



UNIVERSITY OF
GOTHENBURG

IMPROVING DICTIONARIES BY MEASURING ATYPICAL RELATIVE WORD FORM FREQUENCIES

KRISTIAN BLENSENIUS & MONICA VON MARTENS

ELEX 2019, SINTRA, PORTUGAL, 1 OCT. 2019

THE DICTIONARIES

Svensk ordbok & Lexin

SO (Svensk ordbok, 'Dictionary of Swedish')



SVENSKA AKADEMIENS
ORDBÖCKER

HEM ORDBÖCKERNA GRAMMATIK AKADEMIEN KONTAKT

Sök i tre ordböcker på en gång

Sökhjälp

Sök ord ...



SAOL

tryckår: 2015

lex·ik·on [lekˈs-] substantiv ~*et*; pl. ~
• ordbok, uppslagsbok

Visa mer

SO

tryckår: 2009

lex·ikon substantiv ~*et*, plur. ~ *äv*.
lexika, best. plur. ~*en*
ORDLED: lex-ik-on-et

- förteckning över ett språks (viktigare) ord jämte ordförklaringar på samma språk el. genom översättning till annat språk
- JFR ordbok, ordlista, encyklopedi, glossarium, tesaurus

Visa mer

SAOB

tryckår: 1940

LEXIKON *läk⁴sikon*, äv. 302, n.
(**SCHÜCK VittA 1: 138** (*i handl. fr. 1629*) osv.) ((†) r. **CAEHRENSVÅRD Brev 2: 210 (1798)**); best. **-et**; pl. = (**FRÖDING Brev 4 (1878)** osv.) l. **lexika** (**REGNÉR Begr. 11 (1780)** osv.) ((†) **lexiker LEOPOLD 3: 458 (1794, 1816)**). **DENS. i 2SAH 17: 40**

Lexin

Slå upp

Bilder och filmer

Nätbokhandel

Om Lexin

Kontakt

Andra nätlexikon

Till Från Båda

svenska

lexikon

Slå upp!

?

► [Svenskt tangentbord](#)

Spara ordet "lexikon"

Inställningar

▣ **lexikon** [lək:sikån] LYSSNA ↴ subst.
(lexikonet, lexikon eller lexika, lexikonen)

▣ ordbok; uppslagsbok

Sammansättningar:

- konversations|lexikon

Exempel:

- ett svensk-engelskt lexikon
- hon slog upp ordet i lexikonet


Det svenska materialet är sammanställt av Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet.

WHAT'S THE PROBLEM?


Frequency information – rarely explicit

Svensk ordbok

tryckår: 2009

konstighet  [kån`st-] substantiv ~en ~er
ORDLED: konst-ig-het-en

(ofta plur.) konstig företeelse om yttrande, handling, föremål etc.

Döj 


EXEMPEL: *vad är det för konstigheter du har för dig?*

- spec. (i nekande sats) om ngt som är svårt att utföra el. är märkligt svårighet, problem


EXEMPEL: *det här är inga konstigheter, det är bara att skruva av hjulet*

HIST.: sedan 1814

The plural is typically used in certain phrases.

sann  adjektiv *sant*

1 som överensstämmer med verkligheten om påstående e.d.
JFR riktig 1

Döj 

EXEMPEL: *en sann berättelse; en sann historia; reportaget ger en sann bild av händelseförloppet*

- spec. i ett fast uttryck

EXEMPEL: *han fick jobbet och vann hundra tusen på tips – det var nästan för bra för att vara sant*

- ibl. med stark betoning på ärlighet

EXEMPEL: *det är viktigt att vara sann mot sig själv*

KONSTRUKTION: *sann (mot ngn/ngt)*

det var så sant! det höll jag på att glömma!: *det var så sant, det har kommit ett paket med posten till dig!*

det är så sant som det är sagt det stämmer verkligen: *livet består inte bara av arbete, det är så sant som det är sagt*

inte sant? förhåller det sig inte så? som begäran om instämmande, ofta försvagat: *skoltrötta blir väl alla ibland, inte sant?*

HIST.: sedan 1000-talet (runsten, Aspö, Södermanland (Sveriges runinskrifter)); runform *sant* (neutr.), fornsv. *sander*; germ. ord, nära besl. med bl.a. lat. *sum* 'jag är'

The *t*-inflection, *sant*, describes clause content.

