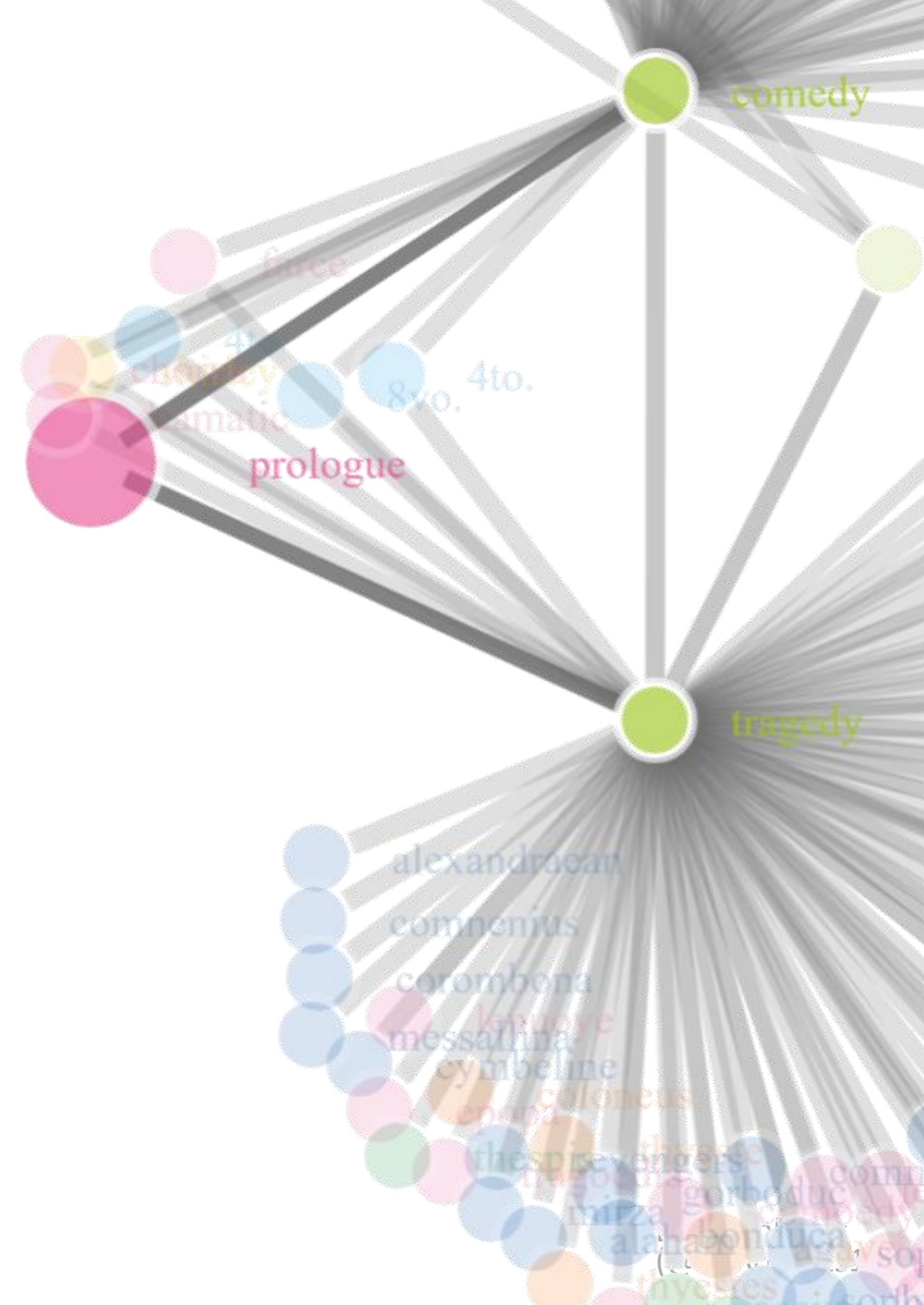


Using Methods from Corpus Linguistics on Historical Texts

Outline of session:

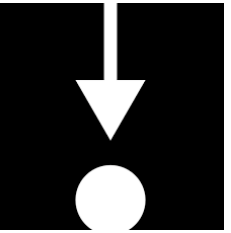
- What is Corpus Linguistics?
- Why is it useful?
- Specific concerns for historical texts and creation of historical corpora
- Hands-on element



What is corpus linguistics?



- Established method in linguistics
- Digital method used to study large volumes of text
- ‘Corpus’ = body
- Specialist software
 - E.g. CQPweb (Andrew Hardie), AntConc (Laurence Anthony), SketchEngine, #LancsBox and many others



Theoretical grounding of corpus linguistics:

- ‘You shall know a word by the company it keeps’
(Firth)
- **Usage-based** to look for patterns in language use
- Interest in how the language ‘functions’ – i.e. how it is used in ‘real language examples’, rather than how it could be used (‘formal’)



Why is corpus linguistics useful?

- Typically based on counting features to tell us more about linguistic phenomenon - quantitative
- Widely used in Digital Humanities
- Provide different perspectives on texts:
*‘Corpus analysis offers a **‘surface’ reading**, a glimpse of the panoramic **wide vista of a word or a concept** across fragments of texts’* (Findlay 2018)
- Can be used to confirm pre-existing literary findings/theory



Things we need:

- A corpus
 - Pre-built or build your own!
 - We're going to focus on using pre-built but can discuss building our own if of interest to group
- A research question
- Appropriate software
- Search strings/terms





Historical Texts and their challenges

Some issues...

1. Converting texts (inc. manuscripts) to a machine-readable format
 - a) Keying
 - b) OCR
 - c) Standardising spellings (where appropriate)
 - d) Grammatical tagging
2. Precision vs recall





Converting texts

Keying, OCR, Annotation

An issue of complexity

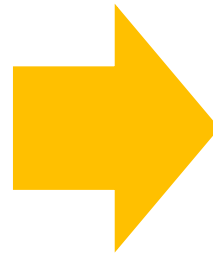
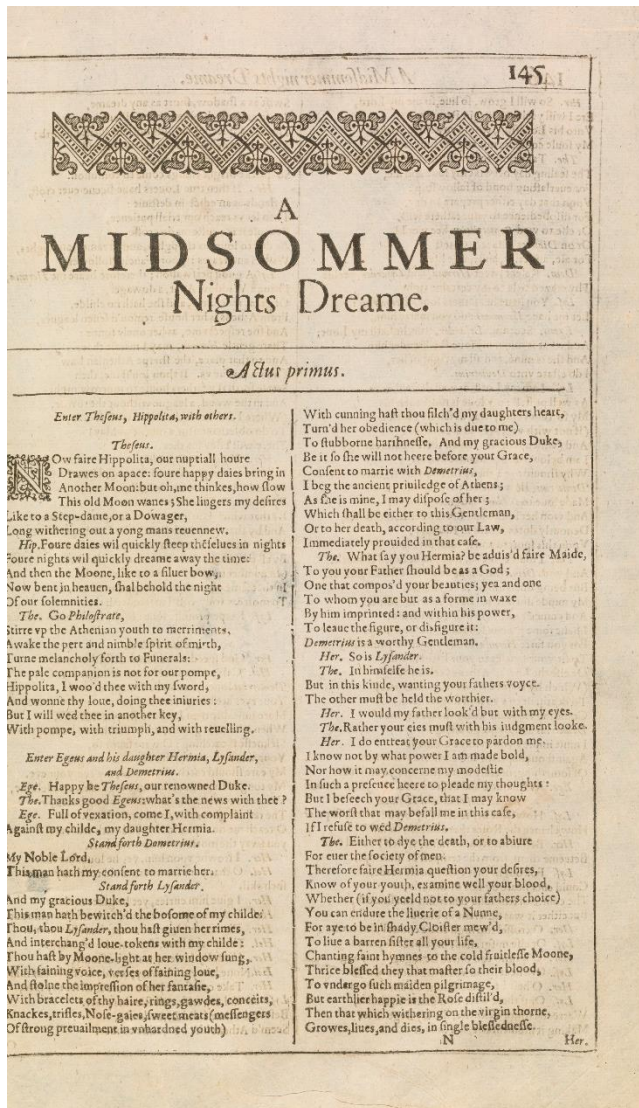
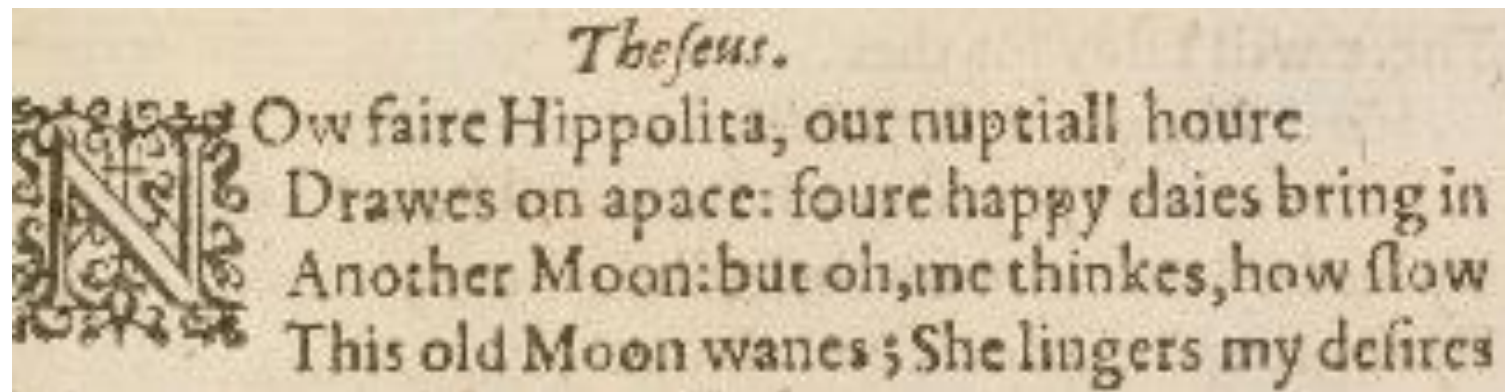


