

types2

Exploring word-frequency differences
in corpora

Tanja Säily & Jukka Suomela



UNIVERSITY OF HELSINKI



Aalto University

Comparing word frequencies

- Corpus linguists do this all the time
- **Significance** of differences observed?
 - **Bag-of-words** tests (e.g. chi-square, log-likelihood ratio test) assume words occur randomly in texts, overestimate significance
 - Tests based on **resampling**: assumption-free, yield confidence intervals
- *types2*: **permutation testing** (resampling)

Exploring word frequencies

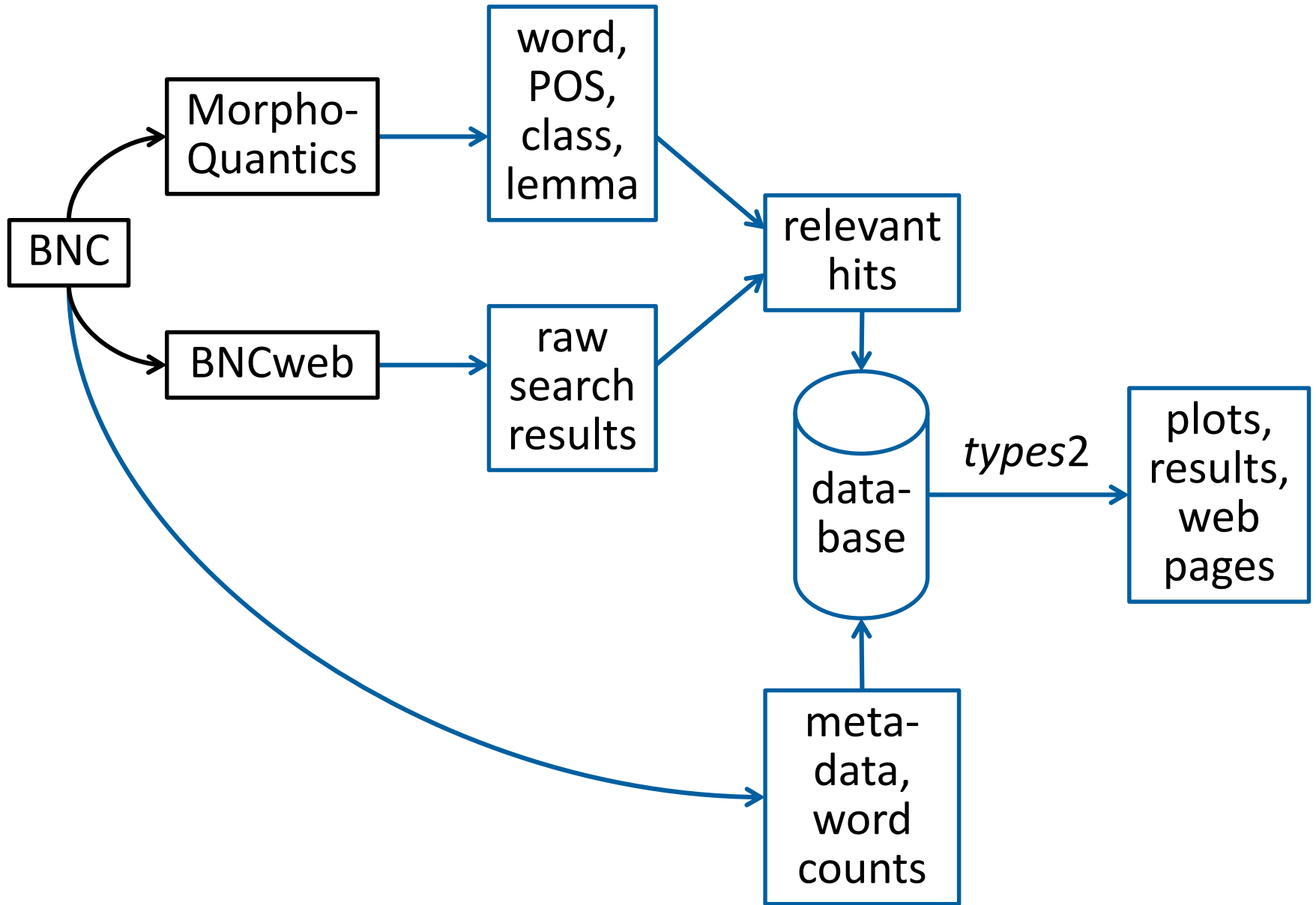
- Typically: static tables, figures
 - Not conducive to rapid exploration
- **Interpretation** of results?
 - Need to go back to the concordances & metadata
- *types2*: online interface with interactive figures, **linked data**

How does it work?

- Do a corpus search
 - BNCweb, WordSmith Tools...
- Narrow down to relevant hits
- **Input** for *types2*: relevant hits + corpus metadata
- **Output**: plots, results, web pages...

Example: Productivity in the BNC

- Derivational suffixes: *-er, -or*
 - **Sociolinguistic variation** in their productivity?
 - Productivity ~ type frequency
- BNC = British National Corpus
 - Demographically sampled spoken component, both gender & social class known: 2.6 Mw
 - **BNCweb** (Lancaster University)
 - **MorphoQuantics** (Laws & Ryder 2014)



Corpora

bnc-spoken-demo

Findings — false discovery rate 0.01

Female	–	types/running words	er	bnc-spoken-demo	0.00001
Female	–	types/running words	er-thing	bnc-spoken-demo	0.00004
Female	–	tokens/running words	er-person	bnc-spoken-demo	0.00006
Male	+	tokens/running words	er-person	bnc-spoken-demo	0.00006

Findings — false discovery rate 0.1

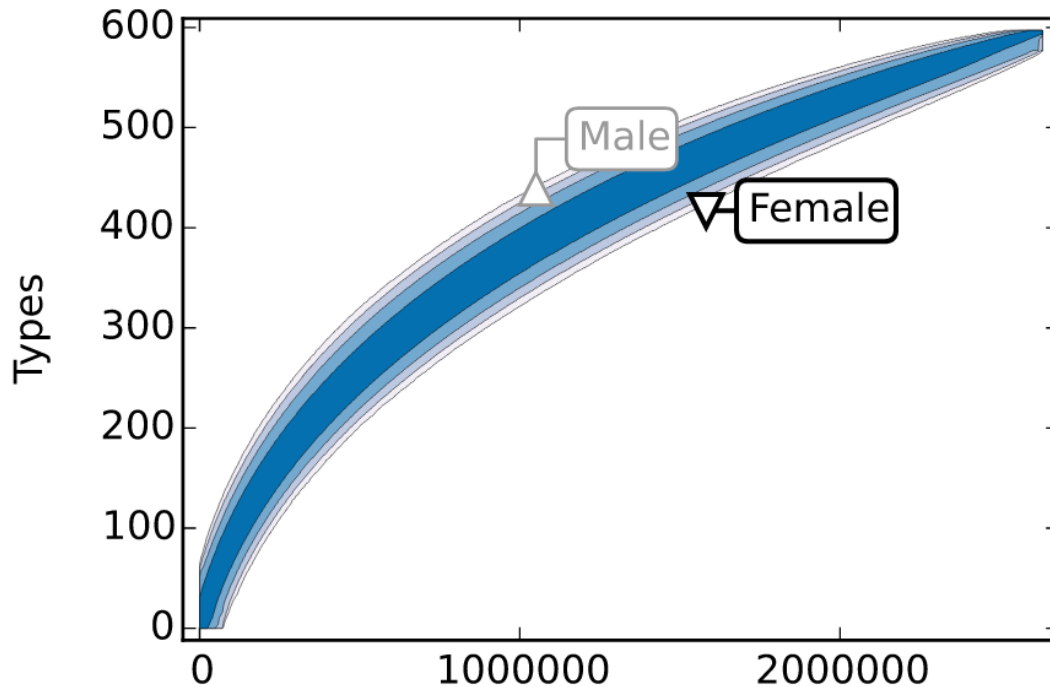
Male	+	types/running words	er	bnc-spoken-demo	0.00022
Male	+	types/running words	er-thing	bnc-spoken-demo	0.00030
Female	–	tokens/running words	er	bnc-spoken-demo	0.00041
Male	+	tokens/running words	er	bnc-spoken-demo	0.00041
Female	–	types/tokens	er-thing	bnc-spoken-demo	0.00115
Female	–	types/running words	er-person	bnc-spoken-demo	0.00239
AB+C1	–	tokens/running words	er-thing	bnc-spoken-demo	0.00377
C2+DE	+	tokens/running words	er-thing	bnc-spoken-demo	0.00378
Male	+	types/tokens	er-thing	bnc-spoken-demo	0.00466

Corpus:

Dataset: [er-person](#) [er-thing](#) [or](#) [or-person](#) [or-thing](#)

Points: none social class

Axes: [hapaxes/tokens](#) [hapaxes/running words](#) [tokens/running words](#) [types/tokens](#)



Collection: [none](#) [Male](#)

Show: [type list](#) [sample list](#)

This collection contains 1582116 running words and 417 types.