Frequency information – rarely explicit

Slå upp Bilder och filmer Nätbokhandel Om Lexin Kontakt Andra nätlexikon

Till Från Båda

Välj språk

konstigheter

Slå upp! ?

► [Svenskt tangentbord](#)

Spara ordet "konstigheter"

Inställningar

Ordet **konstigheter** finns inte i lexikonet, men ordet **kantigheter** finns.

- ▣ **kantighet** [ˈkɑnːtɪgheːt] subst.
(kantigheten, kantigheter, kantigheterna)
- ▣ egenskapen att vara kantig

Exempel:

- för att mjuka upp rummets kantighet lade de ut en rund matta

Konstighet(er) ‘oddnesses’ not included.

Frequency information – rarely explicit

Slå upp Bilder och filmer Nätbokhandel Om Lexin Kontakt Andra nätlexikon

Till Från Båda

Välj språk

sant

Slå upp! ?

► [Svenskt tangentbord](#)

Spara ordet "sant"

Inställningar

sann [san:] LYSSNA adj.
(sant, sanna)

1. som stämmer med verkligheten

Exempel:

- Är det sant att du ska flytta?
- reportaget ger en sann bild av verkligheten i Gaza

2. riktig, äkta

Exempel:

- hon är en sann idealist

Possible to search for word form, but few examples

AIMS

Aims

- Illustrate how focusing on relative word-form frequencies can improve dictionaries
 - How to capture “peculiarities” in terms of which forms are used
 - Which headword forms may be most easily recognized by the dictionary user
 - How to present word forms in the dictionary articles.
- Present a method for examining word form frequencies obtained from our morphological database.
 - Evaluate the results with respect to Svensk ordbok (SO) ‘Dictionary of Swedish’ and, to some extent, with respect to Lexin.
 - Focus: nouns and adjectives.

SO

- Descriptive general definition dictionary
- Intended user: anyone who “wants to know more about the Swedish vocabulary” (p. IX, our translation)
- Some headword forms are adjusted according to dominant morphological form, e.g. plural **döddagar** (‘dying days’) and present **lär** (‘should’).
 - Almost 100 % to 0 % in these cases
- Frequency information not used otherwise.

Lexin

- Learner's dictionary
- “primarily intended for people learning Swedish” (our translation)
- Verbal headwords in the present tense (not frequency motivated; present tense is assumed to make it easier to produce other forms, Gellerstam 1999:7f.)



Some previous word-form studies: lexicography

Selecting and distributing information: which inflections and derivations (regular/irregular) should be included in a dictionary entry/appendix?
Considerations: type of dictionary, user profile, L1/L2, morphological complexity of the language

- Atkins & Rundell (2008:28f.,180); Svensén (2009); Heuberger (2018)

Canonical and special headword forms, e.g. Eng. **arm** vs. **arms** 'weapons'; Sw. **datum** 'date' vs. **data** 'data', suppletive inflections, e.g. *feet* vs. *foot*

- Svensén (2009:106,131)

Reception (finding the appropriate form) vs. production (producing the correct form based on the headword form)

- Svensén (2009:124)

=

- Word-form frequency not much discussed in lexicography
- How word-form frequency may affect lexicographical information (apart from choice of headword form) not much discussed
- Often based on space restrictions in printed dictionaries (condensing text, e.g. *show ~ed ~n* vs. *show showed shown*)

Some previous word-form studies: morphology

Corpora in morphology work, e.g. lemma vs. word-form frequency (Rácz et al. 2016)

Productivity and frequency for roots and affixes (Rácz et al. 2016)

Frequent verbs in *-ed* easier to identify than less frequent in *-ed* (Bybee & Hopper 2001:16)

Response times/processing of word forms depending on affix, lemma size etc.

Type and token productivity. High type freq. => productive etc. (Bybee 1995 m. fl.),
productivity in different text types (Altmann et al. 2009)

Rule-based operations and/or stored as full forms? (Clahsen 2016)

Objective (corpus) frequency vs. perceived frequency/knowledge. Correlation for highly frequent words (Balota et al. 2001)

Practical work

- Choose relevant examples
- ... but do not forget the definition
- Which *blommor* 'flowers' are most frequent?

