kronor utan även till pund. Den fruktan för  
1923 slifför på gälla som norm. Jordbruket. Det **svenska** jordbruket har under de senaste åren kämpat m  
1923 mport av vete under 1923 (se nedan) år det **svenska** jordbrukets läge f. n. bekymmersamt. För at  
1923 t sättningen av åtminstone större delen av det **svenska** vete t de större kvarnarna i september före  
1923 bet betalningsbalans, vilken under den moderna **svenska** industriens grundläggningssperiod i slutet a  
1923 913. Den är så mycket egendomligare som den **svenska** textilindustri i allmänhet synes ha god s  
1923 et givetvis ännu återstår mycket, innan den **svenska** industri en som årsarbete för full drift.  
1923 egeringen, Universitetskonstnärer och den **svenska** minoriteten uppburna s. k. linjeläsningsförel  
1923 , som om det antagits varit något att trygga **svenska** språkets framtida ställning vid universitetet  
1923 av som anknäpningen. Det faktum, att den **svenska** skidspårsrörelsen till en del under universitetet

# ... Parole/SUC and Konkplus, ...



# ...ITG and Litteraturbanken, et cetera, moreover

Modulfönster

- Lära mig mer
- Övningar
- Korpus
  - Korpusökning
  - Rensa söktytan
- Mina korpusurval
- Mina konkorporationer
- Mina annoteringar
- Utskrifter
- Status
- Hjälp

Textvisning | Konkordanser | Annoteringseditor | Inställningar

Unval | Korpusökning | Frekvenser

Ord | Lemma | Grammatisk beskrivning | Intervall

svenska

Sök | Avbryt

☐ Visa endast de fem första resultaten

Beläggstext	Vänsterkontext	Sökt enhet	Tagg	Högerkontext	Ege...
s5-009.6	...t ja kan inte prata bra på svenska	nn			
s5-196.3	...och jag läser författande svenska	nn			
s5-226.6	...Ste på PÅSH KURS LÄRA SVENSKA	pm	OCH SJUKA SH EFTÄR GÅ		
s5-188.1	...jag vet att du är duktig på svenska	nn	och du gått många kurser		
s5-211.8	...många många människor (svenska	nn	människor ) också blev		
s5-107.5	...stad och du ska lära dig svenska	nn	också		
s5-216.5	...ligt och du måste lära dig svenska	nn	men det är inte problem		
s5-108.3	...ka ett jobb måste du vet svenska	nn			
s5-005.3	...jag hoppas att du kan svenska	nn			
s5-154.9	...och han är jätteduktig på svenska	nn	och matematik		

jag vet att du är duktig på svenska och du gått många kurser

Sök i verkstext | Litteraturbanken - Mozilla Firefox

Arkiv | Redigera | Visa | Historik | Bokmärken | Verktyg | Sökp

Sök i verkstext | Litteraturbanken

http://litteraturbanken.se/sök/forfattare=almqvist+almsantal=206+sortering=werkstext+svenska

SALDO de v m Welcome to CLT... | http://it.gu.se/... | Institutionen för... | Lingsstebogen | Language Log | The LINGUIST L... | Vapir E!

Litteraturbanken

Home > sök

Sök i texten

svenska <=> >=>

Ord eller frasa (text)

Visa 20 träffar per sida, eller [anpassa träffar](#)