Only approx. **0.001%** of random collections with 1582116 running words contain at most 417 types.

This finding is probably interesting: the false discovery rate is **smaller than 0.01**.

0

1000000

2000000

Collection: [none](#) [Female](#) [Male](#)

Show: [summary](#) [type list](#) [sample list](#)

This collection contains 1582116 running words and 417 types.

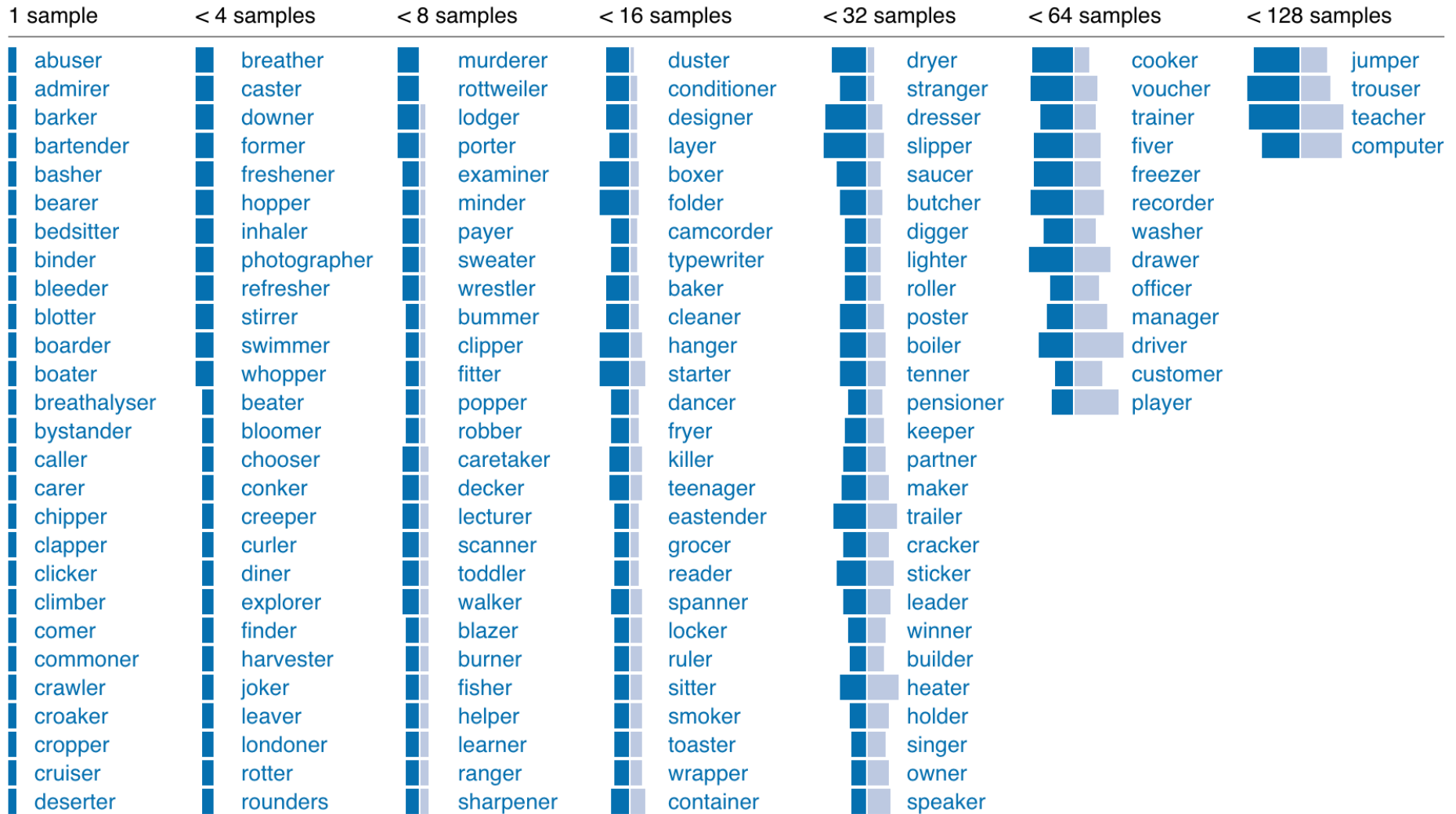
Only approx. **0.001%** of random collections with 1582116 running words contain at most 417 types.

This finding is probably interesting: the false discovery rate is **smaller than 0.01**.

Collection: [none](#) Female Male

Show: [summary](#) type list [sample list](#)

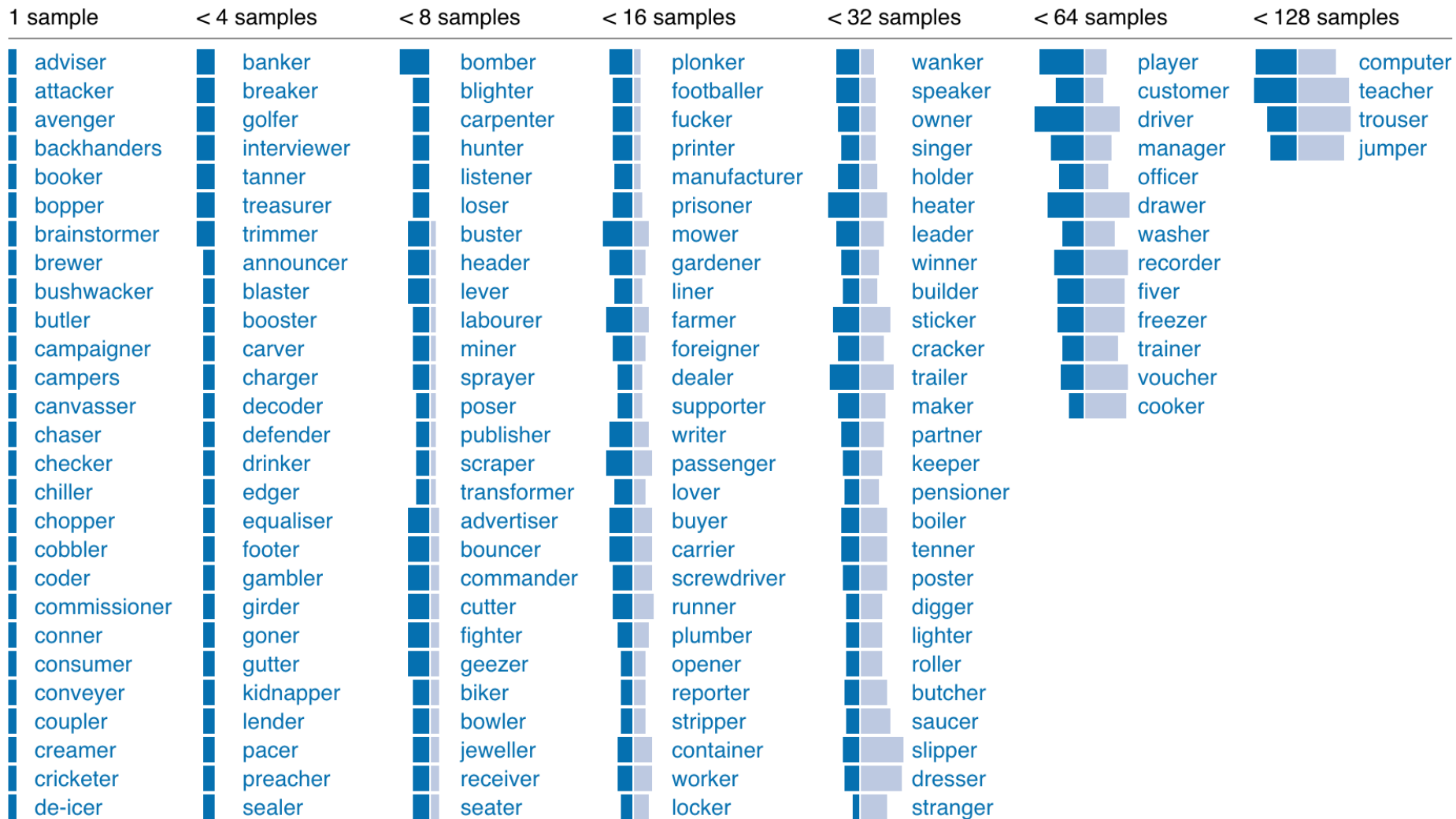
Types that occur in...



Collection: [none](#) [Female](#) [Male](#)

Show: [summary](#) [type list](#) [sample list](#)

Types that occur in...