REDIGERA

blomma substantiv
blomman blommor
blomm`a 🗨

[Sök kommentarer](#)

Bklass: 11a **REDBOJN** **VISA LISTA MED %B**

1 växtedel med (oftast) vacker färg och form, som vanligen sitter väl synlig och som svarar för växtens fortplantning
SAMMANSÄTTN./AVLEDN.: *blomrik; lotusblomma; lönnblomma; smultronblomma*

EXEMPEL: *levande blommor; konstgjorda blommor; plocka blommor; blommorna började slå ut; fruktträdens blommor; blommorna sluter sig när solen går ner; försommarnatten doftade av blommor*

- även med tonvikt på den biologiska funktionen könsorgan hos växt

SAMMANSÄTTN./AVLEDN.: *hanblomma*

EXEMPEL:
en blomma utmärks främst av de sexuella ståndare och fruktblad

- ofta bildligt: *det vackraste eller högsta (i en blomster)*

JFR *blomster*

https://k5.spraakdata.gu.se/li/red/SABOB/sabob_blowup_med_bojn.php?nr=121197&tagg=

nr	tagg	regel	resultat	rensat	FLASHBSAMH-frek	ff
01(11a)	NCUSNI	=	blomma	blomma	663	(2*)
02(11a)	NCUSGI	+=/s	blomma+s	blommas	3	(1)
03(11a)	NCUSND	+=/n	blomma+n	blomman	101	(1)
04(11a)	NCUSGD	+=/n/s	blomma+n+s	blommans	4	(1)
05(11a)	NCUPNI	%sv+/or	blomm+or	blommor	2483	(1)
06(11a)	NCUPGI	%sv+/or/s	blomm+or+s	blommors	0	(1)
07(11a)	NCUPND	%sv+/or/na	blomm+or+na	blommorna	403	(1)
08(11a)	NCUPGD	%sv+/or/na/s	blomm+or+na+s	blommornas	3	(1)

Normalfördelning för substantiv: sg/pl: c. 75/25, bestämd form betydligt ovanligare än obestämd.
Adjektiv: sg genitiv(AQPUSGI/AQPNSGI) har normalt sett frekvensen 0, avvikelser kan tyda på substantivering
Transitiva verb, pres.ind./perf.part.utr.(VOIPA/AFOUSNI) c 90/10, avvikelser kan tyda på lexikalisering
ff= antal superlemman i fullformstabellen som uppvisar denna form treställigt, *= intern homografi för något av dessa

What is the word-form distribution of a “normal” word?

Example: nouns in paradigm #12 (definite: *-en* plur. *-ar*):

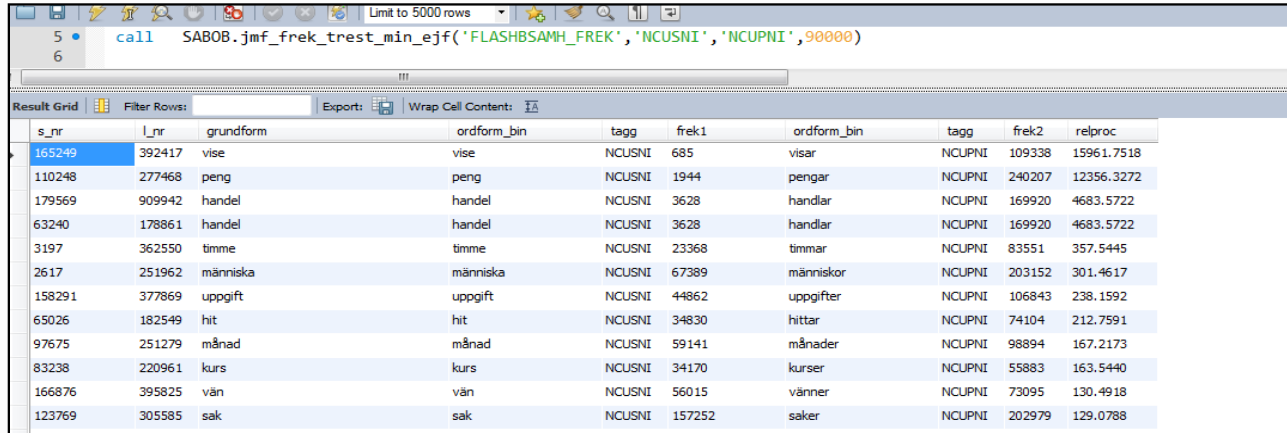
Method:

- Cf. total word-form freqs* (not lemmatized) in the dominating paradigm of the PoS.
- Identify words that “distrurb” the expected distribution.
- Use total frequencies, excluding problematic words
 - Homography?
- Compare the result against some other words.