1	Almqvist, C.J.L.	Anstade Håttor	187	...som på förtäna - så låter min text
2	Almqvist, C.J.L. <th>Anstade Håttor</th> <td>1901</td> <td>...t, s 23-24, en helt ny text, vilket ger en bildning an...</td>	Anstade Håttor	1901	...t, s 23-24, en helt ny text, vilket ger en bildning an...
3	Almqvist, C.J.L. <th>Anstade Håttor</th> <td>1901</td> <td>...sida också utvald. Inom text efter svakt, texten är så g...s</td>	Anstade Håttor	1901	...sida också utvald. Inom text efter svakt, texten är så g...s
4	Thordis, Thomas <th>Att följa svensk</th> <td>1939</td> <td>...sida begrepp skulle här finna text till lågt före betänksl...</td>	Att följa svensk	1939	...sida begrepp skulle här finna text till lågt före betänksl...
5	Asenroth <th>Bullader Sveri</th> <td>19</td> <td>...rydner är språkvetenskapliga frin text till men det karaktärist...</td>	Bullader Sveri	19	...rydner är språkvetenskapliga frin text till men det karaktärist...
6	Asenroth <th>Bullader Sveri</th> <td>15</td> <td>...språkvetenskapliga frin text till text men det karaktärist...</td>	Bullader Sveri	15	...språkvetenskapliga frin text till text men det karaktärist...
7	Asenroth <th>Bullader Sveri</th> <td>16</td> <td>...det lin ut sommarbetsvetenskap - i stället förfrattas men d...</td>	Bullader Sveri	16	...det lin ut sommarbetsvetenskap - i stället förfrattas men d...
8	Asenroth <th>Bullader Sveri</th> <td>16</td> <td>...som på en "bollad" grundad på text - många kortare. När som t...</td>	Bullader Sveri	16	...som på en "bollad" grundad på text - många kortare. När som t...
9	Asenroth <th>Bullader Sveri</th> <td>17</td> <td>...internas om En prokurren text till en skänkes kan t.m. g...</td>	Bullader Sveri	17	...internas om En prokurren text till en skänkes kan t.m. g...
10	Asenroth <th>Bullader Sveri</th> <td>20</td> <td>...angreter att en vis text är text - och iakt. Men för den tröat...</td>	Bullader Sveri	20	...angreter att en vis text är text - och iakt. Men för den tröat...
11	Asenroth <th>Bullader Sveri</th> <td>21</td> <td>TEXT- OCH MEDIOGROUNDS</td>	Bullader Sveri	21	TEXT- OCH MEDIOGROUNDS
12	Asenroth <th>Bullader Sveri</th> <td>198</td> <td>...Uppskickad on J. E. Weber ( text ) och O. R. Sjöberg ( melod ) på...</td>	Bullader Sveri	198	...Uppskickad on J. E. Weber ( text ) och O. R. Sjöberg ( melod ) på...

# ... Dalin and Old Swedish, ...

DALINS ORDBOK (Språkbanken) - Mozilla Firefox

göw Badgera Vga Historik Bökmärken Verktyg Själp

... DALINS ORDBOK (Språkbanken) ...

... språkbanken.gu.se/dalin ...

... DALIN dev ...

**DALINS ORDBOK**

Ordbok öfver svenska språket af A. F. Dalin. (1850-1855)

UPPSLAGSORD

korp

MAX ANTAL TRÄFFAR 10 50K

[Alt om Dalins ordbok](#)

**KORP**

m. 2. (lat. Carvus)

1) Fågel af Sparfågellarnes ordning. 2 fot lång, färgen svart, skimrande i grönt och purpur. Corvus Corax. Svart som en k. (Talesätt) Stjälva som en k., gluskt. (Ordspr.) Den ena k-en hackar ej ut ögat på den andra, tjuftar, roffare och skälmar hålla ihop. -

2) (fig. fam.) Roffare, prejar. — Ss. *K-bo*, *-fjäder*, *-låte*, *-penna*, *-unge*.

FORNSVENSK LEXIKALISK DATABAS - Mozilla Firefox

göw Badgera Vga Historik Bökmärken Verktyg Själp

... FORNSVENSK LEXIKALISK DATABAS ...

... språkbanken.gu.se/fvdbas ...

... DALIN dev ...