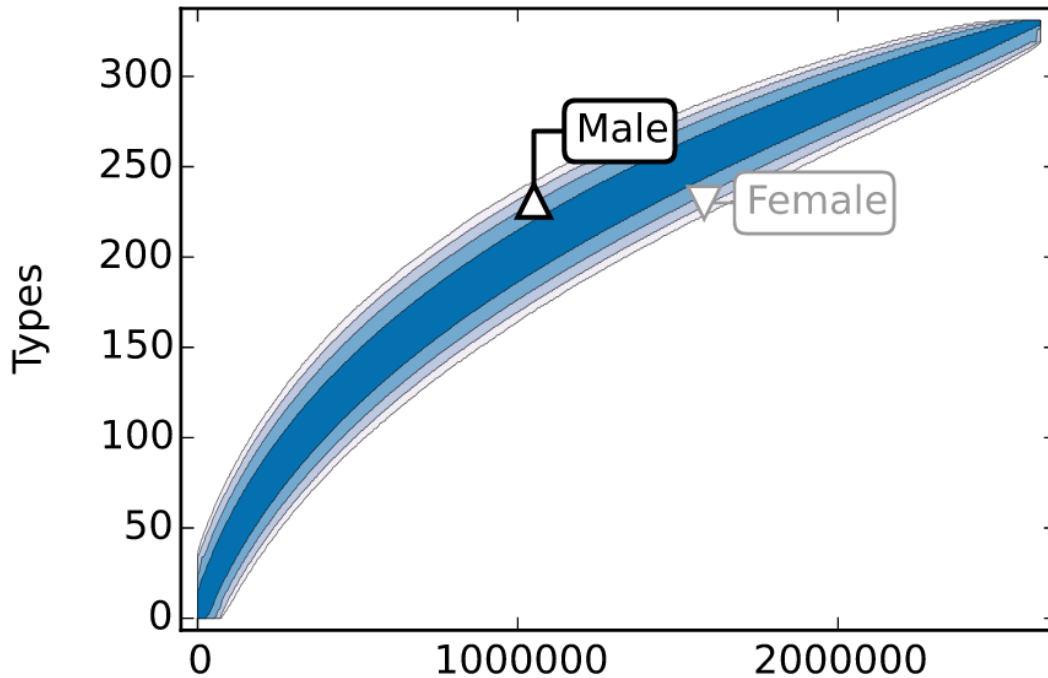


Corpus:

Dataset: or

Points: none social class

Axes:



Collection: none

Show:

Types that occur in...

1 sample

< 4 samples

< 8 samples

< 16 samples

< 32 samples

< 64 samples

< 128 samples

adviser

banker

blighter

plonker

wanker

player

...

Types that occur in...

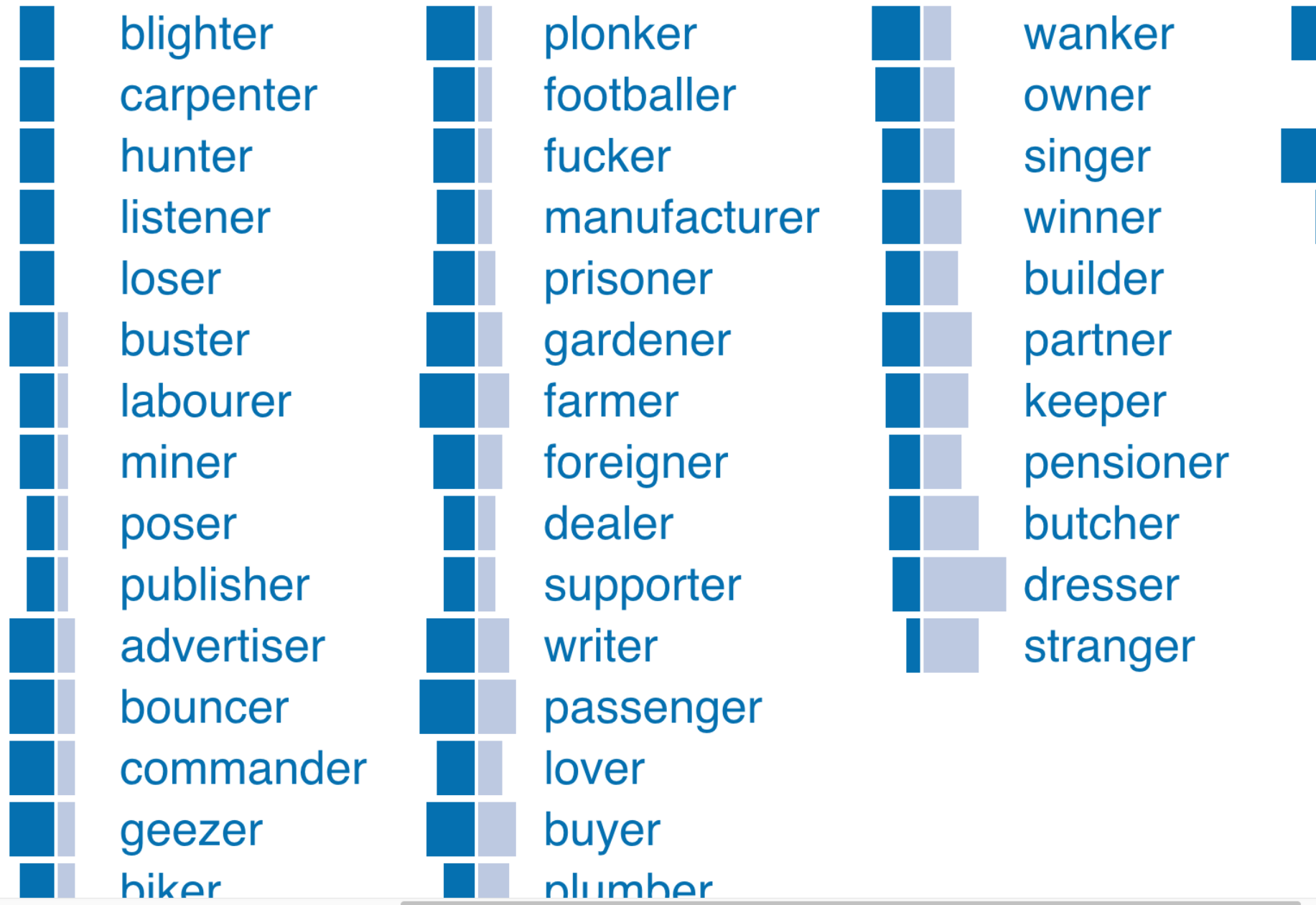


< 8 samples

< 16 samples

< 32 samples

<























































WANKER (NN1) — Male

Sample	Token	Before	Word	After
PS03W	AB Male 35-44	wanker	Do him you	wanker !
PS0JL	C2 Male 25-34	wanker	He were a right	wanker !
PS0TU	AB Male 15-24	wanker	Total	wanker
PS0TU	AB Male 15-24	wanker	And I went fuck you ,	wanker !
PS0TU	AB Male 15-24	wanker	No , he 's a total	wanker !
PS0TU	AB Male 15-24	wanker	He 's a total	wanker .
PS0TU	AB Male 15-24	wanker	Some	wanker in our school paid ten quid for one of those .
PS0TU	AB Male 15-24	wanker	You 'll end up with some real	wanker like .
PS0TU	AB Male 15-24	wanker	It 's just tape it and then some	wanker sits and tries to work it out .
PS0TU	AB Male 15-24	wanker	you talking and then like , but he could still be an old	wanker sure
PS1GE	C2 Male 25-34	wanker	they 're fucking	wankers !
PS1GE	C2 Male 25-34	wanker	I ca n't get out of that car and say , you 're a fucking	wanker I co the only way I could do , the only way
PS1GE	C2 Male 25-34	wanker	Stuart the	wanker [gap:name] .
PS1GE	C2 Male 25-34	wanker	It 's only cos you 're a	wanker is n't he ?
PS1GF	DE Male 15-24	wanker		Wanker it come out winker .
PS1GF	DE Male 15-24	wanker	but if they are	wankers nothing you do , and including , I mean I 've done
PS1GF	DE Male 15-24	wanker	ople use to come out and fucking say to me , fucking	wanker up there , I wonder if he 's gon na jump , [unclear]
PS1GG	DE Male 15-24	wanker	[unclear] went to call someone a	wanker
PS1GG	DE Male 15-24	wanker	Bunch of	wankers ?
PS532	C1 Male -14	wanker	Cos yo , the one in the tie was a right	wanker !
PS532	C1 Male -14	wanker	No , you 're a	wanker on this int ya ?
PS532	C1 Male -14	wanker	You were a fucking	wanker were n't ya ?
PS53C	C1 Male -14	wanker	He 's a fat	wanker as well .
PS53C	C1 Male -14	wanker	He sounds like a such a	wanker when he does it though .
PS59U	C1 Male -14	wanker	and I thought , right , this bloke 's a	wanker