What is the word-form distribution of a “normal” word?

bformsstatistik.xlsx											
	B	C	D	E	F	G	H	I	J	K	L
62	Korpus: Flashback Samhälle (frpn KORP)										
63	Totalfrekvens ordformer som ingår i böjningsmönstret					Totalfrekvens <i>gemena</i> ordformer som ingår i böjningsmönstret,exklusive 'med', 'ur' m fl					
64											
65	sortering	tagg	regel	antal	procent	sortering	tagg	regel	antal	procent	
66	N01a	NCUSNI	'='	33788734	82,259	N01a	NCUSNI	'='	6754268	49,932	
67	N01b	NCUSGI	'=+/s	943136	2,296	N01b	NCUSGI	'=+/s	867810	6,415	
68	N02a	NCUSND	'=+/en	3329244	8,105	N02a	NCUSND	'=+/en	3047546	22,529	
69	N02b	NCUSGD	'=+/en/s	117807	0,287	N02b	NCUSGD	'=+/en/s	103685	0,767	79,643
70	N03a	NCUPNI	'=+/ar	2353015	5,728	N03a	NCUPNI	'=+/ar	2262068	16,723	
71	N03b	NCUPGI	'=+/ar/s	31633	0,077	N03b	NCUPGI	'=+/ar/s	24491	0,181	
72	N05a	NCUPND	'=+/ar/na	496796	1,209	N05a	NCUPND	'=+/ar/na	454031	3,356	
73	N05b	NCUPGD	'=+/ar/na/s	15773	0,038	N05b	NCUPGD	'=+/ar/na/s	13034	0,096	20,357
74											
75				41076138	100				13526933	100	
76		varav: ***									
77			med	4551735		OBS att det också finns ett stort antal homografer NCUPNI=V0IPA					
78			ur	173609		<i>t.ex. delar, bilar, hamnar, vågar</i>					
79			en	6160818		ingen enskild extremt stor dock					
80			för	4944141							
81			man	3421997							
82			var	2256371							
83			vad	1267693							
84			här	1069589							
85			dom	938577							
86											
87			Summa:	24784530							

Finding outliers: indef. plur. vs. indef. sing. nouns



call SABOB.jmf_frekw_trest_min_ejf('FLASHBSAMH_FREK', 'NCUSNI', 'NCUPNI', 90000)

s_nr	l_nr	grundform	ordform_bin	tagg	frek1	ordform_bin	tagg	frek2	relproc
165249	392417	vise	vise	NCUSNI	685	visar	NCUPNI	109338	15961.7518
110248	277468	peng	peng	NCUSNI	1944	pengar	NCUPNI	240207	12356.3272
179569	909942	handel	handel	NCUSNI	3628	handlar	NCUPNI	169920	4683.5722
63240	178861	handel	handel	NCUSNI	3628	handlar	NCUPNI	169920	4683.5722
3197	362550	timme	timme	NCUSNI	23368	timmar	NCUPNI	83551	357.5445
2617	251962	människa	människa	NCUSNI	67389	människor	NCUPNI	203152	301.4617
158291	377869	uppgift	uppgift	NCUSNI	44862	uppgifter	NCUPNI	106843	238.1592
65026	182549	hit	hit	NCUSNI	34830	hittar	NCUPNI	74104	212.7591
97675	251279	månad	månad	NCUSNI	59141	månader	NCUPNI	98894	167.2173
83238	220961	kurs	kurs	NCUSNI	34170	kurser	NCUPNI	55883	163.5440
166876	395825	vän	vän	NCUSNI	56015	vänner	NCUPNI	73095	130.4918
123769	305585	sak	sak	NCUSNI	157252	saker	NCUPNI	202979	129.0788

Mostly plur.

125752	309821	sedan	sedan	NCUSNI	456128	sedaner	NCUPNI	7	0.0015
46158	142881	en	en	NCUSNI	6160818	enar	NCUPNI	73	0.0012
2573	241578	med	med	NCUSNI	4551735	medar	NCUPNI	27	0.0006
146334	352771	så	så	NCUSNI	4339504	sår	NCUPNI	3	0.0001
26928	102160	alla	alla	NCUSNI	864896	allor	NCUPNI	0	0.0000
131948	322685	skulle	skulle	NCUSNI	1374964	skullar	NCUPNI	0	0.0000
2429	202624	khan	kan	NCUSNI	2857104	kaner	NCUPNI	0	0.0000
2737	262549	nåt	nåt	NCUSNI	105958	nåtar	NCUPNI	0	0.0000

Mostly sing.?