Image from the The Bodleian First Folio

Keying



EEBO-TCP

Now faire Hippolita, our nuptiall hower
Draws on apase: fower happy daies bring in
An other Moone: but oh, me thinks, how slow
This old Moone waues! She lingers my desires,

CQPweb

Now fair Hippolyta, our nuptial hour
Draws on apace: four happy days bring in
Another Moon: but oh, me thinks, how slow
This old Moon wanes; She lingers my desires

ORIGINAL

MANUAL KEYING
(TRANSCRIPTION)

(STANDARDISED
USING VARD2)

OCR

ORIGINAL



OCR

any insouciant manner," but now"—and here the words deserve the emphasis of italics—"even spies can't get the necessary travel permits." The remark is so shocking that it is difficult to sort out its full implications. It has always been assumed that the twin duties of a spy were, first, to provide himself with a full cargo of forged papers, among which a mere travel permit would be an inconsiderable detail, knocked off, as it were, in a matter of moments, and, second, to extract genuine papers from other people. And if he cannot perform the first, and less difficult, task, how can he ever proceed with the second?

It is indeed a poor spy who cannot even get to the scene of his activities, and the desperate plea of an eager junior, turned back at his own station by a ticket-collector normally employed in scrutinizing the dates on suburban season-tickets, that he would willingly be parachuted into the appropriate territory would be hardly likely to obtain either sympathy or employment.

The complexity and variety of all that the modern spy needs to pass as a citizen of the modern welfare State may explain the pusillanimity implied in the conductor's

the direct service of the Ministry of Health. But, of an authorized complement of 36 at the present moment, the department has only 11, and it is still losing from that number. The authorized complement is to be increased to 44 in the next financial year, but the present failure of the department, not only in recruitment but in holding the dental staff as it makes this increase in authorized complements seem ludicrous. No one who has the interest of the Civil Service at heart, let alone the National Health Service itself, can envisage this with satisfaction. It arises directly from the fact that dentists in the Civil Service are receiving only two-thirds of what they could readily secure outside of the service.

Much the same picture can be given of the medical service, and each department has a story to tell of the impossibilities of recruiting medical officers of the required standard, salaries so much lower than those paid to medical men outside of the service. The degree of undermanning in the Civil Service which represents merely the general shortage of highly skilled men and women is one thing. The degree of undermanning which is a direct result of the Government providing standards of remuneration outside the service greatly in excess of those they provide inside is merely fanatic.

Yours faithfully,
STANLEY MAYNE, General Secretary,
Institution of Professional Civil
Servants,
17, Hans Place, S.W.1, Jan. 9.

Sir,—As an ex-Civil servant and therefore

it is necessary to treat him as he is used to being treated by his domestic suppliers. The terms are set by him—they are the terms on which he is willing to buy, not those on which his supplier would like to sell. There is no "way we have always done business." There is, indeed, an entirely different relationship between manufacturer, wholesaler, and retailer in the United States from that which exists in other countries.

Whereas in Europe a dog is inclined to eat dog all down the line, each party being satisfied when he has sold his goods to the next in the chain, there is, in the United States, a partnership whose combined activities are directed towards eating one dog—the customer. *Caveat emptor* applies only to the public. The retailer will cancel orders or return unsaleable goods (regardless of whether they were ordered on sale or return or not) because he is the man who sets the trap and a trap baited with unpalatable cheese is a waste of time and money for all concerned.

Once, however, this different relationship is understood, and the intensely competitive conditions which have produced it are realized, the American retailer will be found to be more imaginative and more

most over 10 years in terms of general business activity, with Arizona and California next. In growth of population California, Oregon, and Washington lead, while in the centre of the country, in the agricultural states in the Middle West, there has been a drop in population.

AREAS OF EXPANSION

These developments, if properly understood, can be of immense value to the British exporter. The areas where expansion, both of population and industrial activity, has been greatest are those which can be approached by sea; they are those where the lower freight rates on British vessels reduce the disadvantages of the British manufacturer compared with the manufacturer in the north-east or Middle West, and may in some cases, if rates of tariff are sufficiently reduced, give him a positive advantage. They are also the areas where resentment at "eastern imperialism" is greatest and where, therefore, there is little, if any, more resistance to British goods than to goods from the industrial north-east.

The Far West is becoming every year more self-sufficient and no longer has to import "from the east the finished goods

Times
Digital
Archive

"Exports to America."
Times, 13 Jan. 1950, p.
5. *The Times Digital
Archive*,
[link.gale.com/apps/doc/
CS16992813/TTDA?u=
unilanc&sid=bookmark
-TTDA&xid=7e083e85](https://link.gale.com/apps/doc/CS16992813/TTDA?u=unilanc&sid=bookmark-TTDA&xid=7e083e85).
Accessed 31 May 2022.

spoilt and pampered individual .

But to do business with him it is necessary to treat him as he is used to being treated by his domestic suppliers .

The terms are set by him—they are the terms on which he is willing to buy , not those on which his supplier would like to sell .

There is no " way we have always done business .

" There is , indeed , an entirely different relationship between manufacturer , wholesaler , and retailer in the United States from that which exists in other countries .

Whereas in Europe **dog** is inclined to eat dog all down the line , each party being satisfied when he has sold his goods to the next in the chain , there is , in the United States , a partnership whose combined activities are directed towards eating one dog—the customer .

Caveat emptor applies only to the public .

The retailer will cancel orders or return unsaleable goods (regard- less of whether they were ordered on sale or return or not) because he is the man who sets the trap and a trap baited with unpalatable cheese is a waste of

CQPweb

Times Online:
1950s
Text:
TMZ_1950_01_13
Date: 1950

Ellen Roberts
(@E_Roberts19)

OCR issues

ORIGINAL



OCR (!)

will be
l, and
ie Per-
Wed-
the St.
; and
ice, on
k, 'till
orning,
s Cele-
their

in the course of the evening's Entertainments will be pre-
sented,
The amazing Exhibition of the
DANCING DOGS.
From France and Italy, by Sign. and Signora Maifa.
This Exhibition will conclude with a Variety of Dogs
dressed *En Militaire*, besieging a Town; and two Bull
Dogs, one of them is an English Bull Dog, that rather
than quit his hold suffers himself to be drawn Thirty
Feet high, whilst the Machine is surrounded with Fire-
Works, representing a heavy Discharge of Small Arms
and Artillery.
To which will be added a new Pantomime, called, -
HARLEQUIN'S REVENGE :
Or, The Camp in an Uproar.

employed 814
ported were gr
great beams,
tails, &c.
Copenhagen,
Norway mentio
no ways extrac
that quarter has
Vienna, April
stadt, that on
tion (or enrolli
and that they a
frontier regime



L 't Y T .

With various'forpirifrg EaCperimeats , particularly the Sympathetic Clock , Rings , Box , and Pigeon , Watch , Orange , &c .

Bc. - In the courfS of theevening 's Entertainments will be pre- S uted , The amazing Exhibition of the D A N C I N G D O G S .. Ftem France and Italy , by Sign .

and Signora Maifa .

This Exhibition will coticlude with a Variety of Dogs dreffed E. IdiliC ; ire , befieging a Town ; 2nd two Bell Dogs , one of them is an EngliSh Bull **Dog** , that rather that , quit his hold fuffers hinifelf to be drawn Tisirty Feet high , NvhilSt the Machinre is furrounded with FireWVorks , reprefedtinge a heavy DiScharge of Small Arms and Artillery .

To which will be added a new Pantomime , called , - HARLEQUIN 's REVENGE : Or , The Camp in an Uproar. tVith Additions and .

Alterations , particularly a new Scene of HARLEQTIN 's Marriage .

Boxes 3s ; Pitt as .

Gall. is .

Side Gd'eries 6d .

Doors to be opened at Half paSt Five , to begin at Half paft

Times Digital
Archive

"ROYAL SOCIETY of
MUSICIANS. Under the
Patronage, and by Command
of their." *Times*, 6 May 1785,
p. 1. *The Times Digital Archive*,
link.gale.com/apps/doc/CS16908454/TTDA?u=unilanc&sid=b00kmark-TTDA&xid=6a7ebe3c.
Accessed 31 May 2022.

CQPweb

Times Online: 1780s
Text: TMZ_1785_05_06
Date: 1785

Ellen Roberts
(@E_Roberts19)

Why a 'clean' text is important:

- Most corpus software relies on 'string-matching'
- Computers are 'dumb' and don't **understand** what we are looking for and to 'look' for alternatives
- E.g. if we search for 'dog', the computer will search for a 'd' followed by an 'o' then a 'g'
 - Sees words as 'strings' or consecutive characters
 - Doesn't know that it constitutes a type of animal



Spelt how?

E.g. ‘dog’

Noun or verb?

Quick search results

Showing 1-4 of 4 entries

Widen search? Find ‘dog’ in: » phrases (544) » definitions (1649) » etymologies (424) » quotations (9753) » full text (5378)

View as: List | Timeline Sort by: Entry | Frequency | Date

1. [dog, n.¹](#) View full entry OE

...A domesticated carnivorous mammal, *Canis familiaris* (or *C. lupus familiaris*), which typically has a long snout, an acute sense of smell, non-retractile claws, and a barking, howling, or...