PS1GE	C2 Male 25-34	wanker	're a fucking	wanker
PS1GE	C2 Male 25-34	wanker	Stuart the	wanker
PS1GE	C2 Male 25-34	wanker	cos you 're a	wanker
PS1GF	DE Male 15-24	wanker		Wanker
PS1GF	DE Male 15-24	wanker	out if they are	wankers
PS1GF	DE Male 15-24	wanker	me , fucking	wanker
PS1GG	DE Male 15-24	wanker	ll someone a	wanker
PS1GG	DE Male 15-24	wanker	Bunch of	wankers
PS532	C1 Male -14	wanker	e was a right	wanker
PS532	C1 Male -14	wanker	No , you 're a	wanker
PS532	C1 Male -14	wanker	ere a fucking	wanker
PS53C	C1 Male -14	wanker	He 's a fat	wanker
PS53C	C1 Male -14	wanker	like a such a	wanker
PS59U	C1 Male -14	wanker	his bloke 's a	wanker

Show: [summary](#) [type list](#) [sample list](#)

Sample		Words	Tokens	Types	Unique	Hapaxes	Common types	Unique types
PS0W2	AB Male 60-	41114	30	20	2	1	teacher × 4, player × 3	bowler, adviser
PS50T	C1 Male 35-44	37880	28	16	3	3	manager × 5, farmer × 3, teacher × 3	conveyer, prattler, wanderer
PS03S	C2 Male 60-	36388	51	34	12	11	teacher × 7, painter × 3	quaker, attacker, brewer, mover, poacher, skateboarder, stitcher, voter, whistler
PS0PN	C1 Male 45-59	36385	20	14	2	1	miner × 3, teacher × 3	restorer, zealander
PS03W	AB Male 35-44	32952	36	19	1	1	player × 8, customer × 3, punter × 3	upholsterer
PS05X	AB Male 45-59	27247	39	23	4	3	customer × 7, teacher × 5	commissioner, bopper, punter
PS0HM	C1 Male 35-44	25943	37	16	2	2	teacher × 9, player × 8	goalgetter, passer
PS0FS	C2 Male 35-44	25797	13	11	2	2	manager × 2, worker × 2	avenger, packer
PS02H	DE Male 25-34	24621	25	17	2	2	teacher × 3	no-hoper, taker
PS1BY	AB Male 15-24	24521	16	13	3	3	partner × 3, teacher × 2	jogger, leader, prowler
PS0V4	C1 Male 35-44	24451	15	11	1	1	partner × 3, officer × 2, pensioner × 2	chaser
PS1HH	DE Male 45-59	24164	23	17	2	2	interviewer × 3	conner, promoter
PS1GE	C2 Male 25-34	23488	36	21	1	0	wanker × 4	pretender
PS1GF	DE Male 15-24	22871	26	15	2	2	driver × 5, wanker × 3	butler, snatcher
PS0JX	C1 Male 25-34	19701	16	15	2	2	manufacturer × 2	consumer, mender
PS0H9	C2 Male 35-44	17768	31	15	3	3	dealer × 7, customer × 4, player × 4	coder, repairer, strangler
PS065	DE Male 60-	16050	18	13	2	2	teacher × 3	shifter, waster
PS0GG	DE Male 25-34	15263	22	6	1	1	player × 11, loser × 5	campaigner
PS0F8	C2 Male 45-59	15223	9	6	0	0	driver × 3, advertiser × 2	
PS0KN	DE Male 15-24	14805	21	14	2	2	ripper × 3	booker, hacker
PS0HN	C1 Male -14	14741	10	7	1	1	player × 2, singer × 2, supporter × 2	skier
PS01A	DE Male 45-59	13306	4	3	1	1	builder × 2	hitter
PS089	AB Male -14	12916	2	2	0	0		
PS0JL	C2 Male 25-34	11997	9	9	2	2		discounter, stalker
PS1BT	AB Male 45-59	11813	13	10	0	0	plonker × 3, pensioner × 2	
PS09T	DE Male 60-	10939	8	7	2	2	pensioner × 2	campers, canvasser
PS0N4	C1 Male 45-59	10592	39	6	0	0	customer × 32, punter × 3	
PS0FR	C2 Male -14	10574	7	5	1	1	driver × 3	bushwacker

Sample		Words	Tokens	Types	Unique
PS0W2	AB Male 60-	41114 	30 	20 	2 
PS50T	C1 Male 35-44	37880 	28 	16 	3 
PS03S	C2 Male 60-	36388 	51 	34 	12 
PS0PN	C1 Male 45-59	36385 	20 	14 	2 
PS03W	AB Male 35-44	32952 	36 	19 	1 
PS05X	AB Male 45-59	27247 	39 	23 	4 
PS0HM	C1 Male 35-44	25943 	37 	16 	2 
PS0FS	C2 Male 35-44	25797 	13 	11 	2 
PS02H	DE Male 25-34	24621 	25 	17 	2 
PS1BY	AB Male 15-24	24521 	16 	13 	3 
PS0V4	C1 Male 35-44	24451 	15 	11 	1 
PS1HH	DE Male 45-59	24164 	23 	17 	2 
PS1GE	C2 Male 25-34	23488 	36 	21 	1 

Common types

teacher × 4, player × 3
manager × 5, farmer × 3, teacher × 3
teacher × 7, painter × 3

miner × 3, teacher × 3
player × 8, customer × 3, punter × 3
customer × 7, teacher × 5
teacher × 9, player × 8
manager × 2, worker × 2
teacher × 3
partner × 3, teacher × 2
partner × 3, officer × 2, pensioner × 2
interviewer × 3
wanker × 4

Unique types

bowler, adviser
conveyer, prattler, wanderer
quaker, attacker, brewer, dodge
mover, poacher, skateboarder,
stitcher, voter, whistler
restorer, zealander
upholsterer
commissioner, bopper, puffer, t
goalgetter, passer
avenger, packer
no-hoper, taker
jogger, leader, prowler
chaser
conner, promoter
pretender

PS03S: C2 Male 60-

Sample	Token	Before	Word	After
PS03S	C2 Male 60- attacker	You know , it 's the er	attacker	.
PS03S	C2 Male 60- brewer	like the bit , the	brewers	do
PS03S	C2 Male 60- builder	When we were talking to the	builders	tha he were putting that in .
PS03S	C2 Male 60- butcher	The	butcher	.
PS03S	C2 Male 60- butcher	The	butchers	.
PS03S	C2 Male 60- dancer	Di , was Madge a good	dancer	?
PS03S	C2 Male 60- dancer	oyed though , Phil and George were magnificent	dancers	and the chap I used to walk out with sometime
PS03S	C2 Male 60- dodger	res that majority of missing vouchers are poll tax	dodgers	
PS03S	C2 Male 60- driver	what we call	drivers	
PS03S	C2 Male 60- driver	I 've got plenty of	drivers	like this
PS03S	C2 Male 60- eater	Was your father a good	eater	?
PS03S	C2 Male 60- farmer	four i , about four o'clock and er there was three	farmers	round a big sign , in in there and tha that 's wf
PS03S	C2 Male 60- farmer	And er	farmers	used to throw his cabbage plants in the bottor
PS03S	C2 Male 60- footballer	One of the	footballers	did the same as Evelyn .
PS03S	C2 Male 60- footballer	A	footballer	done the same thing .
PS03S	C2 Male 60- founder	the next then we moved from Archlin Lane to	Founder	Lane , where next Christmas I had Meccano t
PS03S	C2 Male 60- jeweller	, carry all the bags , and er we 're looking in the	jewellers	' shop and some old , some lady stood there ε
PS03S	C2 Male 60- jeweller	since you done it we 've been looking in	jewellers	' shops when I go , I go up town twice a week
PS03S	C2 Male 60- keeper	n most things , I mean now , er none of the shop	keepers	are sa , like they were very helpful in our day '
PS03S	C2 Male 60- lodger	Cos the whole object of the exercise is to have a	lodger	is to pay yo your way .
PS03S	C2 Male 60- maker	machinery catalogue which he examined with a	watchmaker	's glass .
PS03S	C2 Male 60- manaaer	He come across with his un one of the	manaaers	and he savs what 's the idea ?

