- Heads and tails
- Rank vs frequency
- Type vs token
- Frequency of frequencies
- Frequency of intervals
- Historical trends

Watch out for these terms:

- Heads (of a frequency distribution; also tails)
- Hapax legomena (singular, hapax legomenon; "hapaxes" for short)
- Function words
- Stop words
- Rank
- Zipf's law
- Type-token curve
- Frequency of the frequencies (count of counts)

To review ask yourself these questions:

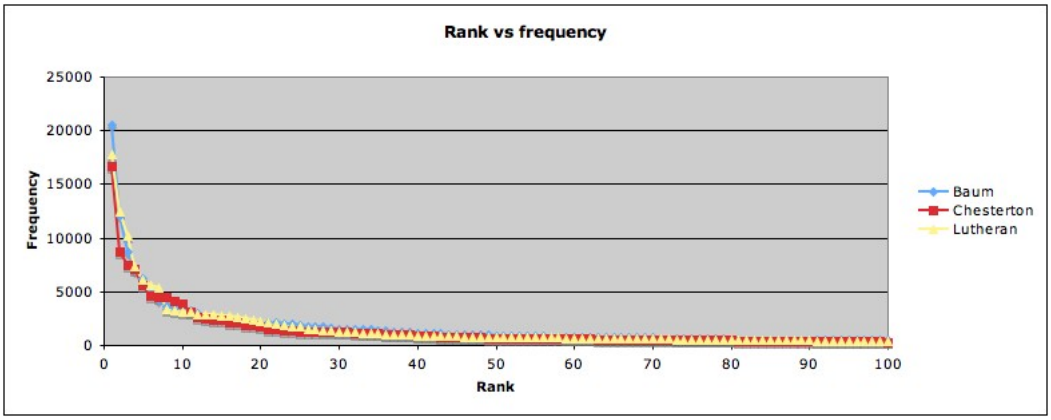
- What does rank vs frequency suggest about the distribution of words?
- What claim did Zipf make about how rank vs frequency exemplifies the principle of least effort?
- What does the type-token curve suggest about a text's lexical richness?
- What does the frequency of the frequencies suggest about a text's lexical richness?
- What does the frequency of the intervals suggest about linguistic "clumpiness" (or "burstiness")?