Finding outliers: adjectival gender

iter Rows: Export: Wrap Cell Content: Fetch rows:

l_nr	grundform	ordform_bin	tagg	frek1	ordform_bin	tagg	frek2	relproc
250449	mycken	mycken	AQPUSNI	266	mycket	AQPNSNI	832851	313101.8797
315789	självfallen	självfallen	AQPUSNI	3	självfallet	AQPNSNI	6848	228266.6667
324087	skämd	skämd	AQPUSNI	30	skämt	AQPNSNI	18700	62333.3333
763907	fattad	fattad	AQPUSNI	40	fattat	AQPNSNI	16016	40040.0000
307932	samtidig	samtidig	AQPUSNI	518	samtidigt	AQPNSNI	98883	19089.3822
346272	studerad	studerad	AQPUSNI	47	studerat	AQPNSNI	8637	18376.5957
118122	betydlig	betydlig	AQPUSNI	381	betydligt	AQPNSNI	56189	14747.7690
199953	jäklig	jäklig	AQPUSNI	53	jäkligt	AQPNSNI	7002	13211.3208
371906	tveklös	tveklös	AQPUSNI	18	tveklöst	AQPNSNI	2171	12061.1111
338591	spårlös	spårlös	AQPUSNI	26	spårlöst	AQPNSNI	2955	11365.3846
194420	innerlig	innerlig	AQPUSNI	34	innerligt	AQPNSNI	3369	9908.8235
271760	osagd	osagd	AQPUSNI	35	osagt	AQPNSNI	3371	9631.4286
395243	väldig	väldig	AQPUSNI	3298	väldigt	AQPNSNI	313153	9495.2395
301256	rund	rund	AQPUSNI	2268	runt	AQPNSNI	214467	9456.2169
133223	definitiv	definitiv	AQPUSNI	400	definitivt	AQPNSNI	32668	8167.0000
236196	lättvindig	lättvindig	AQPUSNI	16	lättvindigt	AQPNSNI	1157	7231.2500

← **betydligt**: headword in SO
betydlig: headword in Lexin

EDITOR INTERPHASE

Nouns

kontrollera/uppdatera böjning - Mozilla Firefox

https://k5.spraakdata.gu.se/ij/red/SABOB/sabob_blowup_med_bojning.php?lnr=26

Nuvarande värden: subst. **närhet** lnr262737

Valj böjningsgrupp: 34a

OBS! Lista med vanligaste böjningsklasserna visas om man tömmer fältet och klickar i det! Ordbildning (för **avstavning och segmentering**) tagg:

GENERERA VISA LISTA MED %NÄRHET

Ordbildning	teknisk stam	böjning	grupp	tagg
Nuvarande:	&en		34a	

Allmän regel för genitiv-s: Om ordet är mer än 1 bokstav långt och slutar på något av s, x, z, S, X, Z: +s regel påförs inget s
i SAOL/SMDB visas + endast om det finns / i regeln, här visas + framför alla ändelser

Allmän regel för n+n: +/na efter ord som slutar på 'n' dubblettar ej 'n'

börjar med kombo tvåställda **34** och treställiga **34a**

nr	tagg	regel	resultat	rensat	FLASHBSAMH-frek	ff
01(34a)	NCUSNI	=	närhet	närhet	9268	(1)
02(34a)	NCUSGI	=+/s	närhet+s	närhets	10	(1)
03(34a)	NCUSND	=+/en	närhet+en	närheten	44386	(1)
04(34a)	NCUSGD	=+/en/s	närhet+en+s	närhetens	4	(1)

Normalfördelning för substantiv: sg/pl: c. 75/25, bestämd form betydligt ovanligare än obestämd.

Adjektiv: sg genitiv(AQPUSGI/AQPNSGI) har normalt sett frekvensen 0, avvikelser kan tyda på substantivering

Transitiva verb, pres.ind./perf.part.utr:(VOIPA/AF0USNI) c 90/10, avvikelser kan tyda på lexikalisering

ff= antal superlemman i fullformstabellen som uppvisar denna form treställigt + tvåställigt, *= intern homografi för något av dessa

prep. phrase *i närheten* 'in the vicinity'?