2. [dog, n.²](#) View full entry ?1550

...In oaths and asseverations: = god...

3. [dog, v.¹](#) View full entry 1519

...transitive. To follow like a dog on the heels of; to track (a person, or his or her trail, footsteps, etc.) closely and persistently. In extended use: to pursue doggedly, esp....

4. [dog, v.²](#) View full entry 1835

...transitive (chiefly in passive). Used as a euphemism for ‘damn’ ...



dog, n.¹

Text size: A A

View as: Outline | Full entry

Quotations: Show all | Hide all

Keywords: On | Off

Pronunciation: Brit. /dɒg/, U.S. /dɔːg/, /dag/

Forms:

α. Old English **doega**, Middle English–1500s **doggue**, Middle English–1600s **doge**, Middle English–1600s **dogge**, Middle English–1800s **dogg**, Middle English– **dog**, 1600s **dogue**, 1600s **togg** (*Welsh English*); *Scottish* pre-1700 **dogge**, pre-1700 1700s **doge**, pre-1700 1700s **dogg**, pre-1700 1700s– **dog**.

β. late Middle English **doog**, late Middle English–1600s **dooge**; *Scottish* pre-1700 **doig**, pre-1700 **doigg**, pre-1700 **doog**, pre-1700 **doogg**.

γ. 1500s–1600s **dodge**.

δ. *Scottish* pre-1700 **dowge**, pre-1700 1800s– **doug**, pre-1700 1800s– **dowg**.

ε. *English regional* 1800s– **doog** /dʊg/ (*Leicestershire*), 1800s– **dug** /dʌg/ (chiefly *north midlands*); *Scottish* 1800s– **dug**; *Irish English* 1800s– **dug** (*Wexford* and *northern*).

ζ. *English regional* 1800s **dorg**, 1800s– **dawg** (*Hampshire*), 1800s– **doag** (*Northumberland*); *U.S. regional* 1800s– **dawg**, 1800s– **dorg**; *Caribbean* 1900s– **dorg**; *Australian* 1800s– **dorg**.

OED (Eng) Variants in EEBOv3

Query "(doggue|doge|dogge|dogg|dog|dogue|doog|dooge)" returned 31,771 matches in 7,930 different texts (in 1,202,214,511 words [44,422 texts]; frequency: 26.427 instances per million words)

Showing frequency breakdown of words in this query, at the query node; there are 8 different types and 31,771 tokens at this concordance position. [9.254 seconds]

No.	Query result	No. of occurrences	Percent
1	Dog	30988	97.54%
2	doge	685	2.16%
3	Dogg	39	0.12%
4	doggue	23	0.07%
5	dogge	17	0.05%
6	dogue	13	0.04%
7	Doog	4	0.01%
8	dooge	2	0.01%

Verges. If you heare a child crie in the night you must call to the nurse, and bid her still it.

Watch. How if the nurse be asleepe and will not heare us?

Dog. Why then depart in peace, and let the childe wake her with crying, for the ewe that will not heare her Lambe when it baes, will neuer answere a calfe when he bleates.

Verges. 'Tis verie true.

Dog. This is the end of the charge: you constable are to present the Princes owne person, if you meete the Prince in the night, you may staie him.

Verges. Nay birladie that I thinke a cannot.

Dog. Five shillings to one on't with anie man that knowes the Statues, he may staie him, marrie not without the prince be willing, for indeede the watch ought to offend no man, and it is an offence to stay a man against his will.

Verges. Birladie I thinke it be so.

Dog. Ha, ah ha, well masters good night, and there be anie matter of weight chances, call vp me, keepe your fellowes counsailes, and your owne, and good night, come neighbour.

blouds, betweene foureteene & fiue & thirtie, sometimes fashioning them like *Pharaos* souldiours in the rechie painting, sometime like god Bels priests in the old Church window, sometime like the shauen *Hercules* in the smircht worm-eaten tapestrie, where his cod-peece seemes as massie as his club.

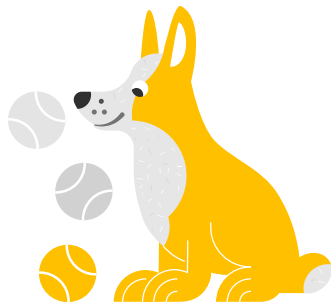
Con. All this I see, and see that the fashion weares out more apparrell then the man; but art not thou thy selfe giddie with the fashion too that thou hast shifted out of thy tale into telling me of the fashion?

Bor. Not so neither, but know that I haue to night wooed *Margaret* the Lady *Heroes* gentle-woman, by the name of *Hero*, she leanes me out at her mistris chamber-vvindow, bids me a thousand times good night: I tell this tale vildly. I should first tell thee how the Prince *Claudio* and my Master planted, and placed, and possessed by my Master *Don Iohn*, saw a far off in the Orchard this amiable incounter.

Con. And thought thy *Margaret* was *Hero*?

Bor. Two of them did, the Prince and *Claudio*, but the diuell my Master knew she was *Margaret* and partly by his oathes, which first posselt them, partly by the darke night which did deceiue them, but chiefly, by my villa-

Even as character abbreviations – a search for 'dog' would return this too, unless we account for it in our digital version of the text



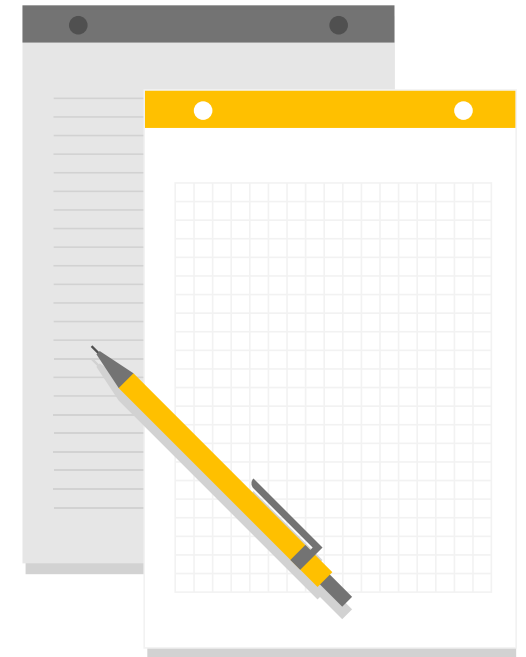
A possible solution?: Corpus Annotation



What is annotation?

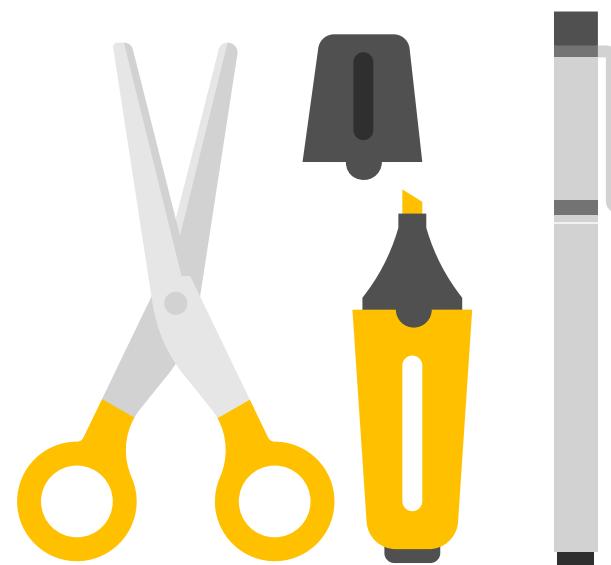
- *'It can be defined as the practice of adding **interpretative, linguistic** information to an electronic corpus'*
- *'Annotation can also refer to the end-product of this process: [...] the electronic **representation** of the language material itself...'*

Geoffrey Leech (1997:2)



In other words...

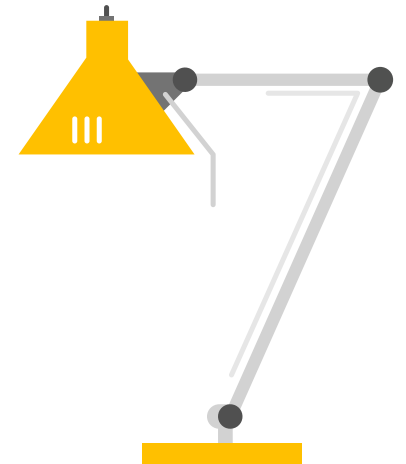
- Annotation is another tool in the corpus toolbox
- Annotation consists of mark-up tags added to the text
 - For both text and word-based levels



Taggers:

- Examples of taggers:
 - CLAWS (Constituent Likelihood Automatic Word-tagging System) – **GRAMMAR/WORD CLASS**
 - Stanford tagger - **GRAMMAR**
 - VARD (Variant Detector) - **SPELLING**
 - USAS (UCREL Semantic Analysis System) - **SEMANTICS**

The Encyclopedia of Shakespeare's language project on CQPweb - **Enhanced** Shakespearean Corpus



Tags 'in the wild':

```
1
2 <text title="A MIDSOMMER Nights Dreame." id="MND">
3 <act n="1" title="Actus primus" id="MND_1">
4 <scene n="1" type="scene" title="" id="MND_1_1">
5 ...
6 <stage>
7 Enter      VV0      enter      M1      VERB      enter_VERB      Enter
8 Theseus    NP1      theseus    A4:1    SUBST      theseus_SUBST   Theseus
9 ,          YCOM     PUNC      STOP    PUNC_STOP    ,
10 Hippolyta  NP1      hippolyta  Z99     SUBST      hippolyta_SUBST  Hippolita
11 ,          YCOM     PUNC      STOP    PUNC_STOP    ,
12 with       IW       with       Z5      PREP       with_PREP        with
13 others     NN2      other      A6:1    SUBST      other_SUBST      others
14 .          YSTP     PUNC      STOP    PUNC_STOP    .
15 </stage>
16 <u who="MND_Theseus" label="Theseus.">
17 <lb/>
18 Now        RT       now        T1:1:2  ADV        now_ADV         Now
19 fair       JJ       fair       G2:2    ADJ        fair_ADJ        faire
20 Hippolyta  NP1      hippolyta  Z99     SUBST      hippolyta_SUBST  Hippolita
21 ,          YCOM     PUNC      STOP    PUNC_STOP    ,
22 our        APPGE    our        Z8      PRON       our_PRON        our
23 nuptial    JJ       nuptial    S4      ADJ        nuptial_ADJ     nuptiall
24 hour       NNT1     hour       T1:3    SUBST      hour_SUBST      heure
25 <lb/>
26 ...
27 </u>
28 <u who="MND_Hippolyta" label="Hip.">
29 Four       MC       four       N1      ADJ        four_ADJ        Foure
30 days       NNT2     day        T1:3    SUBST      day_SUBST       daies
31 will       VM       will       T1:1:3  VERB       will_VERB       wil
32
```



Tags 'in the wild':

dog → Noun or verb?

