

Grammar is to Meaning as the Law is to Good Behaviour

Adam Kilgarriff
Lexical Computing Ltd
Lexicography MasterClass Ltd
Universities of Leeds and Sussex

What is grammar?

- We don't need it

fire fire run quick

Chomsky

- Linguistics
 - Distinguish grammatical from ungrammatical
 - He goes to the park
 - *He goes the park
 - “Starred (non-)sentences”
 - *Competence* not performance
- Dominant since 1960s

Enter the *Corpus*

- Sample of language
 - Newspapers
 - Books
 - Internet
 - Transcripts (of conversations, lectures, meetings..)
- Big
 - Lots of texts
- On computer

Rationalists and Empriricists

- Descartes, Leibniz
- Theory
- Competence
- Rules
- Chomsky
- Locke, Hume
- Data
- Performance
- Statistics
- ...Sampson

Sampson

- “Grammar without grammaticality”
- Empiricist
 - What is found in corpora
 - Some things common, others less so, or absent
 - “the concept of 'ungrammatical' or 'ill-formed' is a delusion, based on a false conception of the kind of thing a human language is”

Many cases

- Grammatical ??????
- Native speakers
 - No problem, communicates clearly
- Linguist
 - Infringes a rule
- Sampson
 - “no strangers, only friends I haven't met yet”

Journal: Corpus Linguistics and Linguistic Theory

- Special Issue
- Sampson and responses
- Sampson: mostly stronger case but

***He goes the park.**

***He goes to park the.**

***He go to the park.**

- Sampson
 - No convincing answer
- A response
 - Grammaticality is a partial function
 - Sentences like these
 - “Is it grammatical” is a question with an answer
 - Others
 - It isn't

How useless!

- Great big question in linguistics
- Partial function: non-answer

An analogy

- **Grammar :: meaning**
- **Law :: good behaviour**

- Meaning/communication, good behaviour
 - Good for society
- Grammar/law
 - Systems of rules for promoting
 - Not a necessary condition, but they help

Round the campfire

- Fred has taken Freda's stone axe
- Acceptable behaviour?
- Long and heated discussion...

Next time

- Frieda has taken Bert's stone axe
- Oh no ... discussion takes too long, we'll starve
 - Rules
 - Define 'theft', 'property', borrowing' ...
 - ***Speeds up decision making***
 - ***Reduces social burden***

Lion!

- Ambiguous
 - Past tense
 - we may be able to steal the kill
 - Present tense
 - Run!
- Grammar (verb + tense)
 - ***Speeds up processing***
 - ***Reduces cognitive burden***

System of rules

- Language:
 - Grammar
- Society
 - Law
- Simple cases: process very fast
- Complex cases
 - Rules don't help much
 - Discussion/analysis still needed

System of rules

- Develops a life of its own
 - Medieval Irish grammar
 - Loopholes, laws to fill them, new loopholes ...
- Pressure to rationalise
 - Invading armies
 - immigration

Where the analogy ends

- Law
 - Money
 - Prison
 - So we pay lawyers
- Language
 - Failed communication
 - Repair and fix
 - :-)

Summary

- Grammar is to meaning as the law is to good behaviour
- Systems of rules for handling common cases efficiently
- Complex cases
 - Law: pay a lawyer
 - Language
 - There isn't an answer: *ungrammaticality* irrelevant



History

- Bible studies
- Literary criticism
 - Shakespeare concordance: 200 years ago
- Dictionary-making
 - Samuel Johnson (1754)
 - Oxford English Dictionary



History

- Psychology
 - How do children learn language?
- Education
 - Teaching to read
 - Thorndike and Lorge, 1940s
 - Word lists for teaching
 - Brown corpus, 1960s, 1m words
 - First modern corpus

Corpus and dictionary

- Dictionary
 - High quality but limited in size
 - might not have what you need
- Corpus
 - vast
 - for *when the dictionary does not tell you enough*

Sketch Engine

- Slovene, English, Chinese, other languages
- In use at
 - Ljubljana, OUP, Macmillan, CUP, Collins
 - Many universities
- *Word sketches*
- Instant web corpora
 - WebBootCaT
- Free trial