Nouns, genitive indefinite sing. and plur.

kontrollera/uppdatera böjning - Mozilla Firefox
https://k5.spraakdata.gu.se/fi/red/SABOB/sabob_blowup_med_bojn.php?lnr=334353

Nuvarande värden:
subst. **sort** lnr334353
Välj böjningsgrupp:
17a

OBS! Lista med vanligaste böjningsklasserna visas om man tömmer fältet och klickar i det!
Ordbildning (för **avstavning och segmentering**) tagg:

GENERERA VISA LISTA MED %SORT

Ordbildning	teknisk stam	böjning	grupp	tagg
Nuvarande:		&en &er	17a	

Allmän regel för genitiv-s. Om ordet är mer än 1 bokstav långt och slutar på något av s, x, z, S, X, Z: +s-regel påförs inget s
SAOL/SMDB visas + endast om det finns / i regeln, här visas + framför alla ändelser
Allmän regel för n+n: +/na efter ord som slutar på 'n' dubbleras ej 'n'

börjar med kombo tvåställa 17 och treställiga 17a

nr	tagg	regel	resultat	rensat	FLASHBSAMH-frek	ff
01(17a)	NCUSNI	=	sort	sort	4315	(1)
02(17a)	NCUSGI	=+/s	sort+s	sorts	25478	(1)
03(17a)	NCUSND	=+/en	sort+en	sorten	2058	(1)
04(17a)	NCUSGD	=+/en/s	sort+en+s	sortens	4744	(1)
05(17a)	NCUPNI	=+/er	sort+er	sorter	1075	(1)
06(17a)	NCUPGI	=+/er/s	sort+er+s	sorters	4018	(1)
07(17a)	NCUPND	=+/er/na	sort+er+na	sorterna	160	(1)
08(17a)	NCUPGD	=+/er/na/s	sort+er+na+s	sorternas	15	(1)

Normalfördelning för substantiv: sg/pl: c. 75/25, bestämd form betydligt ovanligare än obestämd.
Adjektiv: sg genitiv(AQPUSGI/AQPNSGI) har normalt sett frekvensen 0, avvikelse kan tyda på substantivering
Transitiva verb, pres.ind./perf.part.utr.(V0IPA/AF0USNI) c 90/10, avvikelse kan tyda på lexikalisering
ff= antal superlemman i fullformstabellen som uppvisar denna form treställigt + tvåställigt, *= intern homografi för något av dessa

Relevant examples?

Look up in a corpus

en sorts x
olika sorters x

SO

sort [sårt] substantiv ~en ~er
ORDLED: sort-en

- grupp av företeelser som har åtminstone ngt drag gemensamt ofta med tanke på den gemensamma uppsättningen egenskaper
SYN. slag 11 JFR art 1, typ 1, ²form 2, kategori, species 1

Dölj

EXEMPEL: *myntsort*; *stilsort*; *vinsort*; *sju sorters kakor*; *det är bara folk av samma sort som bor runt omkring*; *en bil av den dyra sorten*; *drycker av olika sorter bjöds omkring*; *den sortens omtänksamhet som gör livet uthärdligt*; *jag kan nog din sort!*

Adjectives: superlative outliers

kontrollera/uppdatera böjning - Mozilla Firefox

https://k5.spraakdata.gu.se/fi/red/SABOB/sabob_blowup_med_bojn.php?lnr=120432&tagg=

Nuvarande värden:
adj. **blek** lnr 120432

Välj böjningsgrupp: 41b

OBST Lista med vanligaste böjningsklasserna visas om man tömmer fältet och klickar i det!

Ordbildning (för avstavning och segmentering) tagg GENERERA

VISA LISTA MED %BLEK

Nuvarande:	Ordbildning	teknisk stam	böjning	grupp	tagg
Nuvarande:		&t &a	41b		

Allmän regel för genitiv-s: Om ordet är mer än 1 bokstav långt och slutar på något av s, x, z, S, X, Z, +s-regel påförs inget s /SAOU/SMDB visas + endast om det finns / i regeln, här visas + framför alla ändelser