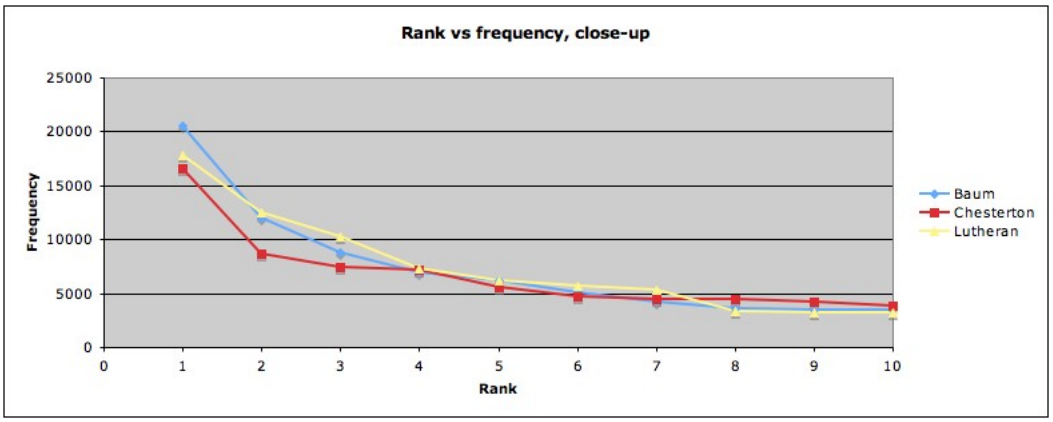
Top 100 types

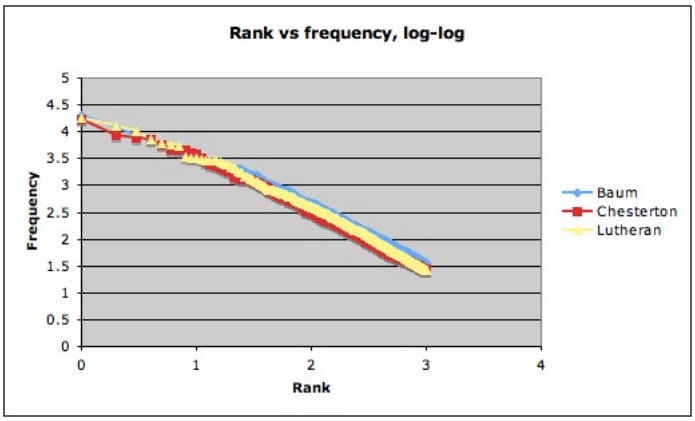
Baum	Chesterton	Lutheran
the 20556	the 16616	the 17827
and 12063	of 8719	and 12480
to 8763	a 7453	of 10247
of 6924	and 7143	to 7351
a 6216	to 5587	that 6144
i 5036	in 4659	in 5639
in 4170	is 4512	is 5299
you 3635	that 4509	not 3317
was 3494	it 4150	for 3249
that 3428	he 3801	it 3157
it 3277	i 3131	god 3126
he 3031	was 2659	we 2897
they 2663	his 2427	as 2860
but 2460	but 2387	be 2844
her 2430	as 2329	this 2795
she 2413	not 2166	a 2713
as 2320	with 2071	are 2592
for 2171	you 1872	they 2444
said 2140	be 1854	but 2385
his 2119	for 1802	by 2181
with 2057	man 1473	which 2135
had 2056	this 1449	christ 1992
so 2035	or 1378	he 1855
is 1983	all 1352	or 1730
all 1927	they 1290	have 1703
s 1765	are 1288	all 1471
not 1721	we 1288	with 1458
at 1692	at 1268	faith 1425
be 1598	had 1261	his 1361
this 1507	have 1260	from 1287
my 1496	on 1210	our 1256
them 1493	which 1132	also 1207
were 1493	if 1129	s 1139
dorothy 1465	one 1093	us 1129
we 1419	said 1061	who 1101
t 1324	an 1008	has 1022
are 1221	there 994	will 1002
me 1202	by 969	one 977
one 1200	like 949	works 975
have 1110	s 899	these 907
on 1110	him 894	do 880
no 1100	who 867	so 832
from 1065	more 807	holy 819
him 1046	has 765	sins 812
then 1029	were 755	them 811
do 1003	only 754	no 795
when 994	from 723	if 794
their 962	so 721	law 781
if 946	about 670	when 773
who 919	can 667	their 758
king 917	what 658	i 754
there 905	will 637	concerning 746
will 878	no 635	word 731
by 862	out 605	on 727
ozma 817	do 602	therefore 713
now 811	when 601	what 702
what 799	would 585	only 701
little 798	than 582	because 689
can 797	some 579	you 673
scarecrow 795	very 571	such 671
your 795	any 569	good 655
could 767	men 562	there 637
would 762	my 543	him 628
which 761	because 528	can 627
oz 757	syme 519	sin 619
tin 756	them 509	been 598
out 739	may 503	church 598
up 694	their 501	should 597
upon 694	could 500	other 595
very 686	up 495	man 593
into 675	even 486	without 578
asked 671	been 484	men 575
us 618	other 478	must 556
good 617	say 477	may 525
or 614	me 462	those 512
see 586	into 458	would 511
man 584	then 454	gospel 510
an 583	t 450	does 497
girl 583	thing 444	at 495
where 559	mr 437	life 492
way 545	things 434	says 490
made 538	great 388	was 479
been 527	world 382	through 474
people 513	must 381	were 472
here 512	our 381	nature 465
well 509	think 378	things 462
time 504	old 365	thus 457
like 502	know 363	even 456
go 498	its 353	upon 445
did 493	something 336	righteousness 437
before 489	really 330	nor 434
over 485	did 329	remission 434
m 484	such 327	doctrine 430
great 480	much 323	true 420
know 477	see 323	before 419
woodman 468	quite 321	now 408
again 457	two 319	grace 406
any 456	us 319	according 404
some 456	most 315	against 402
get 454	might 306	paul 395