**FORNSVENSK LEXIKALISK DATABAS**

UPPSLAGSORD

ramn

MAX ANTAL TRÄFFAR 10 50K

[Alt om Fornsvensk lexikalisk databas](#)

**ramn** *oldswed* ( rampn SD 3: 84 (1312); *Bil* 404; MB 1: 315, 359; Gn 35, 307; Al 890, 894; en MB 1: 171; -a *Al* 2784. ramfn *Bil* 577. rafn. rafnin ib 405; -a *MP* 1: 268; *Bil* 851; -om *MP* 1: 268. raffn *Di* 8. ram *Ea* (Tung) 35; *RK* 3: (sista forts.) 4878. -ari, m. [isl. hrafn] korp. liggia fore thund oc rampn MB 1: 315. therea ben lagho vj dagha for hundom ok ramnum *Bil* 449. *MP* 1: 268. plägdho altidh ther the lagho rafna saman koma ok öpa *Bil* 851. swarter som en ram *Ea* (Tung) 35. thz ar siälsynth at se hwtan rampn *Gl* 35. swa gymnas soldenar godz som rampn atel ib 307; *Bil* 404, 405, 577 MB 1: 171, 359; Al 890, 894, 2784. *Di* 8; *RK* 3: (sista forts.) 4878. *HSBil* 24, 85 (1516). slitz och rifs af heluitis hundom oc rafnom *MP* 1: 268. - ss tillnamn. sigilla ... domini tulonis abbae son et akae rampn fratrurn meorum SD 3: 84 (1312). — *ramna as* ( rampna aas), n. as för korpar. han hafwer än quart naghath rampna aas

**ramn** *oldswed* *swedl* ( rampn: -a *Prosadikt* (Sju vise m) 169. rafn. -om ib (Bart) 65. rapn ib (Karl M D 3) 337). m. korp. nw sa olofernes här mykian som hundrada lotom war swartare än rapn *Prosadikt* (Karl M D 3) 337. -jfr natta ramm.

# ...SALDO and SweFN, et cetera, et cetera

corp...1 - Mozilla Firefox

golv Badgera Vga Historik Bokenmärken Verktyg Bjäp

corp-1

gprsaalbanken.gu.se/salido-es/ld/graph/corp\_1


WELCOME TO COT ... http://ot.gu.se/... Institutionen för a... Kyras II Ungaestblogen Language Log The LINGUIST U... Keep It!

*SALDO*

documentation (pdf)

**corp**

corp (tan)



korpus

http://gprsaalbanken.gu.se/salido-es/ld/graph/corp\_1

Development version of SweFN | Språkbanken - Mozilla Firefox

golv Badgera Vga Historik Bokenmärken Verktyg Bjäp

Development version of SweFN |

http://gprsaalbanken.gu.se/langresearch/swefn/development-version

SALDO dev WELCOME TO COT ... http://ot.gu.se/... Institutionen för a... Kyras II Ungaestblogen Language Log The LINGUIST U... Keep It!

History

Error report

Dependency analysis

Dalia

PM-SBLEX

SALDO

Sweasax

### Abandonment

ram	Abandonment
ömsön	Öpp
semantisk typ	Realisreal_act
körselement	Agent, Theme
periferielement	Depictive, Duration, Explanation, Manner, Place, Time
exempel	<p>Reskittningen på Skandia trögades att överge bilen på grund av stora skador som gjort det omöjligt att navigera.</p> <p>Alla mina vänner kommer att överge mig om de inte får tag på mig.</p> <p>Han glömde inte bara sina glansigen i Helsingfors 1909 utan också en visselpipa i Sverige.</p> <p>Tjänstemäns kan glömma handlingar i en portfölj i taxi eller på tåg.</p> <p>Läkare glömde kvar kompress.</p> <p>Jag har följt bilen på verkstad, säger hon och står st med armarna.</p> <p>En sje dagar genom pojke följades kvar ensan i en varm bil i Bäckebol på Hisingen i går kväll.</p> <p>På denna punkt anser Kurt Johansson att viljarna lämnas i sticket i årets valrörelse.</p>
ans	
sns-exempel	
saldo	vb: glömma.2 överge.1 sv: lämna.2 överge.1 vbw: glömma kvar.1 lämna i sticket.1 lämna kvar.1
saldo (aya)	