...pis
 ...pis
 ...vnik,časopis
 ...pis
 ...er,časopis
 ...pis
 ...pis
 ...vnik,časopis
 ...Tedn,časopis
 ...pis
 ...a
 ... ,časopis
 ... ,časopis
 ...vnik,časopis
 ...Dnev,časopis
 ... ,časopis
 ...vnik,časopis

vrstnicami merila na Češkem in Slovaškem, **fantje** pa so trenirali n
 tudi trije mladi športniki iz Črnomlja. **Fantje** , ki jih trenira M
 tekme, ju ločita le še dva meseca, zato **fanta** že trenirata, a s
 Malnarič. Uspeh ni prišel kar sam, saj **fantje** trdo trenirajo po
 pripravljen skozi celotno sezono. **Fantje** , na Rogli jih tre
 izpolnile sanje, o kakršnih sanja vsak deček. **Fant** trenira na strogo
 Saj vem, da se ne pišemo prav, toda moji **fantje** so trdo trenirali
 potem bomo pa garali! ' Dela pa je veliko: **fantje** tudi po dvakrat
 . Klub deluje v skromnih razmerah. **Fantje** trenirajo v telova
 kategoriji in nastopajo z dvema ekipama. **Fantje** radi trenirajo, ne
 2 ... Tomi, pevec Siddharte. **Fant** je svoje dni treni
 nismo bili tako blizu evropskemu zlatu. **Fantje** trenirajo zjutraj,
 na atletski štadion, so mi padli v oči **fantje** , ki so trenirali s
 delom priprav. " Vse poteka po načrtih, **fantje** pa trenirajo zelo
 brez primere. Tudi v naši deželici so **fantje** , ki prav tako zav
 umirimo, « pravi Miran Pavlin. Naši **fantje** so včeraj dopold
 iavno že dali časa. Zato številni mladi **fantje** trdo trenirajo, d

fant/dekle preloaded/fidaplus2-sld freq = 98530/97288

Common patterns

fant 6.0 4.0 2.0 0 -2.0 -4.0 -6.0 dekle

izmed-d_X	88	65	14.3	10.3
biti	22	33	14.0	19.4
med-d_X	772	1733	4.2	9.3
biti	355	1069	31.5	40.3
razlika	26	15	26.4	16.8
ljubezen	14	13	19.6	15.7
zmagati	14	31	11.9	15.1
odnos	9	16	12.3	14.1

iz-d	3733	3180	8.5	7.1
soseščina	125	29	62.0	37.2
okolica	160	51	45.5	29.8
skupina	293	175	40.5	34.2
vas	155	47	39.4	24.4
sosednja	32	29	33.0	32.4
družina	50	106	22.5	32.9
razred	56	47	27.2	26.0
ansambel	27	10	25.9	15.7
predmestje	16	10	25.6	20.6
soseska	9	16	18.3	25.6
država	20	104	7.8	24.1
sosednji	23	11	23.1	16.0
kraj	16	43	12.3	22.7
vasica	13	8	21.1	16.5
ekipa	38	31	20.7	19.3
Ljubljana	44	38	20.4	19.7
reven	9	13	15.5	19.8

osebek_od	31376	23898	7.8	5.8
zreti	656	10	52.9	6.0
jokati	134	48	26.7	16.9
ljubiti	55	160	9.9	22.3
plesati	39	84	11.7	20.8
zaljubiti	51	93	12.2	19.7
trenirati	110	56	19.3	14.2
starati	34	53	13.3	19.0
odraščati	39	41	16.6	18.6
spoznati	116	226	9.6	18.3
ogovoriti	9	20	9.9	17.8
igrati	464	185	17.3	9.9
prepevati	57	25	17.0	11.0
spogledati	14	17	13.8	16.7
očarati	19	38	9.4	16.6
varati	27	11	16.4	10.0
posiliti	15	23	11.3	16.2
obstopiti	14	9	15.7	12.8

nit [18](#) [14](#) 20.6 17.3

"fant" only patterns

izpod-d	162	64.7
Rogla	87	73.5
Vurberk	29	62.7
lipa	8	27.1
vaški	8	24.2
izmed-d_X	88	14.3
eden	42	48.2

iz-d	3733	8.5
Praprotno	13	37.3
paralelka	12	36.2
celuloid	10	34.1
Harlem	10	29.5
Šentlovrenc	8	28.7
Velenje	29	28.7
Hrušica	10	26.0
bend	17	25.7
Luče	10	22.5
Dol	12	22.0
Piran	13	21.9
laboratorij	17	21.5

osebek_od	31376	7.8
slabeti	188	37.6
tepsti	23	12.8
ušpičiti	10	12.8
prebledeti	8	12.6
dvoriti	8	11.6
zacveteti	13	11.5
prikimati	17	11.2
nalivati	8	11.2
poditi	16	10.7
zafrkavati	8	10.6
zlatiti	18	10.6
zapeti2	15	10.5