Top 100 types excluding stopwords

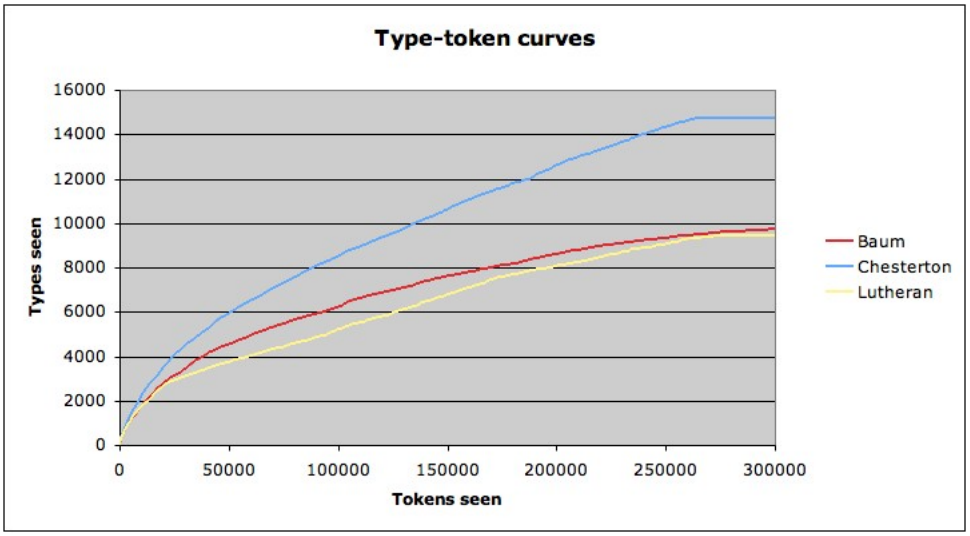
Baum	Chesterton	Lutheran
said 2140	man 1473	god 3126
dorothy 1465	one 1093	christ 1992
one 1200	said 1061	faith 1425
king 917	like 949	also 1207
ozma 817	would 585	us 1129
little 798	men 562	one 977
scarecrow 795	syme 519	works 975
could 767	may 503	holy 819
would 762	could 500	sins 812
oz 757	even 486	law 781
tin 756	say 477	concerning 746
upon 694	thing 444	word 731
asked 671	mr 437	therefore 713
us 618	things 434	good 655
good 617	great 388	sin 619
see 586	world 382	church 598
man 584	must 381	man 593
girl 583	think 378	without 578
way 545	old 365	men 575
made 538	know 363	must 556
people 513	something 336	may 525
well 509	really 330	would 511
time 504	much 323	gospel 510
like 502	see 323	life 492
go 498	quite 321	says 490
m 484	two 319	nature 465
great 480	us 319	things 462
know 477	might 306	thus 457
woodman 468	brown 305	even 456
get 454	little 304	upon 445
magic 452	well 304	righteousness 437
must 449	good 289	remission 434
replied 449	always 282	doctrine 430
head 441	first 279	true 420
away 439	people 274	grace 406
back 426	never 271	according 404
came 416	still 265	paul 395
trot 399	modern 261	believe 393
much 392	father 255	body 390
land 384	face 252	adversaries 377
make 378	rather 247	since 364
long 365	long 240	namely 363
country 359	make 238	love 359
around 358	human 237	lord 357
cap 355	time 236	human 355
never 354	seemed 226	sake 354
ll 345	thought 224	ghost 347
oh 345	life 223	say 347
first 335	right 223	cannot 339
shaggy 325	fact 221	words 332
think 323	made 221	account 326
n 320	whole 220	nothing 325
come 318	cannot 217	power 323
answered 317	almost 215	shall 322
old 315	way 213	many 315
big 313	every 212	teach 312
eyes 312	found 209	divine 310
find 312	black 208	order 306
wizard 312	mean 207	article 304
bill 308	looked 205	first 303
found 307	god 204	heart 302
room 306	head 204	yet 302
looked 301	sense 204	although 300
many 293	upon 200	neither 299
shall 287	christianity 199	christian 295
may 286	nothing 199	eternal 294
betsy 283	come 195	confession 291
queen 280	go 195	ought 291
nome 279	without 194	promise 290
let 275	back 189	work 285
glinda 274	came 189	know 283
place 274	last 189	great 278
sure 272	take 186	necessary 275
lion 266	asked 183	merit 274
princess 266	eyes 182	alone 270
might 259	anything 181	though 262
quite 257	many 181	way 260
ve 257	yet 179	world 251
water 257	matter 178	reason 250
ever 255	also 176	part 249
three 255	another 176	made 248
city 254	new 175	every 246
right 253	place 174	repentance 242
witch 253	shall 173	death 240
woot 252	truth 172	said 240
cried 251	enough 170	rom 239
two 247	certain 163	father 235
friends 244	went 163	truly 234
green 244	find 161	new 231
seemed 243	christian 160	entire 230
another 242	look 160	etc 227
even 239	called 159	time 227
went 236	sort 159	another 223
saw 232	professor 158	given 218
thought 231	everything 156	might 210
take 230	moment 156	let 208
voice 230	though 156	indeed 207
soon 229	let 155	among 205
began 227	kind 153	name 205
em 227	mere 153	commandment 204