*Destruction straight shall **dog** them at the heels*



*Destruction_NN1 straight_RR shall_VM
dog_VVI
them_PPHO2 at_II the_AT heels_NN2*

Richard II (5: 3)

Ellen Roberts
(@E_Roberts19)



Annotation ‘in the wild’:

ESC: First Folio Plus (pre-final version): <i>powered by CQPweb</i>	
Metadata for ESC: First Folio Plus (pre-final version)	
Corpus title	ESC: First Folio Plus (pre-final version)
CQPweb's short handles for this corpus	shakfinc006 / SHAKFINC006
Total number of texts in corpus	38
Total word tokens in all corpus texts	1,038,509
Word types in the corpus	24,942
Standardised type:token ratio (1,000-token basis)	0.3759 types per token
Non-standardised type:token ratio	0.0240 types per token
Text metadata and word-level annotation	
The database stores the following information for each text in the corpus:	Approximate date of first production, in two bins
	Simple genre categories
	Genre categories, taking 'problem plays' into account
	Play code
	Source
The primary classification of texts is based on:	Simple genre categories
Words in this corpus are annotated with:	Part-of-speech tag (C6 tagset)
	Lemma
	Semantic tag (USAS tagset)
	Simple POS (Oxford Simplified Tagset)
	Tagged lemma
	Full USAS analysis (USAS tagset)
	Unregularised spelling
The primary word-level annotation scheme is:	Part-of-speech tag

Textual
annotation

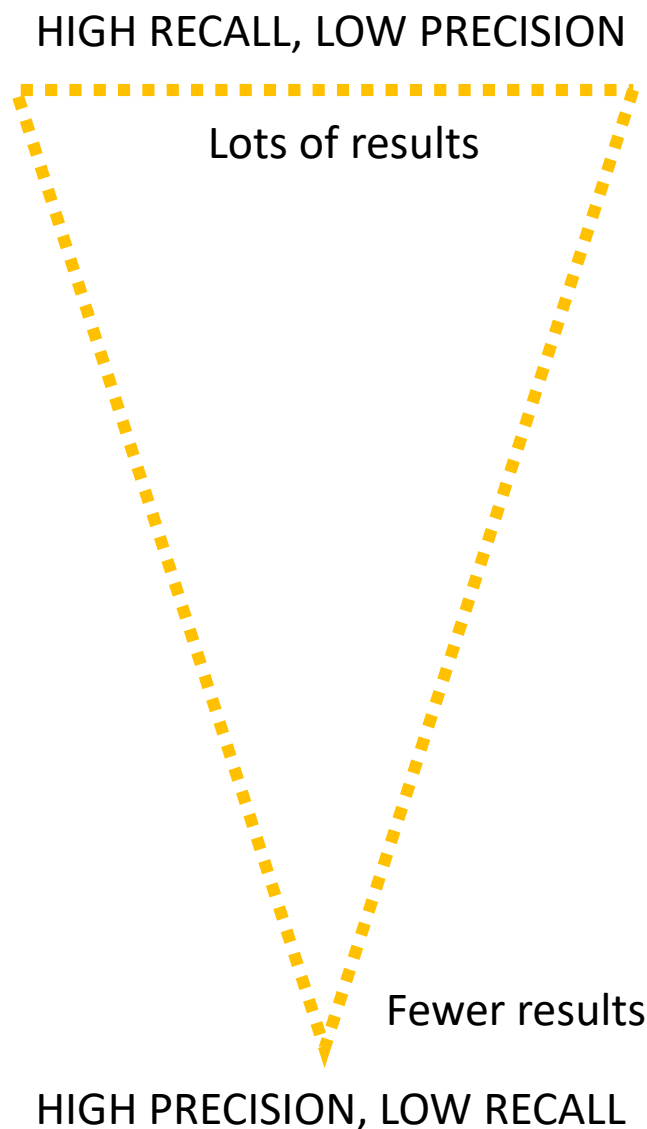
Word-level
annotation



Ellen Roberts
(@E_Roberts19)

Precision vs. recall

- Depends on our aims and quality of data
 - E.g. Use 'dog' in early modern English generally (broad q) vs. as a verb in Shakespeare (narrow q)
- For 'messy' data, high recall – or GENERAL RESULTS may be desired
 - We might want to input all variant spellings in, and then filter the corpus results
- For 'clean' data, high precision – or SPECIFIC RESULTS may be desired
 - We can make use of highly tagged corpora to filter results as an output from the corpus

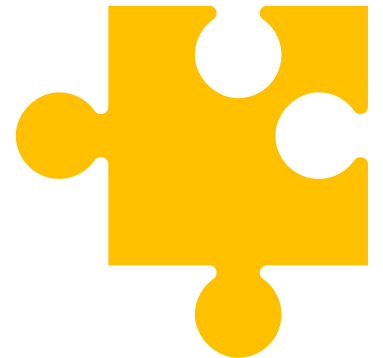




Hands on!

Some preliminaries:

- Session links available at: **github.com/EllenRoberts/QMLHS**
- Sign up to a CQPweb account, using the instructions outlined on GitHub.
- In addition, sign up to have the Encyclopedia of Shakespeare's Language data added to your CQPweb library.



Florence Early English Newspapers	Heisinki Corpus (Diachronic)	HistPop
Lake District Corpus: 28GS	Lancaster Newsbooks Corpus (1654 part)	Paston Letters Vol 1
Sample of ``Phil''	Shakespeare First Folio	
Enhanced Shakespearean Corpus		
ESC: Comparative Plays (pre-final version)	ESC: EEBO-TCP Segment (pre-final version)	ESC: First Folio Plus (pre-final version)
ESC: Poems (pre-final)	ESC: Quartos (pre-final version)	
Historical News Text		
British Library C19 News: Belfast Newsletter	British Library C19 News: The Era	British Library C19 News: Glasgow Herald
British Library C19 News: Hampshire/Portsmouth Telegraph	British Library C19 News: Ipswich Journal	British Library C19 News: Liverpool Mercury (early period)
British Library C19 News: Liverpool Mercury (mid period)	British Library C19 News: Liverpool Mercury (late period)	British Library C19 News: Northern Echo
British Library C19 News: Pall Mall Gazette	British Library C19 News: Reynold's Daily	British Library C19 News: Western Mail
Times Online: 1780s	Times Online: 1790s	Times Online: 1800s
Times Online: 1810s	Times Online: 1820s	Times Online: 1830s
Times Online: 1840s	Times Online: 1850s	Times Online: 1860s

Menu	ESC: First Folio Plus (pre-final version): powered by CQPweb	
Corpus queries	Standard Query	
Standard query	<div></div>	
Restricted query		
Word lookup		
Frequency lists		
Keywords		
Analyse corpus		
Saved query data	Query mode: <div>Simple query (ignore case)</div> Simple query language syntax	
Query history	Number of hits per page: <div>50</div>	
Saved queries	Match strategy: <div>Standard</div>	
Categorised queries	Restriction: <div>None (search whole corpus)</div>	
Upload a query	<div>Start query</div> <div>Reset query</div>	
Create/edit subcorpora		
Corpus info		
<div>View corpus metadata</div>		
C6 tagset		
USAS tagset		
Oxford Simplified Tagset		
About CQPweb		
CQPweb main menu	2021-12-17	Install your own corpus: now available
Help system		The long-anticipated "install your own corpus" system is now up and running, and open to ALL users on this server.
Video tutorials		For how to get started, see the tutorial video at Indonesia Corpus Lab: https://youtu.be/o5bn458OVg8
Who did it?		Maximum corpus size is quite small (1,000,000 words tagged, 3,000,000 words untagged). This is to avoid the system running out of disk space.
Latest news		Please report any issues to me at a.hardie@lancaster.ac.uk .
Report bugs		Have fun everybody!
		best
		Andrew.
	2020-12-16	EEBO data: now more widely available
		Due to the underlying EEBO-TCP data having become open to the public this summer, I'm now able to grant non-restricted access to all users to the "EEBO v2" and "EEBO v3" corpora.
		A V4 will be along Real Soon Now.
		best,
		Andrew.