Speaker information for PS03S (of file KBB)

Name:	Arthur
Number of words:	36,388
Number of turns:	3,980
Average length of turns (<w>-tags per <u>-unit):	9.1427

Standard header information:

Sex:	Male
Age:	60+
Social Class:	C2
Education:	n/a
First Language:	n/a
Dialect/Accent:	Central Midlands

Additional information:

Age:	82
Occupation:	retired (precision engineer)
Role:	self

PS03S: C2 Male 60-

Sample	Token	Before	Word	After
PS03S	C2 Male 60- attacker	You know , it 's the er	attacker	.
PS03S	C2 Male 60- brewer	like the bit , the	brewers	do
PS03S	C2 Male 60- builder	When we were talking to the	builders	tha he were putting that in .
PS03S	C2 Male 60- butcher	The	butcher	.
PS03S	C2 Male 60- butcher	The	butchers	.
PS03S	C2 Male 60- dancer	Di , was Madge a good	dancer	?
PS03S	C2 Male 60- dancer	oyed though , Phil and George were magnificent	dancers	and the chap I used to walk out with sometime
PS03S	C2 Male 60- dodger	res that majority of missing vouchers are poll tax	dodgers	
PS03S	C2 Male 60- driver	what we call	drivers	
PS03S	C2 Male 60- driver	I 've got plenty of	drivers	like this
PS03S	C2 Male 60- eater	Was your father a good	eater	?
PS03S	C2 Male 60- farmer	four i , about four o'clock and er there was three	farmers	round a big sign , in in there and tha that 's wf
PS03S	C2 Male 60- farmer	And er	farmers	used to throw his cabbage plants in the bottor
PS03S	C2 Male 60- footballer	One of the	footballers	did the same as Evelyn .
PS03S	C2 Male 60- footballer	A	footballer	done the same thing .
PS03S	C2 Male 60- founder	the next then we moved from Archlin Lane to	Founder	Lane , where next Christmas I had Meccano t
PS03S	C2 Male 60- jeweller	, carry all the bags , and er we 're looking in the	jewellers	' shop and some old , some lady stood there ε
PS03S	C2 Male 60- jeweller	since you done it we 've been looking in	jewellers	' shops when I go , I go up town twice a week
PS03S	C2 Male 60- keeper	n most things , I mean now , er none of the shop	keepers	are sa , like they were very helpful in our day '
PS03S	C2 Male 60- lodger	Cos the whole object of the exercise is to have a	lodger	is to pay yo your way .
PS03S	C2 Male 60- maker	machinery catalogue which he examined with a	watchmaker	's glass .
PS03S	C2 Male 60- manaaer	He come across with his un one of the	manaaers	and he savs what 's the idea ?



File info for KBB



Go!

Show POS-tags

Colour wordclass



Your browser appears to be Chrome. The functionality of the audio features is likely to be limited. Firefox and Safari do limitation.

[Evelyn](#) 1642 **Drink** your tea then.

[Arthur](#) 1643 Thrush, they're not fighters but they're, they're, not cowards they'll <pause> you know, have a go.

[Evelyn](#) 1644 Who, the thrushes?

1645 I've never seen them.

[Arthur](#) 1646 They do, he lowers his wings and crouches.

1647 You know, it's the er <pause> attacker.

[Evelyn](#) 1648 They've eat all that seed you <pause> you're going to run out before we get <pause> <-|-> another <-|->

[Arthur](#) 1649 <-|-> Well <-|->

[Evelyn](#) 1650 lot.

[Arthur](#) 1651 At least they'll die happy won't they?

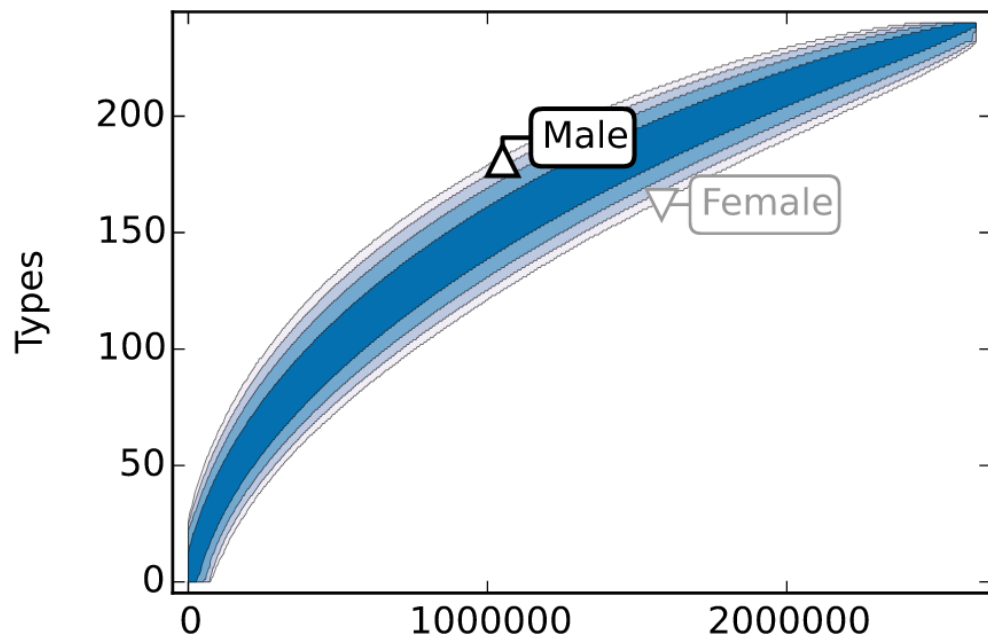
[Evelyn](#) 1652 <voice quality: laughing>I don't know<end of voice quality>!

Corpus:

Dataset:

Points:

Axes:



Collection:

Show:

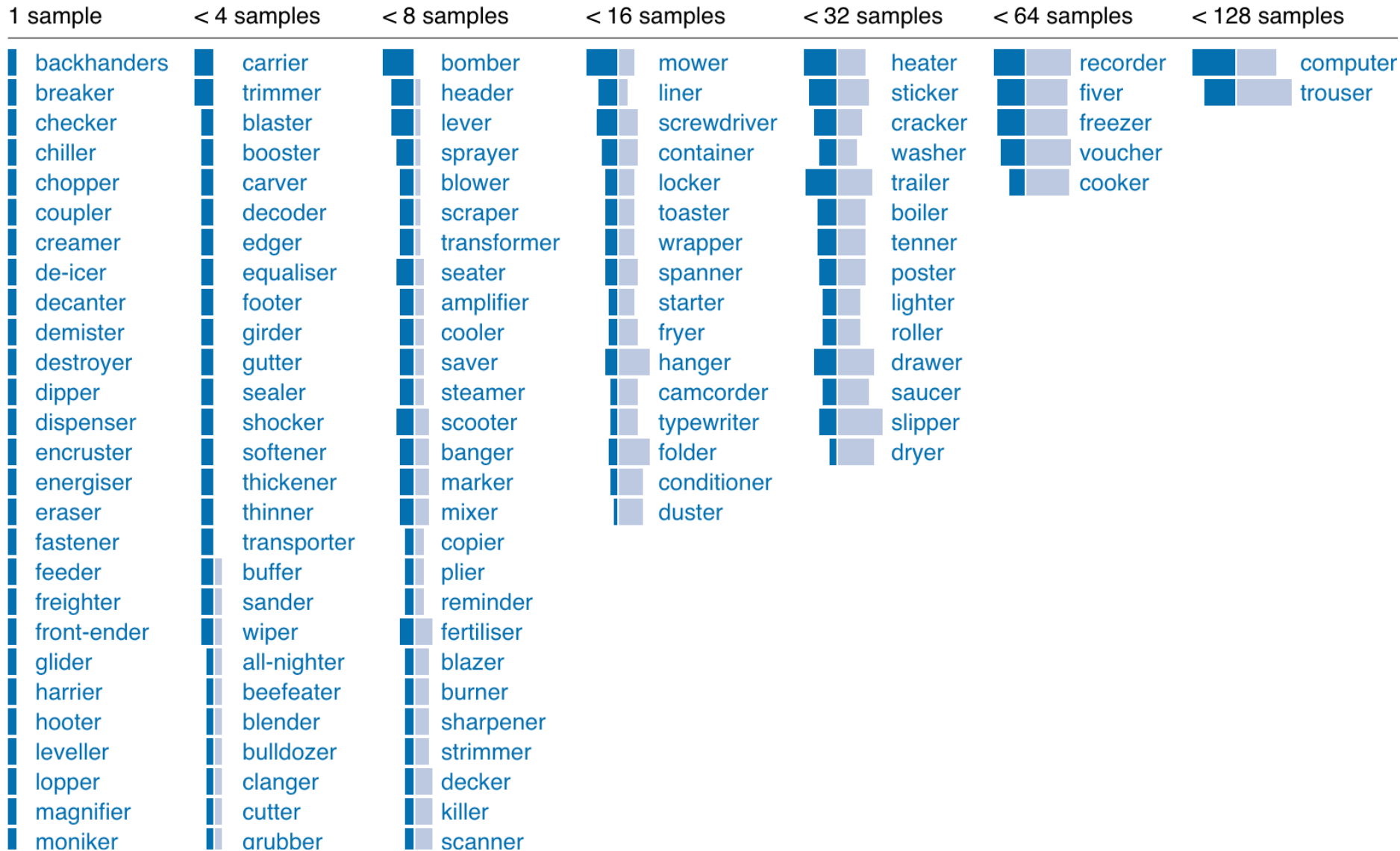
Types that occur in...