20080

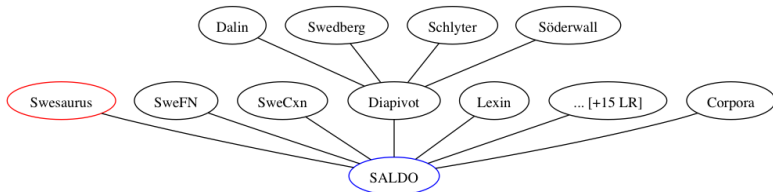


What to do?



## Changing the situation

- Put the resources in the center, not the interfaces (**downloadable** resources in a common format, so far IPR permits)
- Centralize and think in term of research infrastructure (= technological solutions that try to enable as much new research as possible)
  - **Korp** – corpora infrastructure
  - **Karp** – lexical infrastructure
- Link all the resources to a **pivot resource** (GF speak: a lexical abstract syntax), SALDO; that is, create a large LT resource network (a macro-resource).



# SALDO

- **SALDO** is a full-scale ( $\sim 130k$  word senses, 2M word forms) lexical-semantic resource for Swedish with semantic relations between all word senses (including MWE).
- Available under an open license: CC-BY.
- SALDO is a directed graph with so called **primary** and **secondary** relations.
- The fundamental unit is the **word sense** (the first version of SALDO did only contain word senses).
- All word senses is given one or more formal descriptions, referred to as **lemgrams** (lemgram=paradigm+lemma  $\rightarrow$  inflection table)

# SALDO “PIDs”

- SALDO has id's for:
  - senses (`grad..1`)
  - lemmagrams (`grad..nn.1`)
  - parts of speech (`nn`)
  - paradigms (`nn_3u_film`)
- the id's are designed to be
  - **unique** (no other id's should be necessary, e.g., database keys)
  - **atomic** (no built-in assumptions about sense–subsense relationships, etc.)
  - **usable in Semantic Web formalisms** (RDF, OWL): id's are well-formed XML names
  - **human-readable** (makes resources easier to work with)


## Details about SALDO

- All (except a few) have a obligatory **primary descriptor**, and an optional set of **secondary descriptors**.
- 41 senses lack primary descriptor, joined together with an artificial zero-sense *PRIM..1* (E.g., färg 'color', 'rak' 'straight', tänka 'think', ...)
- A primary descriptor should be **semantically close** and more **central**: more frequent, stylistically more neutral, morphologically simpler, and more.
- The secondary descriptors help discriminate the sense (no special criteria).

# SALDO example: bota 'cure'

Saldo (1)					
Sense	Lemgram	POS	Primary	Children (primary)	Children (secondary)
<b>bota</b> 	<b>bota (verb)</b>  (77767)	verb	<b>frisk</b>	<ul style="list-style-type: none"><li>• <b>behandla<sup>2</sup></b> -</li><li>• <b>behandlingsbar</b> möjlig</li><li>• <b>botande</b> -</li><li>• <b>botlig</b> möjlig</li><li>• <b>doktor</b> -</li><li>• <b>elektroterapi</b> elektricitet</li><li>• <b>hela</b> -</li><li>• <b>kramplösande</b> <b>kramp</b></li><li>• <b>kurativ</b> -</li><li>• <b>kurera</b> -</li><li>• <b>läka<sup>2</sup></b> -</li><li>• <b>läkare</b> -</li><li>• <b>medicin</b> -</li><li>• <b>obotad</b> inte</li><li>• <b>obotlig</b> omöjlig</li><li>• <b>operera</b> -</li><li>• <b>psykiatri</b> psykisk</li><li>• <b>sjukhus</b> -</li><li>• <b>svårbotad</b> svår</li></ul>	<ul style="list-style-type: none"><li>• <b>kvacka</b> fuska</li><li>• <b>medicin<sup>2</sup></b> medel</li></ul>
N corpora hits				secondary descriptors	
<div>Hide</div>					