Menu	ESC: First Folio Plus (pre-final version): <i>powered by CQPweb</i>	
Corpus queries	Metadata for ESC: First Folio Plus (pre-final version)	
Standard query	Corpus title	ESC: First Folio Plus (pre-final version)
Advanced query	Query and short handles for this corpus	shakfinc006 / SHAKFINC006
Word lookup	Total number of texts in corpus	38
Frequency lists	Total word tokens in all corpus texts	1,038,509
Keywords	Word types in the corpus	24,942
Analyse corpus	Standardised type:token ratio (1,000-token basis)	0.3759 types per token
Saved query data	Non-standardised type:token ratio	0.0240 types per token
Query history	Text metadata and word-level annotation	
Saved queries	The database stores the following information for each text in the corpus:	Approximate date of first production, in two bins
Categorised queries		Simple genre categories
Upload a query		Genre categories, taking 'problem plays' into account
Create/edit subcorpora		Play code
Corpus info		Source
View corpus metadata	The primary classification of texts is based on:	Simple genre categories
Corpus manual unavailable	Words in this corpus are annotated with:	Part-of-speech tag (C6 tagset)
C6 tagset		Lemma
USAS tagset		Semantic tag (USAS tagset)
Oxford Simplified Tagset		Simple POS (Oxford Simplified Tagset)
About CQPweb		Tagged lemma
CQPweb main menu		Full USAS analysis (USAS tagset)
Help system	The primary word-level annotation scheme is:	Unregularised spelling
Video tutorials		Part-of-speech tag
Who did it?		
Latest news		
Report bugs		

Let's search for the word:
'will'

Your query "will" returned 5,182 matches in 38 different texts (in 1,038,509 words [38 texts]; frequency: 4,989.846 instances per million words) [0.081 seconds - retrieved from cache]			
<	<<	>>	>
Show Page:	1	Line view	Show in random order
Choose action...			
No	Text	Solution 1 to 50	Page 1 / 104
1	1H4 1H4_1_1	a twelvemonth old , And bootless t is to tell you we	will go : Therefore we meet not now . Then let me hear
2	1H4 1H4_1_1	to Jerusalem . Cousin , on Wednesday next , our Council we	will hold At Windsor , and so inform the Lords : But come
3	1H4 1H4_1_1	be done , Than out of anger can be uttered . 1H4 Westmorland :I	will my Liege . Exeunt Enter Henry Prince of Wales , Sir John
4	1H4 1H4_1_2	none . 1H4 Prince :What , none ? 1H4 Falstaff :No , not so much as	will serve to be Prologue to an Egg and Butter . 1H4 Prince :Well ,
5	1H4 1H4_1_2	the wicked . I must give over this life , and I	will give it over : and I do not , I am a
6	1H4 1H4_1_2	for he was never yet a Breaker of Proverbs : He	will give the devil his due . 1H4 Poins :Then art thou damned for keeping
7	1H4 1H4_1_2	we may do it as secure as sleep : if you	will go , I will stuff your Purses full of Crowns : if
8	1H4 1H4_1_2	it as secure as sleep : if you will go , I	will stuff your Purses full of Crowns : if you will not ,
9	1H4 1H4_1_2	I will stuff your Purses full of Crowns : if you	will not , tarry at home and be hanged . 1H4 Falstaff :Hear ye Yedward
10	1H4 1H4_1_2	and go not , I 'll hang you for going . 1H4 Poins :You	will chops . 1H4 Falstaff :Hal , wilt thou make one ? 1H4 Prince :Who , I
11	1H4 1H4_1_2	1H4 Falstaff :Why , that 's well said . 1H4 Prince :Well , come what	will , I 'll tarry at home . 1H4 Falstaff :I 'll be a Traitor
12	1H4 1H4_1_2	John , I prithee leave the Prince & me alone , I	will lay him down such reasons for this adventure , that he shall
13	1H4 1H4_1_2	those men that we have already waylaid , yourself and I ,	will not be there : and when they have the booty , if
14	1H4 1H4_1_2	shall we part with them in setting forth ? 1H4 Poins :Why , we	will set forth before or after them , and appoint them a place
15	1H4 1H4_1_2	wherein it is at our pleasure to fail ; and then	will they adventure upon the exploit themselves , which they shall have no
16	1H4 1H4_1_2	set upon them . 1H4 Prince :Aye , but t is like that they	will know us by our horses , by our habits , and by
17	1H4 1H4_1_2	I 'll tie them in the wood , our visors we	will change after we leave them : and sirrah , I have Cases
18	1H4 1H4_1_2	to immask our noted outward garments . 1H4 Prince :But I doubt they	will be too hard for us . 1H4 Poins :Well , for two of them
19	1H4 1H4_1_2	reason , I 'll forswear Arms . The virtue of this Jest	will be , the incomprehensible lies that this fat Rogue will tell us
20	1H4 1H4_1_2	this Jest will be , the incomprehensible lies that this fat Rogue	will tell us , when we meet at Supper : how thirty at
21	1H4 1H4_1_2	my Lord . Exit Poins 1H4 Prince :I know you all , and	will awhile uphold The unyoked humour of your idleness : Yet herein will
22	1H4 1H4_1_2	will awhile uphold The unyoked humour of your idleness : Yet herein	will I imitate the Sun , Who doth permit the base contagious clouds
23	1H4 1H4_1_2	offence a skill , Redeeming time , when men think least I	will . Enter the King , Northumberland , Worcester , Hotspur , Sir
24	1H4 1H4_1_3	You tread upon my patience : But be sure , I	will from henceforth rather be myself , Mighty , and to be feared
25	1H4 1H4_1_3	Or you shall hear in such a kind from me As	will displease ye . My Lord Northumberland , We License your departure with
26	1H4 1H4_1_3	King . 1H4 Hotspur :And if the devil come and roar for them I	will not send them . I will after straight And tell him so
27	1H4 1H4_1_3	come and roar for them I will not send them . I	will after straight And tell him so : for I will ease my
28	1H4 1H4_1_3	I will after straight And tell him so : for I	will ease my heart , Although it be with hazard of my head
29	1H4 1H4_1_3	Uncle . Enter Worcester . 1H4 Hotspur :Speak of Mortimer ? Yes , I	will speak of him , and let my soul Want mercy , if
30	1H4 1H4_1_3	my dear blood drop by drop in the dust , But I	will lift the downfall Mortimer As high in the Air , as this
31	1H4 1H4_1_3	mad 1H4 Worcester :Who struck this heat up after I was gone ? 1H4 Hotspur :He	will (forsooth) have all my Prisoners : And when I urged
32	1H4 1H4_1_3	say --- 1H4 Worcester :Peace Cousin , say no more . And now I	will unclasp a Secret book , And to your quick-conceiving Discontents , I
33	1H4 1H4_1_3	my purposes . Those Prisoners you shall keep . 1H4 Hotspur :Nay , I	will ; that 's flat : He said , he would not ransom
34	1H4 1H4_1_3	Mortimer : Forbade my tongue to speak of Mortimer . But I	will find him when he lies asleep , And in his ear ,
35	1H4 1H4_1_3	for diverse reasons Which I shall send you written , be assured	Will easily be granted you , my Lord . Your Son in Scotland
36	1H4 1H4_1_3	it on . 1H4 Hotspur :I smell it : Upon my life , it	will do wondrous well . 1H4 Northumberland :Before the game 's afoot , thou still

Displaying extended context for query match # 10 in text 1H4, at position 1H4_1_2

<< Prev

Next >>

Choose action... ▾

Show tags

Switch to alternative view (Unregularised spelling)

Pilgrims going
to Canterbury with rich Offerings , and Traders riding
to London with fat Purses . I have visors for you
all ; you have horses for yourselves ; Gadshill lies tonight
in Rochester , I have bespoke Supper tomorrow in
Eastcheap ; we may do it as secure as sleep : if you will
go , I will stuff your Purses full of Crowns : if you will
not , tarry at home and be hanged .

1H4_Falstaff:

Hear ye Yedward , if I tarry at home and go not ,
I 'll hang you for going .

1H4_Poins:

You **will** chops .

1H4_Falstaff:

Hal , wilt thou make one ?

1H4_Prince:

Who , I rob ? I a Thief ? Not I .

1H4_Falstaff:

There 's neither honesty , manhood , nor good fellowship
in thee , nor thou camst not of the blood royal ,
if thou darst not stand for ten shillings .

1H4_Prince:

Well then , once in my days I 'll be a madcap .

1H4_Falstaff:

Why , that 's well said .

1H4_Prince:

Well , come what will , I 'll tarry at home .

1H4_Falstaff:

I 'll be a Traitor then , when thou art King .

1H4_Prince:

I care not .

Displaying extended context for query match # 10 in text 1H4, at position 1H4_1_2

<< Prev

Next >>

Choose action...

Show tags

Switch to alternative view (Unregularised spelling)

Pilgrims going
to Canterbury with rich Offerings , and Traders riding
to London with fat Purses . I have visors for you
all ; you have horses for yourselves ; Gadshill lies tonight
in Rochester , I have bespoke Supper tomorrow in
Eastcheap ; we may do it as secure as sleep : if you will
go , I will stuff your Purses full of Crowns : if you will
not , tarry at home and be hanged .

1H4_Falstaff:

Hear ye Yedward , if I tarry at home and go not ,
I 'll hang you for going .

1H4_Poins:

You **will** chops .

1H4_Falstaff:

Hal , wilt thou make one ?

1H4_Prince:

Who , I rob ? I a Thief ? Not I .

1H4_Falstaff:

There 's neither honesty , manhood , nor good fellowship
in thee , nor thou camst not of the blood royal ,
if thou darst not stand for ten shillings .

1H4_Prince:

Well then , once in my days I 'll be a madcap .

1H4_Falstaff:

Why , that 's well said .