1 sample	< 4 samples	< 8 samples	< 16 samples	< 32 samples	< 64 samples	< 128 samples
<input type="checkbox"/> backhanders <input type="checkbox"/> breaker	<input type="checkbox"/> carrier <input type="checkbox"/> trimmer	<input type="checkbox"/> bomber <input type="checkbox"/> header	<input type="checkbox"/> mower <input type="checkbox"/> liner	<input type="checkbox"/> heater <input type="checkbox"/> sticker	<input type="checkbox"/> recorder <input type="checkbox"/> fiver	<input type="checkbox"/> computer <input type="checkbox"/> trouser

Collection: [none](#) [Female](#) [Male](#)

Show: [summary](#) [type list](#) [sample list](#)

Types that occur in...



1 sample

< 4 samples

< 8 samples

< 16 samples

backhanders
breaker
checker
chiller
chopper
coupler
creamer
de-icer
decanter
demister
destroyer
dipper
dispenser
encruster
energiser

carrier
trimmer
blaster
booster
carver
decoder
edger
equaliser
footer
girder
gutter
sealer
shocker
softener
thickener

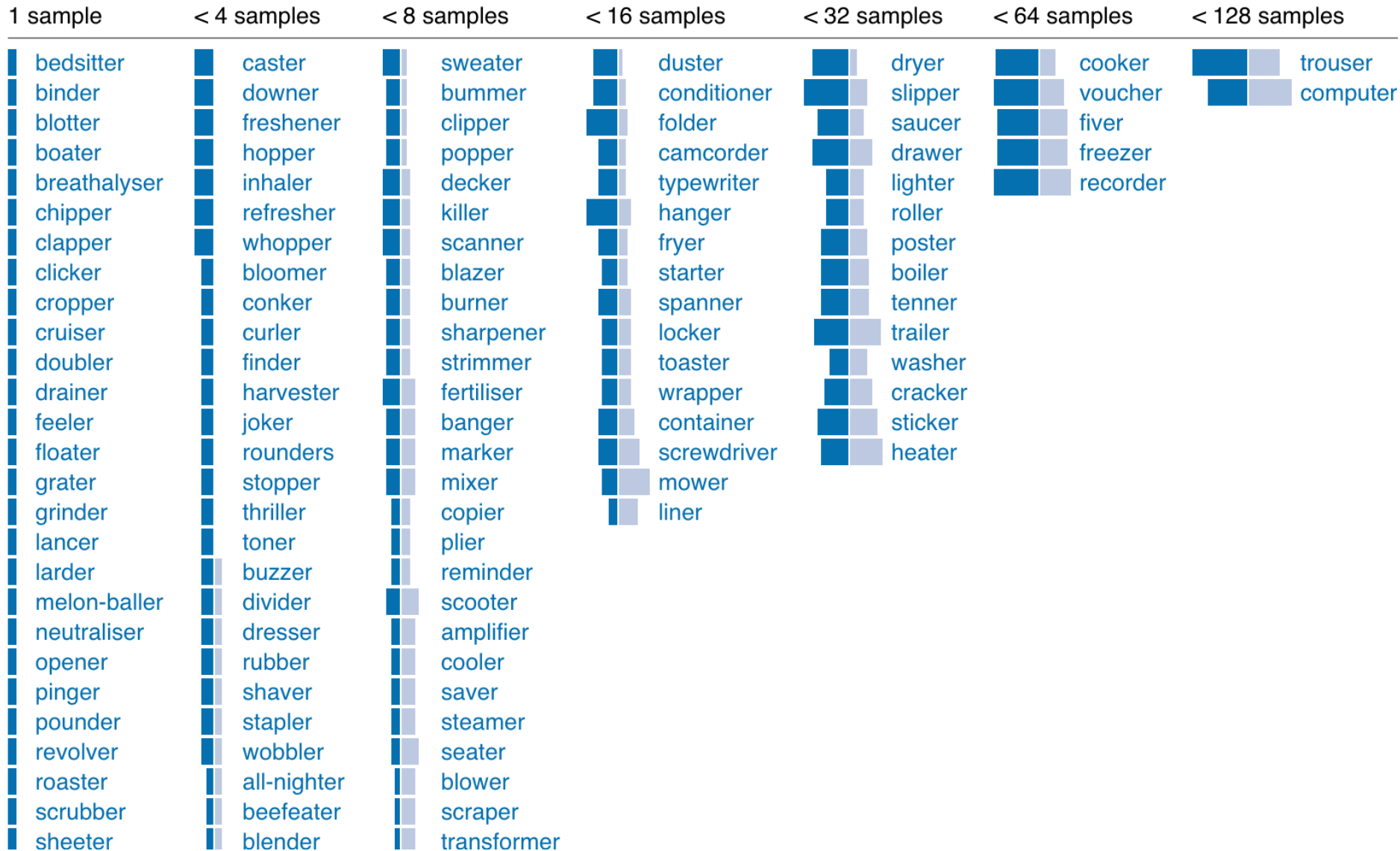
bomber
header
lever
sprayer
blower
scraper
transformer
seater
amplifier
cooler
saver
steamer
scooter
banger
marker

mower
liner
screwdriver
container
locker
toaster
wrapper
spanner
starter
fryer
hanger
camcorder
typewriter
folder
conditioner

Collection: [none](#) Female Male

Show: [summary](#) type list [sample list](#)

Types that occur in...

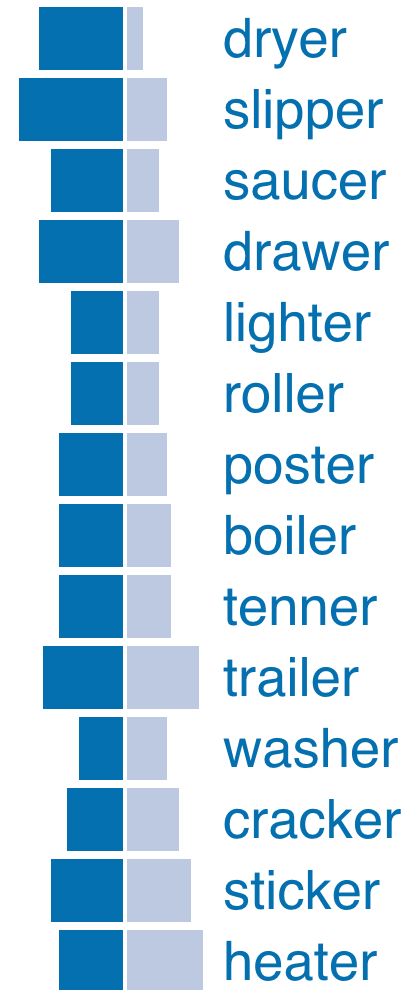
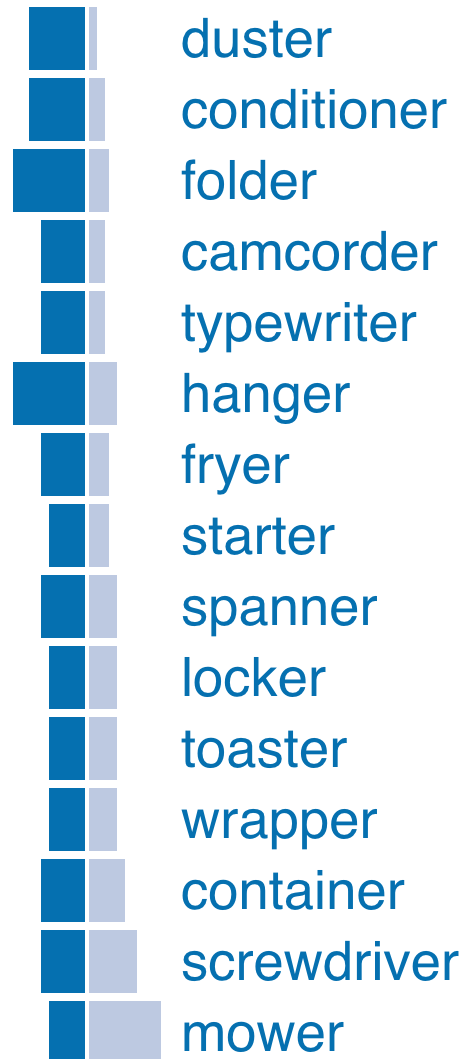


< 8 samples

< 16 samples

< 32 samples

< 64 samples



Conclusion

- Linked data helps with
 - Generating hypotheses
 - Age? Setting? Relationship?
 - Interpreting results
 - Male overuse of *-er*: playful name-calling, focus on tools & occupations?
- *types2*: free tool, works with multiple corpora & concordancers

