

Ordtavlor och Talspråksfrekvenser från GSLC, sammanställda med tanke på användning i talande samtalsjälpmedel. Arbetsmaterial, Bitte Rydeman 2009.

Listorna i det här dokumentet kommer från Göteborgs Talspråskkorpus (GSLC) och sammanställdes 2009 som ett arbetsmaterial i samband med arbetet med min doktorsavhandling. För beskrivning av talspråskorpusen och introduktion till referenserna i det här dokumentet hänvisas till avhandlingen (Rydeman, 2010). Jag har ställt samman materialet i detta dokument som inspiration och underlag för vokabulär i samtalshjälpmedel. Även om språket ständigt utvecklas tror jag att de flesta av de mest frekventa orden håller sig ganska stabila.

Lund, 2013-10-01
Bitte Rydeman

Innehåll

Ordfrekvenser	2
De 235 mest frekventa orden	2
De 60 mest frekventa orden.....	4
Ordklasser och ordtavlor.....	5
Talspråkets grammatik	5
Ordklasser	5
Ordtavlor	6
Frekventa kollokationer	7
Referenser	7

Ordfrekvenser

De 235 mest frekventa orden

(disambiguerade skriftspråksammanförda).

Korpusens storlek 1 263 408 tokens.

det 1	av 41	hon 81
är 2	<u>också 42</u>	sådant 82
och 3	ha 43	jo 83
ja 4	får 44	upp 84
att 5	när 45	menar 85
jag 6	skulle 46	<u>någonting 86</u>
så 7	väl 47	nej 87
som 8	bara 48	säger 88
inte 9	tycker 49	mer 89
<u>har 10</u>	säga 50	naturen 90
vi 11	sen 51	väldigt 91
på 12	vara 52	få 92
man 13	finns 53	mej 93
då 14	alltså 54	två 94
i 15	hur 55	utan 95
du 16	mycket 56	se 96
en 17	kommer 57	sätt 97
ju 18	något 58	andra 98
men 19	lite 59	tar 99
dom 20	sej 60	sade 100
för 21	tror 61	från 101
den 22	vet 62	hm 102
här 23	<u>någon 63</u>	jaha 103
om 24	hade 64	kronor 104
m 25	kanske 65	ingen 105
<u>eh 26</u>	öh 66	gå 106
kan 27	göra 67	ser 107
med 28	gör 68	dej 108
där 29	blir 69	fall 109
var 30	går 70	sådana 110
ska 31	vill 71	sådan 111
till 32	just 72	<u>exempel 112</u>
nä 33	ut 73	okej 113
vad 34	liksom 74	än 114
eller 35	ni 75	oss 115
va 36	måste 76	de 116
ett 37	bra 77	nog 117
äh 38	ta 78	detta 118
nu 39	alla 79	olika 119
han 40	in 80	kunna 120

varit 121	femtio 163	folk 205
hela 122	riktigt 164	innan 206
många 123	hej 165	liten 207
precis 124	gång 166	stor 208
egentligen 125	gäller 167	börja 209
helt 126	min 168	nästan 210
tack 127	fem 169	vem 211
saker 128	aldrig 170	tid 212
tre 129	idag 171	sak 213
några 130	<u>svårt</u> 172	åt 214
bli 131	känner 173	fel 215
samma 132	kunde 174	tillbaka 216
utav 133	vidare 175	heller 217
över 134	första 176	haft 218
rätt 135	heter 177	använda 219
faktiskt 136	mot 178	jaså 220
annat 137	människor 179	blev 221
bort 138	igen 180	sin 222
komma 139	även 181	sitter 223
ner 140	vid 182	fyra 224
<u>klart</u> 141	tänka 183	ifrån 225
visst 142	tänkte 184	förstår 226
behöver 143	inget 185	tio 227
kom 144	fått 186	själva 228
natur 145	varför 187	vår 229
ganska 146	annan 188	naturlig 230
tänker 147	vissa 189	sett 231
fram 148	mera 190	titta 232
år 149	fast 191	noll 233
allt 150	börjar 192	kvar234
ligger 151	tjugo 193	först 235
ändå 152	stora 194	
fick 153	människan 195	
del 154	gick 196	
efter 155	frågan 197	
själv 156	alltid 198	
står 157	mellan 199	
fråga 158	direkt 200	
gjort 159	gjorde 201	
bättre 160	<u>håller</u> 202	
därför 161	tiden 203	
under 162	sagt 204	