1H4_Prince:

Well , come what will , I 'll tarry at home .

1H4_Falstaff:

I 'll be a Traitor then , when thou art King .

1H4_Prince:

I care not .

Displaying extended context for query match # 10 in text 1H4, at position 1H4_1_2

Currently in alternative view (*Unregularised spelling*)

<< Prev

Next >>

Choose action...

Show tags

Leave alternative view

Pilgrimes going
to Canterbury with rich Offerings , and Traders riding
to London with fat Purses . I haue vizards for you
all ; you haue horses for your selues : Gads-hill lyes to night
in Rochester , I haue bespoke Supper to morrow in
Eastcheape ; we may doe it as secure as sleepe : if you will
go , I will stuffe your Purses full of Crownes : if you will
not , tarry at home and be hang'd .

1H4_Falstaff:

Heare ye Yedward , if I tarry at home and go not ,
Ile hang you for going .

1H4_Poins:

You **will** chops .

1H4_Falstaff:

Hal , wilt thou make one ?

1H4_Prince:

Who , I rob ? I a Theefe ? Not I .

1H4_Falstaff:

There 's neither honesty , manhood , nor good fellowship
in thee , nor thou cam'st not of the blood-royall ,
if thou dar'st not stand for ten shillings .

1H4_Prince:

Well then , once in my dayes Ile be a mad-cap .

1H4_Falstaff:

Why , that 's well said .

1H4_Prince:

Well , come what will , Ile tarry at home .

1H4_Falstaff:

Ile be a Traitor then , when thou art King .

1H4_Prince:

I care not .

Displaying extended context for query match # 10 in text 1H4, at position 1H4_1_2

<< Prev

Next >>

Choose action... ▾

Show tags

Switch to alternative view (Unregularised spelling)

Pilgrims going
to Canterbury with rich Offerings , and Traders riding
to London with fat Purses . I have visors for you
all ; you have horses for yourselves ; Gadshill lies tonight
in Rochester , I have bespoke Supper tomorrow in
Eastcheap ; we may do it as secure as sleep : if you will
go , I will stuff your Purses full of Crowns : if you will
not , tarry at home and be hanged .

1H4_Falstaff:

Hear ye Yedward , if I tarry at home and go not ,
I 'll hang you for going .

1H4_Poins:

You **will** chops .

1H4_Falstaff:

Hal , wilt thou make one ?

1H4_Prince:

Who , I rob ? I a Thief ? Not I .

1H4_Falstaff:

There 's neither honesty , manhood , nor good fellowship
in thee , nor thou camst not of the blood royal ,
if thou darst not stand for ten shillings .

1H4_Prince:

Well then , once in my days I 'll be a madcap .

1H4_Falstaff:

Why , that 's well said .

1H4_Prince:

Well , come what will , I 'll tarry at home .

1H4_Falstaff:

I 'll be a Traitor then , when thou art King .

1H4_Prince:

I care not .

Displaying extended context for query match # 10 in text 1H4, at position 1H4_1_2

<< Prev

Next >>

Choose action... ▾

Show tags

Switch to alternative view (Unregularised spelling)

Pilgrims going
to Canterbury with rich Offerings , and Traders riding
to London with fat Purses . I have visors for you
all ; you have horses for yourselves ; Gadshill lies tonight
in Rochester , I have bespoke Supper tomorrow in
Eastcheap ; we may do it as secure as sleep : if you will
go , I will stuff your Purses full of Crowns : if you will
not , tarry at home and be hanged .

1H4_Falstaff:

Hear ye Yedward , if I tarry at home and go not ,
I 'll hang you for going .

1H4_Poins:

You will chops .

1H4_Falstaff:

Hal , wilt thou make one ?

1H4_Prince:

Who , I rob ? I a Thief ? Not I .

1H4_Falstaff:

There 's neither honesty , manhood , nor good fellowship
in thee , nor thou camst not of the blood royal ,
if thou darst not stand for ten shillings .

1H4_Prince:

Well then , once in my days I 'll be a madcap .

1H4_Falstaff:

Why , that 's well said .

1H4_Prince:

Well , come what will , I 'll tarry at home .

1H4_Falstaff:

I 'll be a Traitor then , when thou art King .

1H4_Prince:

I care not .

Displaying extended context for query match # 10 in text 1H4, at position 1H4_1_2

<< Prev

Next >>

Choose action... ▾

Hide tags

Switch to alternative view (Unregularised spelling)

Pilgrims_NN2 going_VVG
to_II Canterbury_NP1 with_IW rich_JJ Offerings_NN2 ,_YCOM and_CC Traders_NN2 riding_VVG
to_II London_NP1 with_IW fat_JJ Purses_NN2 ,_YSTP I_PPIS1 have_VH0 visors_NN2 for_IF you_PPY
all_DB ;_YSCOL you_PPY have_VH0 horses_NN2 for_IF yourselves_PPX2 ;_YCOL Gadshill_NP1 lies_VVZ tonight_RT
in_II Rochester_NP1 ,_YCOM I_PPIS1 have_VH0 bespoke_VVN Supper_NN1 tomorrow_RT in_II
Eastcheap_NP1 ;_YSCOL we_PPIS2 may_VM do_VDI it_PPH1 as_RG secure_JJ as_CSA sleep_NN1 ;_YCOL if_CS you_PPY will_VM
go_VVI ,_YCOM I_PPIS1 will_VM stuff_VVI your_APPGE Purses_NN2 full_JJ of_IO Crowns_NN2 ;_YCOL if_CS you_PPY will_VM
not_XX ,_YCOM tarry_VV0 at_II home_NN1 and_CC be_VBI hanged_VVN ,_YSTP

1H4_Falstaff:

Hear_VV0 ye_PPY Yedward_NP1 ,_YCOM if_CS I_PPIS1 tarry_VV0 at_II home_NN1 and_CC go_VV0 not_XX ,_YCOM
I_PPIS1 'll_VM hang_VVI you_PPY for_IF going_VVG ,_YSTP

1H4_Poins:

You_PPY will_VM chops_NN2 ,_YSTP

1H4_Falstaff:

Hal_NP1 ,_YCOM wilt_VMT thou_PPYS1 make_VVI one_PN1 ?_YQUE

1H4_Prince:

Who_PNQS ,_YCOM I_PPIS1 rob_VV0 ?_YQUE I_PPIS1 a_AT1 Thief_NN1 ?_YQUE Not_XX I_PPIS1 ,_YSTP

1H4_Falstaff:

There_EX 's_VBZ neither_RR honesty_NN1 ,_YCOM manhood_NN1 ,_YCOM nor_CC good_JJ fellowship_NN1
in_II thee_PPYO1 ,_YCOM nor_CC thou_PPYS1 camst_VVDT not_XX of_IO the_AT blood_NN1 royal_JJ ,_YCOM
if_CS thou_PPYS1 darst_VMT not_XX stand_VVI for_IF ten_MC shillings_NN2 ,_YSTP

1H4_Prince:

Well_RR then_RT ,_YCOM once_RR in_II my_APPGE days_NNT2 I_PPIS1 'll_VM be_VBI a_AT1 madcap_NN1 ,_YSTP

1H4_Falstaff:

Why_RRQ ,_YCOM that_DD1 's_VBZ well_RR said_VVN ,_YSTP

1H4_Prince:

Well_RR ,_YCOM come_VV0 what_DDQ will_VM ,_YCOM I_PPIS1 'll_VM tarry_VVI at_II home_NN1 ,_YSTP

1H4_Falstaff:

I_PPIS1 'll_VM be_VBI a_AT1 Traitor_NN1 then_RT ,_YCOM when_CS thou_PPYS1 art_VBT King_NN1 ,_YSTP

1H4_Prince:

I_PPIS1 care_VV0 not_XX ,_YSTP

Displaying extended context for query match # 10 in text 1H4, at position 1H4_1_2

[<< Prev](#)[Next >>](#)

Choose action...

[Show tags](#)[Switch to alternative view \(Unregularised spelling\)](#)

Choose action

Text info for 1H4

[Back to concordance](#)[New query](#)

Pilgrims going
to Canterbury with rich Offerings , and Trave
to London with fat Purses . I have visors for more context
all ; you have horses for yourselves : Gadst
in Rochester , I have bespoke Supper tomo
Eastcheap ; we may do it as secure as sleep
go , I will stuff your Purses full of Crowns : i
not , tarry at home and be hanged .
1H4_Falstaff:
Hear ye Yedward , if I tarry at home and go not ,
I 'll hang you for going .
1H4_Poins:
You **will** chops .
1H4_Falstaff:
Hal , wilt thou make one ?
1H4_Prince:
Who , I rob ? I a Thief ? Not I .
1H4_Falstaff:
There 's neither honesty , manhood , nor good fellowship
in thee , nor thou camst not of the blood royal ,
if thou darst not stand for ten shillings .
1H4_Prince:
Well then , once in my days I 'll be a madcap .
1H4_Falstaff:
Why , that 's well said .
1H4_Prince:
Well , come what will , I 'll tarry at home .
1H4_Falstaff:
I 'll be a Traitor then , when thou art King .
1H4_Prince:
I care not .

Displaying extended context for query match # 10 in text 1H4, at position 1H4_1_2

[<< Prev](#)
[Next >>](#)

Choose action...

[Show tags](#)
[Switch to alternative view \(Unregularised spelling\)](#)

Choose action...