References

- Laws, J.V. & C. Ryder (2014) *MorphoQuantics*.
<http://morphoquantics.co.uk>
- Laws, J.V. & C. Ryder (2014) Getting the measure of derivational morphology in adult speech: A corpus analysis using *MorphoQuantics*. *Language Studies Working Papers*: University of Reading, Vol. 6, 3–17.
- Suomela, J. (2014) *types2*: Type and hapax accumulation curves.
<http://users.ics.aalto.fi/suomela/types2/>
- Suomela, J. (2015) *bnc-affix*: Analysing productivity of affixes with BNC & MorphoQuantics data.
<https://github.com/suomela/bnc-affix>

Example: Productivity in the CEEC

- Derivational suffixes: *-ness, -ity*
 - **Sociolinguistic variation** in their productivity?
- CEEC = Corpora of Early English Correspondence
 - Long 18th century, 1680–1800: 2.2 Mw
 - **WordSmith Tools** (Mike Scott)
 - Pruned down to relevant hits in **Excel**

Corpora

ceec-1680-1800	CEEC & CEECE, 1680-1800, sample = letters with a certain sender & period
ceec-1680-1800-relcode	CEEC & CEECE, 1680-1800, sample = letters with a certain sender & period & relcode

Findings — false discovery rate 0.001

1680-1719	–	types/tokens	ity	ceec-1680-1800	0.00000
1680-1719	–	types/running words	ity	ceec-1680-1800	0.00000

Findings — false discovery rate 0.01

M, TC	+	types/running words	ity	ceec-1680-1800-relcode	0.00006
-----------------------	---	---------------------	-----	------------------------	---------

Findings — false discovery rate 0.1

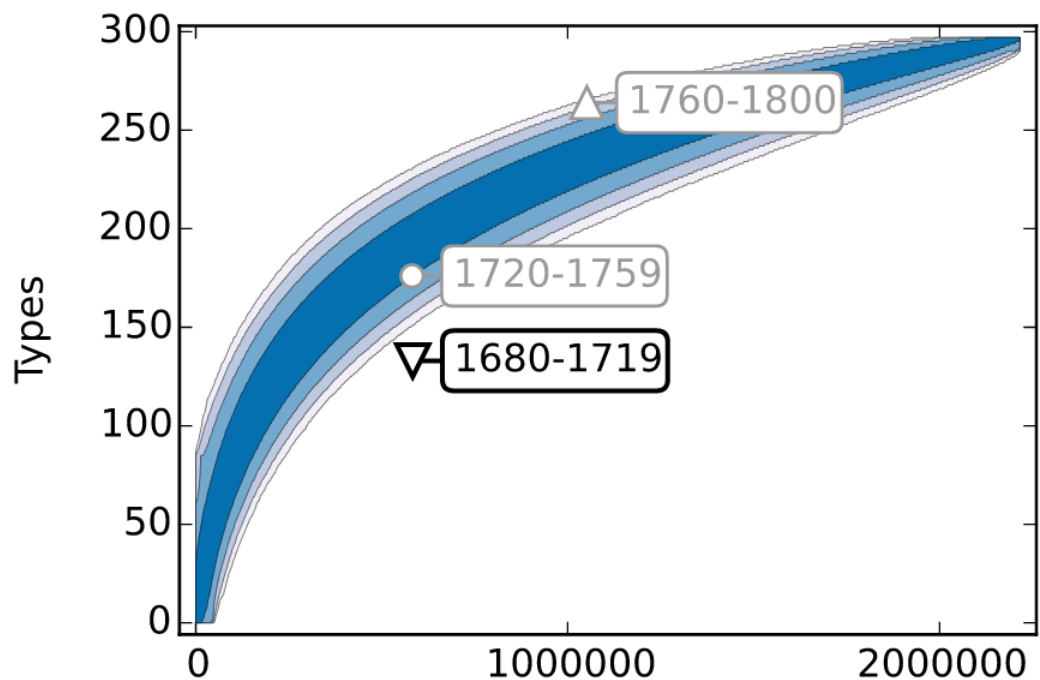
1680	–	types/tokens	ity	ceec-1680-1800	0.00028
1680	–	types/running words	ity	ceec-1680-1800	0.00030
1760-1800	+	types/running words	ity	ceec-1680-1800	0.00030
M, TC	+	types/tokens	ity	ceec-1680-1800-relcode	0.00039
1700	–	types/running words	ity	ceec-1680-1800	0.00057
rank R	–	types/tokens	ness	ceec-1680-1800	0.00140
1760-1800	+	types/tokens	ity	ceec-1680-1800	0.00219
TC	+	types/running words	ity	ceec-1680-1800-relcode	0.00270

Corpus: ceec-1680-1800-relcode

Dataset: [ness](#)

Points: none education gender period rank-current socmob

Axes: [types/tokens](#)



Collection: [none](#) [1720-1759](#) [1760-1800](#)

Written in 1680-1719

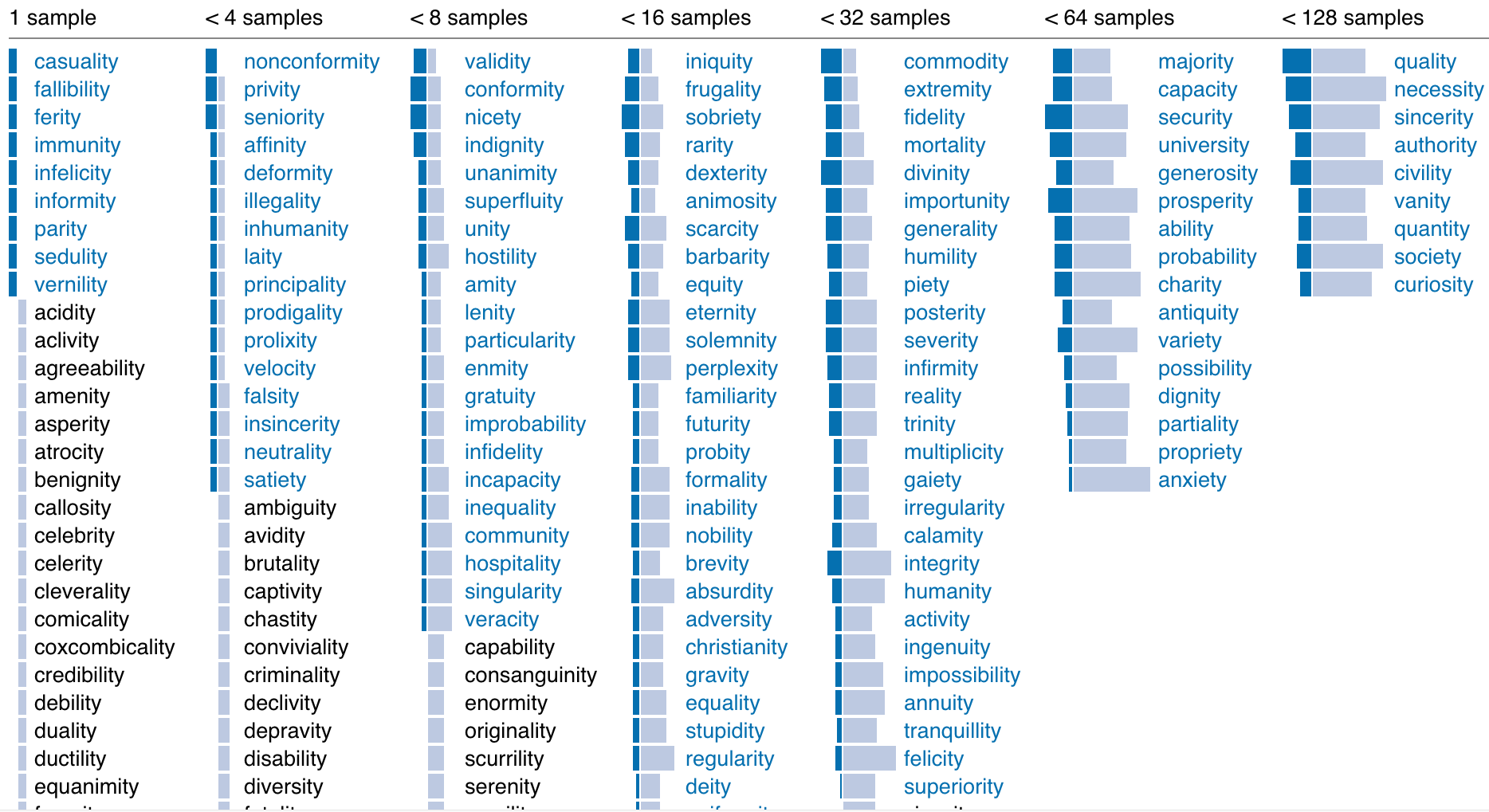
Show: [type list](#) [sample list](#)

This collection contains 583086 running words and 133 types.

Only approx. **0.000%** of random collections with 583086 running words contain at most 133 types.

This finding is probably interesting: the false discovery rate is **smaller than 0.001**.

Types that occur in...