De 60 mest frekventa orden

(disambiguerade och skriftspråksammanförda, minus m, eh, va och äh, som inte låter sig uttalas så bra med talsyntes). Korpusens storlek 1 263 408 tokens

Ord/rank	Antal ggr	Delsumma	Procent				
det 1	77810			vad 34	7629		
är 2	36843			eller 35	6394		
och 3	35471			ett 37	5979		
ja 4	32404			nu 39	5090		
att 5	30440			han 40	5072		
jag 6	28628			av 41	4729		
så 7	26059			<u>också 42</u>	4626		
som 8	19205			ha 43	4535		
inte 9	18681			får 44	4377	653846	51,80%
<u>har 10</u>	18469			när 45	4289		
vi 11	18421			skulle 46	4202		
på 12	17719			väl 47	3997		
man 13	17377			bara 48	3802		
då 14	17343			tycker 49	3771		
i 15	17039			säga 50	3768		
du 16	16040			sen 51	3753		
en 17	15506			vara 52	3712		
ju 18	15286			finns 53	3592	692201	54,80%
men 19	14623			alltså 54	3469		
dom 20	14534			hur 55	3392		
för 21	14120			mycket			
den 22	14098			56	3391		
här 23	13573			kommer			
om 24	12708			57	3349		
kan 27	10875			något 58	3348		
med 28	10531			lite 59	3337		
där 29	9470			sej 60	3319		
var 30	8199			tror 61	3314		
ska 31	8105			vet 62	3271		
till 32	8075	597652	47%	<u>någon 63</u>	3208		
nä 33	7763			hade 64	3096	725226	57,40%
					725226		

Ordklasser och ordtavlor

Ordklasser och försök att skapa ordtavlor som kan komplettera bokstavstavlor och vokabulär i samtalshjälpmedel. Utifrån frekvenserna i talspråkskorpussen har jag laborerat med ordtavlor på olika sätt.

Talspråkets grammatik

Expanded clause	Inner clause									Post field
Pre field	Initial field			Mid field				End field		Post field
	inl/subj	alt fun	finit		satsadv	Konj/pr	infin	nom.l	advl	
ja	det	vad	är	en	inte	för	ha	den här	bra	eller
m	den	var	var	ett	så	på	få	en sån	lite	så
okej	jag	när	har	nån	här	till	köpa	såna	mer	tack
nä	du	hade	kan	nåt	bara	i	kolla	sånt	nya	va
hej	ni	skulle	ska	ingen	väl	med	ta	kort	in	äh

Figur 1. Ordtavla inspirerad av talspråkets grammatik (Jfr. 1.6.1, sidan 17-18 i avhandlingen, The Growth of Phrases, Rydeman 2010 och Lindström, 2002).

jag	vi	är	inte	att	då	nu	ha
du	dom	får	och	av	för	om	säga
han	en	har	också	här	i	på	vara
den	ett	kan	där	ju	med	sen	bara
det	skulle	ska	tycker	som	men	till	väl
man	när	var	vad	så	eller	ja	nej

Figur 2. Ordtavla som är inspirerad av talspråkets grammatik, med de mest frekventa orden.

Ordklasser

pron	verb	adverb	subst			konj	prep	återkoppl		adjektiv		interjek	
det	är	så	naturen	tid	sidan	och	på	ja	bra	sånt	väldig	tack	fy
jag	har	inte	sätt	saker	göteborg	att	i	m	nja	bra	stort	hej	amen
vi	kan	då	kronor	grejer	stycken	men	för	nå	javisst	sånt	god	oj	hah
man	ska	ju	fall	ord	kultur	om	med	va	nåhå	såna	viktigt	fan	äh
du	var	här	exempel	problem	sättet	eller	till	jo	aha	olika	gamla	varsågod	hejdå
en	får	där	saker	dag	språk	som	om	nej	jaja	hela	mindre	vadå	välkommen
dom	skulle	nu	natur	massa	hand	så	av	jaha	yes	själv		å	vadå
som	tycker	också	år	pengar	djur	än	från	okej	nåmen	stora		hallå	dadda
den	säga	när	del	barn	person	utan	liksom	just	absolut	liten		usch	ojdå
vad	har	väl	gång	sverige		fast	utan	ha		stor		aj	ojoj
ett	vara	bara	människor	slags		eftersom	utav	visst		bättre		hejsan	pa
han	finns	sen	människan	typ		innan	efter	kanske		svårt		förlåt	välkomna
sej	kommer	alltså	frågan	början		både	över	hm		själva		oh	god morgon
nåt	tror	hur	tiden	gång		dels	under	precis		naturlig		he	
nån	vet	lite	folk	teknik		medan	mot	jaså		nya		herregud	
ni	hade	mycket	frågan	grann		antingen	vid	aha		större			