Text info for 1H4

More context

Back to concordance

New query

Pilgrims going
to Canterbury with rich Offerings , and Tra
to London with fat Purses . I have visors for
all ; you have horses for yourselves : Gadst
in Rochester , I have bespoke Supper tomo
Eastcheap ; we may do it as secure as sleep
go , I will stuff your Purses full of Crowns : i
not , tarry at home and be hanged .
1H4_Falstaff:
Hear ye Yedward , if I tarry at home and go not ,
I 'll hang you for going .
1H4_Poins:
You **will** chops .
1H4_Falstaff:
Hal , wilt thou make one ?
1H4_Prince:
Who , I rob ? I a Thief ? Not I .
1H4_Falstaff:
There 's neither honesty , manhood , nor good fellowship
in thee , nor thou camst not of the blood royal ,
if thou darst not stand for ten shillings .
1H4_Prince:
Well then , once in my days I 'll be a madcap .
1H4_Falstaff:
Why , that 's well said .
1H4_Prince:
Well , come what will , I 'll tarry at home .
1H4_Falstaff:
I 'll be a Traitor then , when thou art King .
1H4_Prince:
I care not .

Metadata for text 1H4

Text identification code	1H4
MLA labels	1H4
Short title	Henry IV, Part 1
Simple genre categories	History
Genre categories, taking 'problem plays' into account	H
Date of first publication	1598
Date range of first production	1596-1597
Approximate date of first production	1597
Approximate date of first production, in two bins	1592-1600
Source	First Folio
Source for digital text	Developed from transcriptions provided by Internet Shakespeare Editions.
Play code	1H4
Source edition publication date	1623
No. words in text	29,724

More advanced queries

- Simple Query mode:
 - Search for exact words or phrases
 - E.g. cat
 - Queries using wildcard searches
 - E.g. ca*t, ca?t, ca+t
 - Queries using POS tags
 - E.g. dog_VV*
 - Queries using lemmas (care with this – depends on accuracy!)
 - E.g. {dog}
 - Queries using OR searches
 - E.g. (cat|dog)
- Using freq breakdown:

Quantifier	Explanation
*	Match 0 or more times
+	Match 1 or more times
?	Match 1 or 0 times

Ca*t

No.	Query result	No. of occurrences	Percent
1	Canst	178	37.47%
2	Cast	110	23.16%
3	cat	33	6.95%
4	caught	33	6.95%
5	camst	31	6.53%

Ca+t

No.	Query result	No. of occurrences	Percent
1	Canst	178	40.27%
2	Cast	110	24.89%
3	caught	33	7.47%
4	camst	31	7.01%
5	Capulet	23	5.20%

Ca?t

No.	Query result	No. of occurrences	Percent
1	Cast	110	95.65%
2	Cart	5	4.35%

Uniting queries 1

- E.g. *_JJ* dog
- In other words, any adjective preceding 'dog'



Your query “*_JJ* dog” returned 35 matches in 19 different texts (in 1,038,509 words [38 texts]; frequency: 33.702 instances per million words) [0.088 seconds – retrieved from cache]

<div> < << >> > Show Page: 1 Line view Show in random order Choose action... </div>				
No	Text	Solution 1 to 35 Page 1 / 1		
1	2H4 2H4_1_3	thyselt to cast him up . So , so , (thou	common Dog) didst thou disgorge Thy glutton bosom of the Royal Richard ,
2	2H4 2H4_4_3	from curbed License plucks The muzzle of Restraint ; and the	wild Dog	Shall flesh his tooth in every Innocent . O my poor Kingdom
3	AYL AYL_1_1	becomes me for my good . <i>AYL_Oliver</i> :Get you with him , you	old dog	. <i>AYL_Adam</i> :Is old dog my reward : most true , I have
4	AYL AYL_1_1	good . <i>AYL_Oliver</i> :Get you with him , you old dog . <i>AYL_Adam</i> :Is	old dog	my reward : most true , I have lost my teeth in
5	CE CE_3_2	and my heart of steel , she had transformed me to a	Curtal dog	, & made me turn in the wheel . <i>CE_AntiphoSyracuse</i> :Go hie thee
6	CE CE_5_1	venom clamours of a jealous woman , Poisons more deadly than a	mad dog	's tooth . It seems his sleeps were hindered by thy railing
7	Cor Cor_1_1	especially against Caius Martius . <i>Cor_All</i> :Against him first : He 's a	very dog	to the Commonalty . <i>Cor_2Citizen</i> :Consider you what Services he has done for
8	Cym Cym_2_1	<i>Cym_SecondLord</i> :No my Lord ; nor crop the ears of them . <i>Cym_Clotten</i> :	Whoreson dog	: I gave him satisfaction ? would he had been one of
9	H5 H5_2_1	you shog off ? I would have you solus . <i>H5_Pistol</i> :Solus ,	egregious dog	? O Viper vile ; The solus in thy most marvellous face
10	H5 H5_2_3	Straws , men 's Faiths are Wafer-Cakes , and holdfast is the	only Dog	: My Duck , therefore Caveto be thy Counsellor . Go ,

Uniting queries 2

- E.g. *_JJ* (dog|cat)
- In other words, any adjective preceding 'dog' OR 'cat'



Your query <code>"*_JJ* (dog cat)"</code> returned 39 matches in 21 different texts (in 1,038,509 words [38 texts]; frequency: 37.554 instances per million words) [0.477 seconds]				
<div> < << >> > Show Page: 1 Line view Show in random order Choose action... </div>				
No	Text	Solution 1 to 39 Page 1 / 1		
1	1H4 1H4_3_1	and a moulten Raven , A couching Lion , and a	ramping Cat	, And such a deal of skimble-skamble Stuff , As puts me
2	2H4 2H4_1_3	thyself to cast him up . So , so , (thou	common Dog) didst thou disgorge Thy glutton bosom of the Royal Richard ,
3	2H4 2H4_4_3	from curbed License plucks The muzzle of Restraint ; and the	wild Dog	Shall flesh his tooth in every Innocent . O my poor Kingdom
4	AYL AYL_1_1	becomes me for my good . <i>AYL_Oliver</i> :Get you with him , you	old dog	. <i>AYL_Adam</i> :Is old dog my reward : most true , I have
5	AYL AYL_1_1	good . <i>AYL_Oliver</i> :Get you with him , you old dog . <i>AYL_Adam</i> :Is	old dog	my reward : most true , I have lost my teeth in
6	CE CE_3_2	and my heart of steel , she had transformed me to a	Curtal dog	, & made me turn in the wheel . <i>CE_AntiphoSyracuse</i> :Go hie thee
7	CE CE_5_1	venom clamours of a jealous woman , Poisons more deadly than a	mad dog	's tooth . It seems his sleeps were hindered by thy railing
8	Cor Cor_1_1	especially against Caius Martius . <i>Cor_All</i> :Against him first : He 's a	very dog	to the Commonalty . <i>Cor_2Citizen</i> :Consider you what Services he has done for
9	Cym Cym_2_1	<i>Cym_SecondLord</i> :No my Lord ; nor crop the ears of them . <i>Cym_Clotten</i> :	Whoreson dog	: I gave him satisfaction ? would he had been one of
10	H5 H5_2_1	you shog off ? I would have you solus . <i>H5_Pistol</i> :Solus ,	egregious dog	? O Viper vile ; The solus in thy most marvellous face

Have a go

- Try searching for the following things – how many results do you get?:
- ‘love’
 - As a noun
 - As a verb
 - Followed by a noun (hint: can include proper nouns!)
- ‘fool’
 - As a noun
 - As a lemma
 - Pre-modified by an adjective



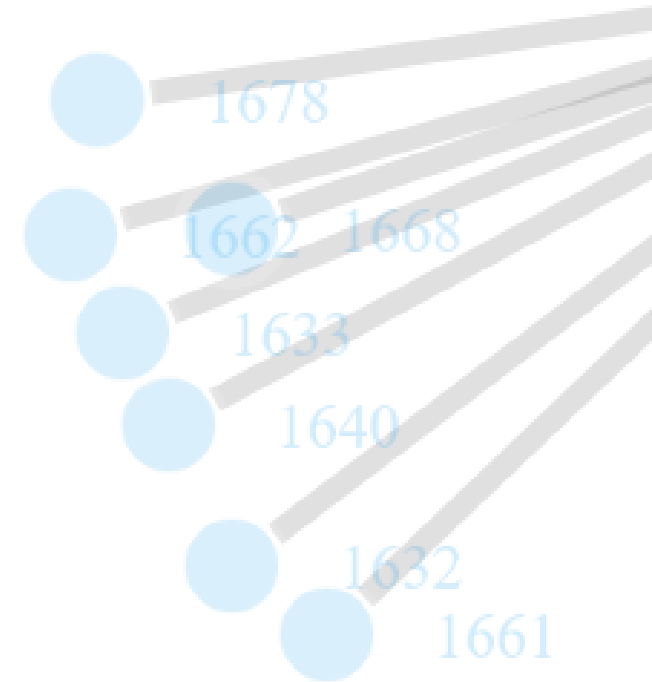
Answers

- Try searching for the following things – how many results do you get?:
- ‘love’
 - As a noun
 - As a verb
 - Followed by a noun (can include proper nouns!)
- ‘fool’
 - As a noun
 - As a lemma
 - Pre-modified by an adjective

Option 1	Option 2	Result for option 1	Result for option 2
love_N*	love_NN*	1,428	1,428
love_V*	love_VV*	625	625
love *_N*	love *_NN*	62	33
fool_N*	fool_NN*	388	388
{fool}	NA	506	-
_JJ fool	NA	71	-

Collocation

- Frequent or habitual co-occurrence of two words/lexemes
 - E.g.
 - ‘fish and ...’
 - ‘save ...’
- A more formal definition:
- ‘*COLLOCATION is the relation of co-occurrence between an obligatory core word or phrase (the node) and individual COLLOCATES: word-tokens which are directly observable and countable in texts.*’ (Stubbs)
- Number of uses: linguistic analysis, lexicography, etc