## Saldo morphology (1)

Lemgram	POS	Paradigm	Morphology
<b>bota (verb)</b>  (77367)	verb	vb_1a_hitta	<div> bota ... </div> <div> pres ind aktiv pres ind s-form pret ind aktiv pret ind s-form </div> <div> <b>botar</b> <b>botas</b> <b>botade</b> <b>botades</b> </div>
			Inflectional forms

# Linking backwards in time (I)

- Linking SALDO and Dalin (19th century Swedish) is relatively straightforward.
- The vocabulary differences are mainly in the compounds, e.g.:
  - *bäfverhund* 'dog used for beaver hunt'
  - *bäfverhund* → **modernize** → *bäverhund* → **compound analysis** → *bäver..nn.1+hund..nn.1*

## Linking backwards in time (II)

- Linking the Old Swedish to SALDO is more challenging.  
An illustrative example:
  - **bakvapi** *fatal accident resulting from a sword being struck backwards without the striker looking in that direction beforehand*
- Link to what? Accident? Sword? Both? Others?



# Korp pipeline: the annotation lab

Svenska English



## KORP Annotation Laboratory

Language of analysis: Swedish

Load example: Drama

Åtta sidor

Talbanken

Läsbart

Exempelkorpus

Editor

Upload

1 Alla deltagare på GFs sommarskola älskar pizza.

☒ Lexical analysis ☒ Compound analysis ☒ Dependency analysis

Show advanced settings

Run!

<paragraph> [Show XML]

<sentence id="8f7-83b"> [Show XML]



word	msd	lemma	lex	saldo	prefix	suffix	ref	dephead	deprel
Alla	DT. UTR+NEU. PLU. IND+DEF	all	all..pn.1	all..1			1	2	DT
deltagare	NN. UTR. PLU. IND. NOM	deltagare	deltagare..nn.1	deltagare..1			2	6	SS
på	PP	på	på..pp.1	på..1			3	2	ET
GFs	PM. GEN						4	5	DT
sommarskola	NN. UTR. SIN. IND. NOM				sommar..nn.1	skola..nn.1	5	3	PA
älskar	VB. PRS. AKT	älska	älska..vb.2, älska..vb.1	älska..2, älska..1			6		ROOT
pizza	NN. UTR. SIN. IND. NOM	pizza	pizza..nn.1	pizza..1			7	6	OO
.	MAD						8	6	IP

</sentence>

</paragraph>

# Korp: the corpora infrastructure

Modern | Parallel | Old Swedish | Litteraturbanken | Kubhist | Old texts | Fler ▾

Cite Korp Svenska | English  



122 of 198 corpora selected — 1.64G of 9.10G tokens

Search history 




Simple Extended Advanced Compare 

bota (verb)

Search ▾

also as ☐ initial part ☐ final part and ☐ case-insensitive

Relaterade ord (SWE-FN)  
bota, hela, läka  
botande, helande, botlig...

KWIC: hits per page: 25  sort within corpora: not sorted  Statistics: compile based on: word  ☐ Show word picture

KWIC Statistics Word picture

Results: 13,192

« 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... » Go to page  of 528 [Show context](#)

ÄRO Underrättelser 2012

Till slut undrar bladet om det möjligen finns medicin som kan **bota** den rondellfeber som grasserar i Kimito förrän den sprider sig till Dragsfjärd.

Såsom Allianssi konstaterade i sin undersökning: Det blir alltid dyrare att **bota** än att förebygga.

– Ifall det har bildats metastaser är prognosen dålig och cancer är svår att **bota**.