1 sample

casualty
fallibility
ferity
immunity
infelicity
informity
parity
sedulity
vernility
acidity
aclivity
agreeability
amenity
asperity
atrocit

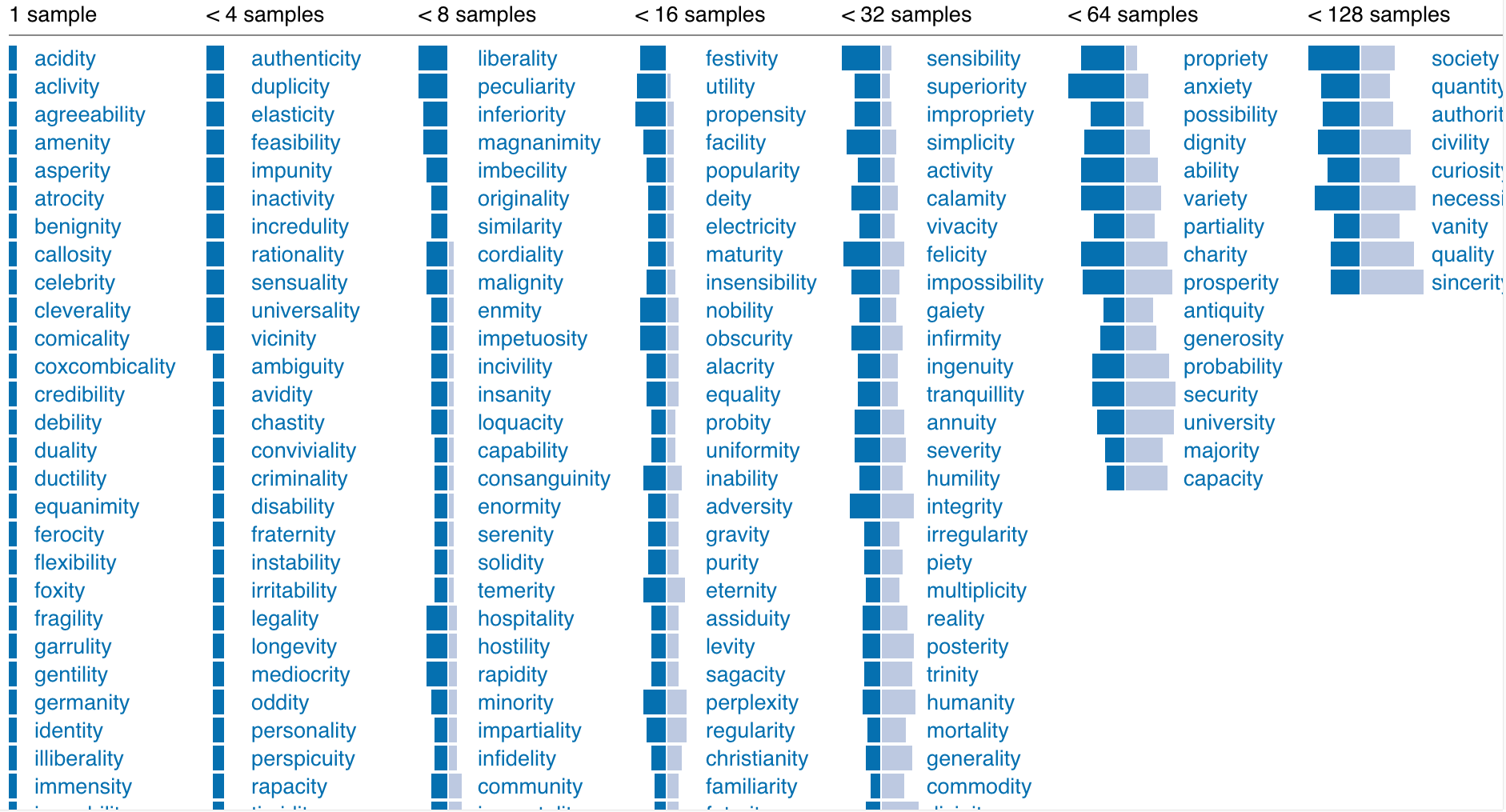
< 4 samples

nonconformity
privity
seniority
affinity
deformity
illegality
inhumanity
laity
principality
prodigality
prolixity
velocity
falsity
insincerity
neutrality

< 8 samples

validity
conformity
nicety
indignity
unanimity
superfluity
unity
hostility
amity
lenity
particularity
enmity
gratuity
improbability
infidelity

Types that occur in...



1 sample

< 4 samples

< 8 samples
































































■ acidity
■ acivity
■ agreeability
■ amenity
■ asperity
■ atrocity
■ benignity
■ callosity
■ celebrity
■ cleverality
■ comicality
■ coxcombicality
■ credibility
■ debility
■ duality

■ authenticity
■ duplicity
■ elasticity
■ feasibility
■ impunity
■ inactivity
■ incredulity
■ rationality
■ sensuality
■ universality
■ vicinity
■ ambiguity
■ avidity
■ chastity
■ conviviality

■ liberality
■ peculiarity
■ inferiority
■ magnanimity
■ imbecility
■ originality
■ similarity
■ cordiality
■ malignity
■ enmity
■ impetuosity
■ incivility
■ insanity
■ loquacity
■ capability

Written in 1760-1800

Show: [summary](#) [type list](#) [sample list](#)

Sample	Words	Tokens	Types	Unique	Hapaxes	Common types
WCOWPER-1780	45350 	173 	85 	6	5	opportunity × 22, abi
FBURNEY-1780	44781 	121 	56 	4	3	opportunity × 13, anx
TTWINING-1780	38896 	105 	58 	7	5	curiosity × 13, oppo
HPIOZZI-1780	38409 	92 	49 	1	1	quality × 7, society ×
SLENNOX-1780	34645 	94 	41 	1	1	society × 14, anxiety
SLENNOX-1760	31032 	54 	26 	2	1	vanity × 8, civility × 5
WCLIFT-1780	28984 	43 	12 	1	1	opportunity × 28, pro
MWOLLSTONE-1780	28709 	109 	45 	1	1	opportunity × 13, ser
JAUSTEN-1780	27955 	46 	30 	1	1	opportunity × 8, dign
GCULLEY-1780	24753 	32 	13 	0	0	quantity × 13, quality
GEORGE4-1780	23756 	77 	37 	0	0	opportunity × 12, soc
DGARRICK-1760	22880 	39 	27 	2	2	security × 6, curiosity
CBURNEY-1780	20196 	55 	42 	3	3	society × 6, antiquity
EDRAPER-1760	19996 	125 	61 	3	3	society × 17, dignity
CBURNEY-1760	19255 	44 	33 	1	1	curiosity × 4, opportu
J1WEDGWOOD-1760	18516 	54 	32 	0	0	society × 4
WJONES-1780	18077 	84 	38 	2	1	society × 11, felicity ×
ISANCHO-1760	17618 	84 	44 	2	2	charity × 5, felicity × 5
TTWINING-1760	17454 	54 	34 	2	2	opportunity × 9, curio
JBENTHAM-1780	16603 	54 	28 	2	2	necessity × 10, oppo
RNEWDIGATE-1760	15030 	44 	26 	0	0	university × 6, curios

Hapaxes	Common types	Unique types
5	opportunity × 22, ability × 7	spirituality, duality, ductility, equanimity, impiety, sterility
3	opportunity × 13, anxiety × 11, necessity × 6	agreeability, activity, atrocity, notability
5	curiosity × 13, opportunity × 7, charity × 5	incredibility, suavity, coxcombicality, foxity, germanity, incommodity, nudity
1	quality × 7, society × 6	credibility
1	society × 14, anxiety × 8	mobility
1	vanity × 8, civility × 5, propriety × 5	gentility, incapability
1	opportunity × 28, prosperity × 4, capacity × 2	materiality
1	opportunity × 13, sensibility × 9	sociability
1	opportunity × 8, dignity × 3	impurity
0	quantity × 13, quality × 5, propriety × 3	
0	opportunity × 12, society × 8, integrity × 5	
2	security × 6, curiosity × 5	comicality, scoundrellity
3	society × 6, antiquity × 3	illiberality, notoriety, publicity
3	society × 17, dignity × 7	flexibility, venality, volatility
1	curiosity × 4, opportunity × 3, quality × 3	puerility
0	society × 4	
1	society × 11, felicity × 8, opportunity × 6	identity, stability
2	charity × 5, felicity × 5, humanity × 5	benignity, cleverality
2	opportunity × 9, curiosity × 5, university × 4	amenity, ridiculability
2	necessity × 10, opportunity × 8, authority × 5	immobility, incombustibility
0	university × 6, curiosity × 4, opportunity × 4	

Sample	Token	Before	Word	After
TTWINING-1780	coxcombicality	the least appearance of	coxcombicality	, affectation, impo

“If you find any thing that has the least appearance of **coxcombicality**, affectation, importance, conceit, &c., have no mercy upon it.”

Thomas Twining to Richard Twining, 1788