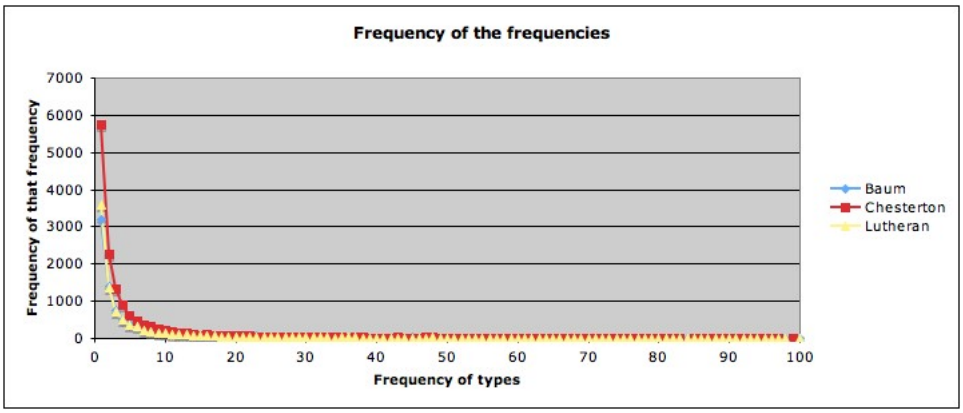


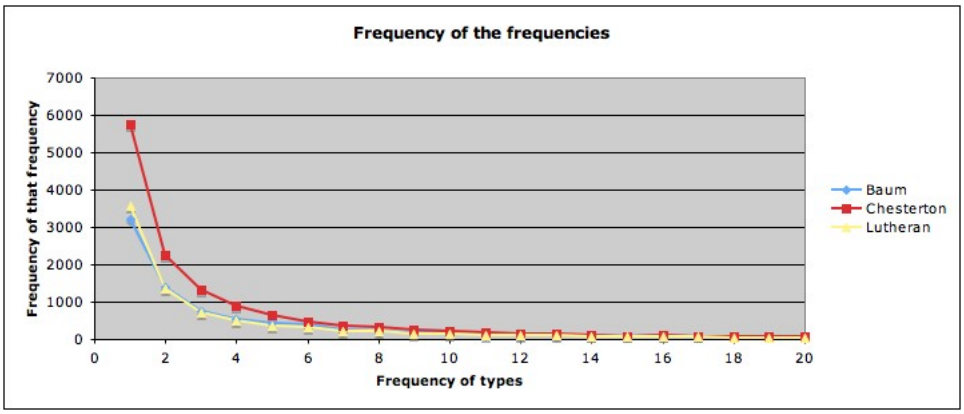


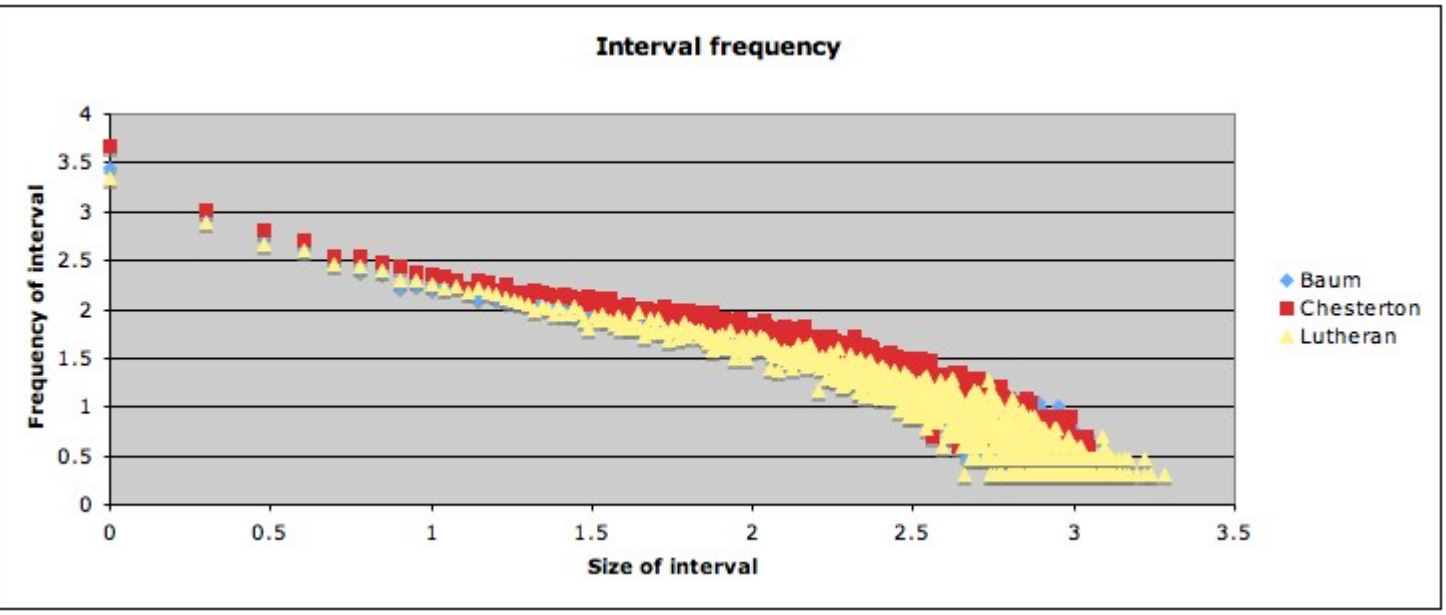


Donning his new canine decoder, Professor Schwartzman becomes the first human being on Earth to hear what barking dogs are actually saying.


Type-token curves
Types seen
Tokens seen
0
2000
4000
6000
8000
10000
12000
14000
16000
0
50000
100000
150000
200000
250000
300000
Baum
Chesterton
Lutheran


Frequency of the frequencies
Frequency of that frequency
Frequency of types
0
1000
2000
3000
4000
5000
6000
7000
0
10
20
30
40
50
60
70
80
90
100
Baum
Chesterton
Lutheran


Frequency of the frequencies
Frequency of that frequency
Frequency of types
0
1000
2000
3000
4000
5000
6000
7000
0
2
4
6
8
10
12
14
16
18
20
Baum
Chesterton
Lutheran


Interval frequency
Frequency of interval
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3
3.5
Size of interval
Baum
Chesterton
Lutheran