Figur 3. Frekventa ord från talspråkskorpussen indelade efter ordklasser

Ordtavlor

Här följer ordtavlor innehållande frekventa ord, med notering om hur stor del av förekomsterna i talspråkskorpusen som de representerar:

det	en	är	inte	att	då
jag	man	har	och	i	på
du	vi	men	som	så	ja
18 av de mest frekventa orden i talspråket, ca. 36%					

Figur 4: 18 av de mest frekventa orden.

det	en	är	inte	om	att	i
den	ett	har	och	så	för	på
jag	vi	kan	där	som	då	med
du	dom	ska	här	men	nu	till
man	när	var	ju	eller	ja	nej
35 av de mest frekventa orden i talspråket, ca. 50%						

Figur 5: 35 av de mest frekventa orden.

det	en	är	inte	här	att	i	ha
den	ett	har	och	om	av	på	säga
jag	vi	kan	också	så	för	med	vara
du	dom	ska	där	som	då	till	bara
han	skulle	får	tycker	sen	nu	eller	väl
man	när	var	vad	ju	men	ja	nej
48 av de mest frekventa orden i talspråket, ca. 54%							

Figur 6: 48 av de mest frekventa orden.

det	hur	är	kan	en	inte	att	i	bli	kunna	mej	alla
den	när	blir	ska	ett	och	av	på	vara	veta	dej	bra
jag	vad	finns	vet	ingen	också	för	med	ha	tro	sej	lite
du	var	var	vill	inget	där	då	till	exempel	se	oss	mer
han	hade	har	tycker	någon	här	nu	från	få	ta	saker	mycket
hon	kom	får	tror	något	om	sen	in	göra	vänta	sätt	många
man	måste	gör	menar	några	så	ju	ut	säga	bara	fall	olika
vi	sa	säger	ser	annat	som	jo	upp	komma	nog	sån	samma
ni	skulle	kommer	tar	andra	utan	just	än	gå	väl	såna	kanske
dom	tänkte	går	.	?	!	men	väldigt	alltså	eller	sånt	liksom

Figur 7: 117 av de mest frekventa orden.

Frekventa kollokationer

I talspråskorpusen kan man se att vissa kombinationer av ord (kollokationer) förekommer väldigt frekvent. Det skulle ju kunna innebära att den som använder samtalshjälpmiddel där hela ord förekommer (med och utan symbolstöd) kan ha nytta av att ord som ingår i sådana kollokationer finns intill varandra.

det är	ja det
är det	att man
att det	om man
men det	för att
är ju	om det
så att	det finns
det var	det kan
och så	det är
det här	vi har
och det	

Figur 8: Frekventa kollokationer

Referenser

- Allwood, J. (ed.) (2000). *Talspråksfrekvenser*. Gothenburg Papers in Theoretical Linguistics S21, University of Gothenburg, Department of Linguistics.
- Lindström, J. (2005). Grammatiken i interaktionens tjänst. In J. Anward & B. Nordberg (Eds.), *Samtal och grammatik. Studier i svenskt samtalsspråk* (pp. 11-53). Lund: Studentlitteratur.
- Lindström, J. (2008). *Tur och ordning. Introduktion till svensk samtalsgrammatik*: Norstedts Akademiska Förlag.
- Rydeman, B. (2010). *The growth of Phrases. User-centred Design for Activity-based Voice Output Communication Aids*. Monographs in Linguistics 42. Department of Philosophy, Linguistics and the Theory of Science. Gothenburg University (monografi). Available as colour pdf at: <http://hdl.handle.net/2077/22204>
- Todman, J., & Alm, N. (2003). Modelling conversational pragmatics in communication aids. *Journal of Pragmatics*, 35(4), 523-538.
- Todman, J., Alm, N., Higginbotham, J., & File, P. (2008). Whole Utterance Approaches in AAC. *Augmentative and Alternative Communication*, 24(3), 235-254.