Det handlar om att hitta mediciner som **bota** eller förhindrar sjukdomar, eller som aktiverar kroppens egen ämnesproduktion.

Kimitoöns biblioteksbesökare kan **bota** eventuella höstdpressioner samtidigt som de sitter skönt och läser.

ÄRO Underrättelser 2013

Enligt dem går cancer att **bota** och han ska bli helt återställd.

Wallgren en kolumn där han kallade sig själv för en rasist som kan bli bättre, men aldrig **bota**.

ASTRA 1960–1979 (does not support extended context)

Men så, att man lär sig, i vilken förvånansvärd grad sexualproblemen kan **bota**.

## Corpus

Äro Underrättelser 2012

### Text attributes

date: 2012-10-23

### Word attributes

part-of-speech: verb

baseform:

bota


lemgram:

bota (verb)

sense:

bota

# Korp: word picture

122 of 198 corpora selected — 1.64G of 9.10G tokens

Search history

SimpleExtendedAdvancedCompare 2

bota (verb)Search

also as ☐ initial part ☐ final part and ☐ case-insensitive

Relaterade ord (SWE-FN)  
bota, hela, läka  
botande, helande, botlig...

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture

KWICStatisticsWord picture

SubjectbotaObjectAdverbial

1. sjukdom	116	1. sjukdom	707	1. på 1700-tal	14
2. medicin	37	2. cancer	222	2. från sjukdom	12
3. cancer	22	3. förkylning	142	3. från cancer	8
4. horn	16	4. huvudvärk	142	4. i bukspottskörteln	6
5. bajs	18	5. fanatiker	72	5. till	8
6. diabetes	12	6. diabetes	81	6. till och med	8
7. behandling	18	7. bakfylla	54	7. till och med	8
8. pbl-rädsla	8	8. baksmälla	51	8. med ibt	4
9. patient	21	9. homosexualitet	46	9. utav genterapi	4
10. läkemedel	15	10. patient	72	10. via nojesguiden	4
11. huvudvärk	15	11. flygrädsla	36	11. i stadion <sup>2</sup>	4
12. spädbarn	10	12. krämpa	35	12. i stadium	4
13. baksmälla	10	13. åkomma	36	13. före avspark	3
14. homosexualitet	11	14. autism	37	14. under tid	7
15. åkomma	8	15. depression	39	15. numera	7

Show word class

# Karp

## KARP

[Log in](#) [Download resources](#) [Cite Karp](#) [Svenska](#) [English](#) [About Karp](#)

21 lexical resources selected — 460,765 lexical entries

[Search history](#)

Simple

Extended

Editor



bota (verb)

Search

through Diapivot



Hits per page: 25

Full Hits (9)

Statistics

Swesaurus

SweFN

Parole+

Kelly

LWT

WordNet-Saldo

Lexin

Diakronisk pivot

### Swesaurus (1)

Sense Degree Sense Type Source

bota ◀ 90 ▶ *hela* syn fslbota ◀ 100 ▶ *kurera* syn fslbota ◀ 60 ▶ *läka* syn dfsbota ◀ 80 ▶ *läka*<sup>2</sup> syn dfs