Allmän regel för n+n: +/na efter ord som slutar på 'n' dubbleras ej 'n'

börjar med kombo tvåställda 41 och treställda 41b

nr	tagg	regel	resultat	rensat	FLASHBSAMH-frek	ff
01(41b)	AQPUSNI	=	blek	blek	566	(2)
02(41b)	AQPNSNI	=+t	blekt	blekt	157	(2 ⁺)
03(41b)	AQP**ND	=+a	bleka	bleka	405	(3)
04(41b)	AQPMSND	=+e	bleke	bleke	21	(3 ⁺)
05(41b)	AQC**N*	=+/are	blek+are	blekare	43	(2 ⁺)
06(41b)	AQS**NI	=+/ast	blek+ast	blekast	7	(1)
07(41b)	AQS**ND	=+/aste	blek+aste	blekaste	1765	(1)
s1(41)	AQPUSGI	=+s	blek+s	bleks	17	(1+1)
s2(41)	AQPNSGI	=+/t/s	blek+t/s	blekts	17	(1+2)
s3(41)	AQP**GD	=+/as	blek+as	blekas	10	(2+1)
s6(41)	AQPMSGD	=+/e/s	blek+e+s	blekes	0	(3 ⁺ +1)
s7(41)	AQC**G*	=+/are/s	blek+are+s	blekares	0	(1 ⁺ +1)
s9(41)	AQS**GD	=+/aste/s	blek+ast+e+s	blekastes	0	(0+1)
x1(41)	NCUSNI	=+/het	blek+het	blekhet	153	(1+1)
x2(41)	NCUSGI	=+/het/s	blek+het+s	blekhets	0	(1+1)
x3(41)	NCUSND	=+/het/en	blek+het+en	blekheten	1	(1+1)
x4(41)	NCUSGD	=+/het/en/s	blek+het+en+s	blekhetens	1	(1+1)
x5(41)	NCUPNI	=+/het/er	blek+het+er	blekheter	0	(0+1)
x6(41)	NCUPGI	=+/het/er/s	blek+het+er+s	blekheters	0	(0+1)
x7(41)	NCUPND	=+/het/er/na	blek+het+er+na	blekheterna	0	(0+1)
x8(41)	NCUPGD	=+/het/er/na/s	blek+het+er+na+s	blekheternas	0	(0+1)

Normalfördelning för substantiv. sg/pl: c. 75/25, bestämd form betydligt ovanligare än obestämd.

Adjektiv. sg genitiv(AQPUSGI/AQPNSGI) har normalt sett frekvensen 0, avvikelse kan tyda på substantivering

blek adjektiv ~t

tryckår: 2009

1 som har svag, vitaktig färg i förhållande till den naturliga; särsk. om ansikte (el. om person, med tanke på detta)

JFR färglös

Visa mer

2 som gör ett kraftlöst intryck om person, handling o.d. <särsk. i sportjargong>

JFR ²matt 1

Döj

EXEMPEL: *partiet har ibland kallats en blek kopia av moderaterna; hemmalaget var ovanligt blekt igår; hans dikter är bara blek romantisk efterklang*

o i vissa uttryck äv. som rent förstärkningsord

EXEMPEL: *bleka döden; den bleka nöden; jag har inte den blekaste aning om vad klockan är*

HIST.: sedan 1623



blekaste aning
'slightest clue'

Ordering senses

- Should the most frequent form always be used as the headword form?
 - For example, the distribution of sing. and plur. forms may vary between senses ▶ using the most frequent form may be problematic

tryckår: 2009

1 blomm`a substantiv ~n
blommor
ORDLED: blomm-an

1 växt del med (oftast) vacker färg och form, som vanligen sitter väl synlig och som svarar för växtens fortplantning [Visa mer](#)

2 växt med vackert färgade och formade (fortplantnings)delar spec. om örter
JFR *blomster* [Visa mer](#)

3 bild som återger färg eller form hos de vackraste växtdelarna särsk. som detalj i utsmyckning el. mönster

handel Om Lexin Kontakt Andra nätlexikon

blomma [ˈblɔmːa] LYSSNA subst.
(blomman, blommor, blommorna)

1. växt, ört
BILD SVENSKA
Sammansättningar:

- blom|bukett

Exempel:

- barnen plockade blommor på ängen
- Kan du vattna mina blommor när jag är bortrest?