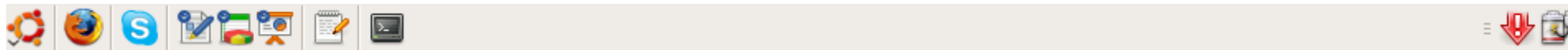
biti_kakšen?	803	6.9
prizadeven	11	27.8
samozavesten	11	26.6
nadarjen	12	25.6
zadovoljen	15	23.8
razpoložen	9	22.3
spreten	8	21.3
kriv	10	20.8
prijazen	10	18.7
hiter	13	17.9
vesel	8	16.8
močen	13	16.6
star	13	13.5

kot-d_X	797	6.2
dekle	60	39.2

komu-čemu	1637	4.8
čestitati	117	42.9

od-d_X	613	4.3
eden	135	57.5

trpnik	3015	3.9
mobilizirati	15	22.4



trenta	10	23.2
Jersey	12	23.6
vzhoden	26	23.2
torta	14	23.2

osebek_od 23898 5.8	komu-čemu 1922 5.5	p-koga-česa 1189 4.4	koga-česa 2207 3.9				
zanositi	50 23.8	dvoriti	18 30.9	mlad	290 54.2	pokvariti	109 40.9
možiti	15 17.3	šepniti	8 18.8	briljanten	19 37.9	vpisovati	13 18.3
omožiti	12 15.5	nagajati	11 18.3	simpatičen	29 36.7	vabiti	31 15.9
vzdihovati	14 15.3	oddati	27 17.9	gol	22 32.0	vročati	10 15.8
hihitati	8 13.0	podarjati	8 14.8	fantov	9 29.4	mrgoleti	8 15.3
obletavati	8 12.8	svetovati	17 13.4	mnog	32 28.3	poljubiti	10 15.1
pozirati	13 12.8	nasmehniti	10 12.8	sv.	25 27.6	prevladovati	15 13.5
sanjariti	14 12.4	upreti	9 11.1	željen	10 23.2	pomeriti	17 13.1
zdrzniti	11 12.2	izročiti	11 10.6	lep	17 17.1	iskati	31 11.8
krikniti	8 11.5	peti	10 8.9	cel	13 16.0	popestrili	10 11.8
odkimati	10 11.1	prinesti	15 8.4	dober	27 15.3	spoznavati	9 11.7
smučati	17 10.0	verjeti	13 8.0	star	17 14.2	obleči	13 11.6

količinski 1844 3.7	biti_kakšen? 436 3.7	veznik 12878 3.0	trpnik 2252 2.8				
pomanjkljivo	16 34.9	lep	21 23.7	en	8 19.6	posiliti	18 24.0
rosno	12 33.1	priden	8 23.4	nek	11 16.5	naličiti	9 19.3
ogromno	15 25.1	srečen	8 19.2	kolikor	8 12.6	razgaliti	11 19.0

Keywords

Corpus: **preloaded/fidaplus2-sld**

Subcorpus: **hopla**

Reference corpus: **preloaded/fidaplus2-sld**

Reference subcorpus: **delo**

lemma	<i>preloaded/fidaplus2-sld: hopla</i>			<i>preloaded/fidaplus2-sld: delo</i>			Score
	Freq	ARF	ARF/mill	Freq	ARF	ARF/mill	
hopla	4858	308.3	240.2	59	32.9	0.9	3.4
am.	4749	264.9	206.4	7	2.0	0.1	3.1
kupon	1745	254.1	198.0	411	167.1	4.7	2.8
Kopitarjev	942	234.0	182.3	70	41.0	1.1	2.8
hoplavrtec	808	223.9	174.4	0	0.0	0.0	2.7
dag	1373	223.5	174.1	369	80.8	2.3	2.7
20.00	2153	254.3	198.1	950	410.9	11.5	2.7
1001	694	218.0	169.8	93	53.4	1.5	2.7
Jennifer	1346	229.9	179.1	589	255.0	7.1	2.6
detektiv	903	228.6	178.1	645	256.1	7.1	2.6
križanka	1308	238.0	185.4	769	372.8	10.4	2.6
zvezdniški	929	230.1	179.3	685	354.8	9.9	2.5
zaljubiti	2081	262.8	204.7	1672	740.8	20.6	2.5
p.	638	212.5	165.6	354	187.8	5.2	2.5
Mimi	771	221.1	172.2	509	288.6	8.0	2.5
žrebanje	743	217.6	169.6	619	307.4	8.6	2.5
izžrebanec	491	191.7	149.4	72	33.7	0.9	2.5
žlica	1093	220.6	171.9	1265	396.1	11.0	2.4

Done