### SweFN (1)

ID Cure\_mod

Domain Med

Core elements Affliction, Body\_part, Healer, Medication, Patient, Treatment

Peripheral elements Degree, Duration, Manner, Motivation, Place, Purpose, Time

Example sentences (Salvan)<sup>Medication</sup> [läker]<sup>LU</sup> [skrubbsår och brännsår]<sup>Affliction</sup> ·[Läkaren]<sup>Healer</sup> [botade]<sup>LU</sup> både [ryggskott och vatten i knät]<sup>Affliction</sup>, innan hon hamnade i svårigheter efter att ha utfört ett föryngringsexperiment med patienten.[Genterapi]<sup>Treatment</sup> [botade]<sup>LU</sup> [dödssjuka i cancer]<sup>Patient</sup> ·[Läkaren]<sup>Healer</sup>1 [botar]<sup>LU</sup>1 [kroppen]<sup>Body\_part</sup>1 och [filosofen]<sup>Healer</sup>2 [botar]<sup>LU</sup>2 [själen]<sup>Body\_part</sup>2, men det krävs ett engagemang för att lyckas.Traditionellt har [stafylokockinfektioner]<sup>Affliction</sup> enkelt [botats]<sup>LU</sup> [med antibiotika]<sup>Medication</sup> ·Syftet med terapi för en personlighetsstörning är inte att [fullständigt]<sup>Degree</sup> [bota]<sup>LU</sup> [patienten]<sup>Patient</sup>, eftersom det varken är möjligt eller eftersträvaransvärt.Han ansåg även att [Gud]<sup>Healer</sup> hade [helat]<sup>LU</sup> [honom]<sup>Patient</sup> [från cancer]<sup>Affliction</sup> ·[Transplantation]<sup>Treatment</sup> kan ha [botat]<sup>LU</sup> [hiv-smittad]<sup>Patient</sup> ·

Comment New frame / Ny ram. Den modifierade ramen bygger på ramen Cure i BFN, men den ges här en snävare tolkning som implicerar att ett positivt resultat av medicinsk behandling föreligger.

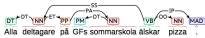
Lexical units

[Sort by POS](#)

# An quick introduction to Corpus Workbench

- A database system for querying annotated texts.
- Uses regular expressions over attributed words.
- Part of the backend of Korp.
- Input format:

```
<paragraph> [Vinea XRG]
<sentence id="8f7-83b"> [Vinea XRG]
PM: Egennamn
```



Alla deltagare på GFs sommarskola älskar pizza .

ord	msd	lemma	lex	saldo	prefix	suffix	ref	dephead	deprel
Alla	DT, UTR+NEU, PLU, IND+DEF	all	<a href="#">all_pn.1</a>	<a href="#">all_.1</a>			1	2	DT
deltagare	NN, UTR, PLU, IND, NOM	deltagare	<a href="#">deltagare_nn.1</a>	<a href="#">deltagare_.1</a>			2	6	SS
på	PP	på	<a href="#">på_pp.1</a>	<a href="#">på_.1</a>			3	2	ET
GFs	PM, GEN						4	5	DT
sommarskola	NN, UTR, SIN, IND, NOM				<a href="#">sommär_.nn.1</a>	<a href="#">skola_.nn.1</a>	5	3	PA
älskar	VB, PRS, AKT	älska	<a href="#">älska_vb.2</a> , <a href="#">älska_vb.1</a>	<a href="#">älska_.2</a> , <a href="#">älska_.1</a>			6		ROOT
pizza	NN, UTR, SIN, IND, NOM	pizza	<a href="#">pizza_nn.1</a>	<a href="#">pizza_.1</a>			7	6	OO
.	MAD						8	6	IP

```
</sentence>
</paragraph>
```

# Corpus Query Language (CQL)

- Basic form, a box = word/token

[attr=value]

- Example:

[word="pizza"]

- Regular expression:

[word="pizz(a|or)"]

- Boolean expression:

[word="pizz(a|or)" & (pos="VB" | pos="JJ")]

# Corpus Query Language (CQL)

- Comparisons: =, !=, <=, >=, !<=, !>=, ( ==, != )  
[c >= 5]
- Sequences of tokens/words  
[word="älskar" []{0,3} [word="pizza"]
- A longer example  
"catch|caught" [tag="DT"] [tag="JJ"]\* [tag="N.\*"] |  
[tag="N.\*"] "was|were" "caught"

# Demonstration: overview

1. **Korp annotation lab**

<<http://spraakbanken.gu.se/korp/annoteringslabb>>

2. **Korp**

<<http://spraakbanken.gu.se/korp>>

3. **Karp**

<<http://spraakbanken.gu.se/karp>>