2. (den vackert färgade) kronan av en blomma, blomkrona
BILD SVENSKA
Sammansättningar:

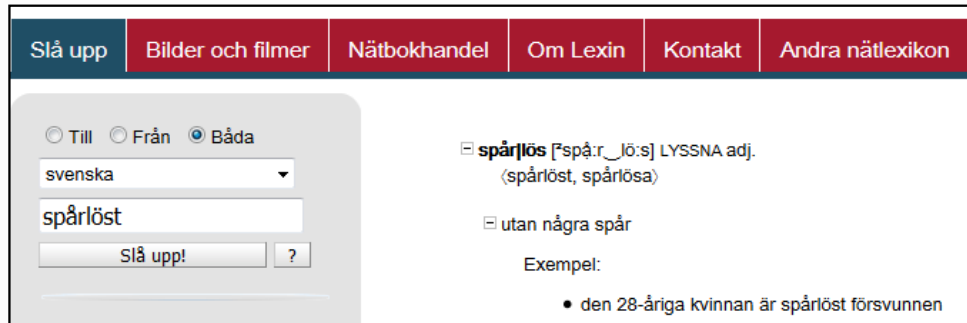
- blom|ställning

Exempel:

- fruktträdens blommor

Is the headword a good representative?

- Lexin: Adjectives which are almost always used as adverbials (*t*-form): should be made clear that the headword form is infrequent
- SO: OK ordering of senses?



Slå upp Bilder och filmer Nätbokhandel Om Lexin Kontakt Andra nätlexikon

Till Från Båda

svenska

spår|lös

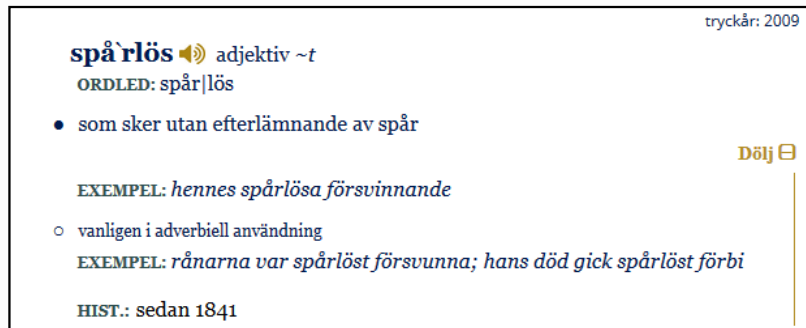
Slå upp! ?

spår|lös [ˈspɑːr_ˈløːs] LYSSNA adj.
(spår|lost, spår|lösa)


utan några spår

Exempel:

- den 28-åriga kvinnan är spår|lost försvunnen




tryckår: 2009

spår|lös  adjektiv ~t

ORDLED: spår|lös

- som sker utan efterlämnande av spår

Dölj 

EXEMPEL: hennes spår|lösa försvinnande

- vanligen i adverbial användning

EXEMPEL: rånarna var spår|lost försvunna; hans död gick spår|lost förbi

HIST.: sedan 1841

Practical consequences

- Increased awareness about frequency distribution when structuring the database and the dictionary entries.
- Particularly frequent word forms which are not headwords should have a reference to the appropriate entry.

Demoversion SABOB (Kommentera? [Tryck här](#))

Sökord: pengar

Öppna gärna svenska.se i eget fönster för jämförelse.

Artikel

Frassök

Lista

Konstr.

pengar → se [peng](#)

Ordet **pengar** finns inte som eget uppslagsord.

© Svenska Akademien och Göteborgs universitet, Institutionen för svenska språket

References

- Altmann, Eduardo G.; Janet B. Pierrehumbert & Adilson E. Motter. 2009. Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLoS One* 4/11, e7678.
- Atkins BTS, M. Rundell (2008). *The Oxford Guide to Practical Lexicography*. OUP.
- Balota, D. A., M. Pilotti & M. J Cortese (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition* 29/4, 639–47.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10/5, 425–55.
- Bybee, J., and P. Hopper (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins.
- Clahsen, H. (2016). Experimental Studies of Morphology. In: Hippiusley, A. & G. Stump, *The Cambridge Handbook of Morphology*. CUP, 792–819.
- Gellerstam, M. (1999). LEXIN – lexikon för invandrare. *LexicoNordica* 6, 3–17.
- Heuberger, R. (2018). Dictionaries to assist teaching and learning. I: Fuertes-Olivera, Pedro A. *The Routledge Handbook of Lexicography*. Milton Park, Abingdon, Oxon: Routledge. S. 300–316.
- Rácz, P., V. Papp & J. Hay (2016). Frequency and corpora. In: Hippiusley, A. & G. Stump, *The Cambridge Handbook of Morphology*. CUP, 685–709.
- Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. CUP.



UNIVERSITY OF
GOTHENBURG

kristian.blensenius@gu.se
monica.von.martens@gu.se