Collocation

Min. freq cut off of 5
Stat used: Log Ratio

Showing collocate 'life' for node 'save'

- The co-occurrence or relationship between two tokens/words can be measured in different ways
- We set a span to the left and right of the **node**/word of interest

For query "save" returned 219 matches in 38 different texts (in 1,038,509 words [38 texts]; frequency: 210.879 instances per million words), sorted on position +1 (219 hits), collocating with life (21 hits) [0.076 seconds]

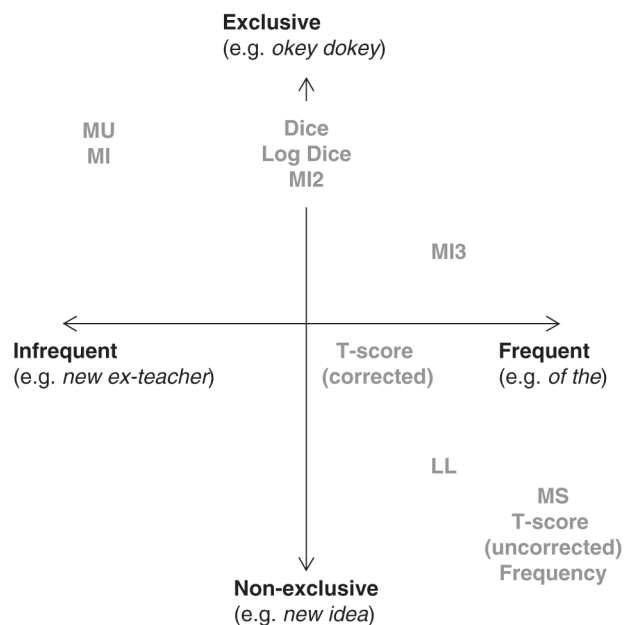
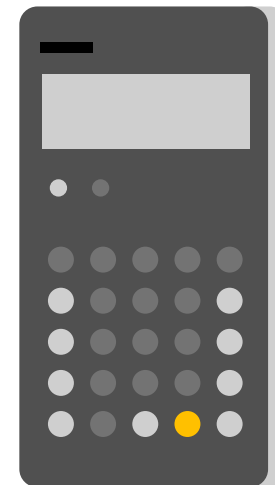
Text	Solution 1 to 21	Page 1 / 1
1H6 1H6_3_2	Talbot ? 1H6_Falstaff :Aye, all the Talbots in the world, to	save my life . Exit . 1H6_Captain :Cowardly Knight, ill fortune follow thee
1H6 1H6_4_6	On that advantage, bought with such a shame, To	save a paltry life, and slay bright Fame, Before young Talbot
2H6 2H6_3_1	that were no policy: The King will labour still to	save his Life, The Commons haply rise, to save his Life
2H6 2H6_3_1	still to save his Life, The Commons haply rise, to	save his Life; And yet we have but trivial argument, More
2H6 2H6_4_7	fare with your departed souls, And therefore yet relent, and	save my life . 2H6_Cade :Away with him, and do as I command
AW AW_4_1	on To gather from thee. Haply thou mayst inform Something to	save thy life . AW_Paroles :O let me live, And all the secrets
AYL AYL_3_5	in you than in the ordinary Of Nature's sale-work? God	save my little life, I think she means to tangle my eyes
CE CE_5_1	I bestrid thee in the wars, and took Deep scars to	save thy life; even for the blood That then I lost for
CE CE_5_1	vouchsafe me speak a word: Haply I see a friend will	save my life, And pay the sum that may deliver me.
Cor Cor_4_5	Hearth, not out of Hope (Mistake me not) to	save my life: for if I had feared death, of all
Cor Cor_5_4	not us. Enter a Messenger. Cor_Messenger :Sir, if you'd	save your life, fly to your House, The Plebeians have got
H5 H5_4_4	cent ecus. H5_Pistol :What are his words? H5_Boy :He prays you to	save his life, he is a Gentleman of a good house,
LLL LLL_2_1	your eye. LLL_Rosaline :No point, with my knife. LLL_Berowne :Now God	save thy life. LLL_Rosaline :And yours from long living. LLL_Berowne :I can not
LLL LLL_4_2	duty, adieu. LLL_Jaquenetta :Good Costard go with me: Sir God	save your life. LLL_Dull :Have with thee my girl. Exit. LLL_Holofernes :Sir
MA MA_5_4	good day, I yield upon great persuasion, & partly to	save your life, for I was told, you were in a
MM MM_2_4	upon that pain. MM_Isabella :True. MM_Angelo :Admit no other way to	save his life (As I subscribe not that, nor any other
MM MM_5_1	And you may marvel, why I obscured myself, Labouring to	save his life: and would not rather Make rash remonstrance of my
R3 R3_4_3	R3_Richard :Wrong not her Birth, she is a Royal Princess. R3_QueenElizabeth :To	save her life, I'll say she is not so. R3_Richard :Her
TS TS_1_3	therefore frame your manners to the time Your fellow Tranio here to	save my life, Puts my apparel, and my countenance on,
TS TS_1_3	you, as becomes: While I make way from hence to	save my life: You understand me? TS_Biondello :I sir, never a
TS TS_4_2	as an apple doth an oyster, & all one. TS_TraAsLucentio :To	save your life in this extremity, This favour will I do you

← Node →

SPAN: 3 words to left and right

Quick note on statistics

- We can use different statistics in collocation
- They can be used for different things (again depending on our research priority)



Statistic ID	Statistic name	Statistic cut-off value	L and R span	Minimum collocate freq. (C)	Minimum collocation freq. (NC)	Filter
4b	MI2	3	L5-R5	5	1	function words removed

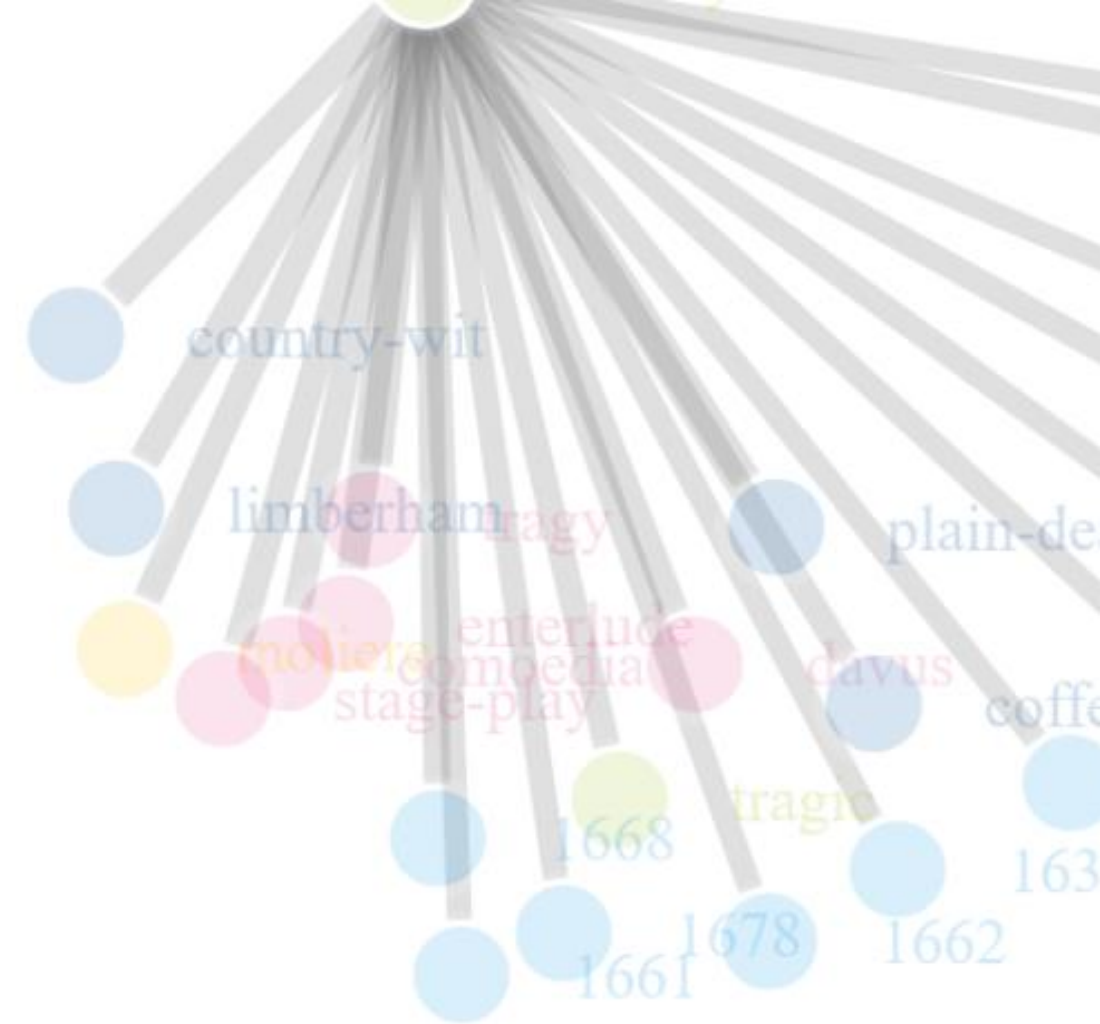
4b-MI2(3), L5-R5, C5-NC1; function words removed

Example

CPN or equivalent is highly encouraged - readers know **exactly** what parameters were used and can **replicate**

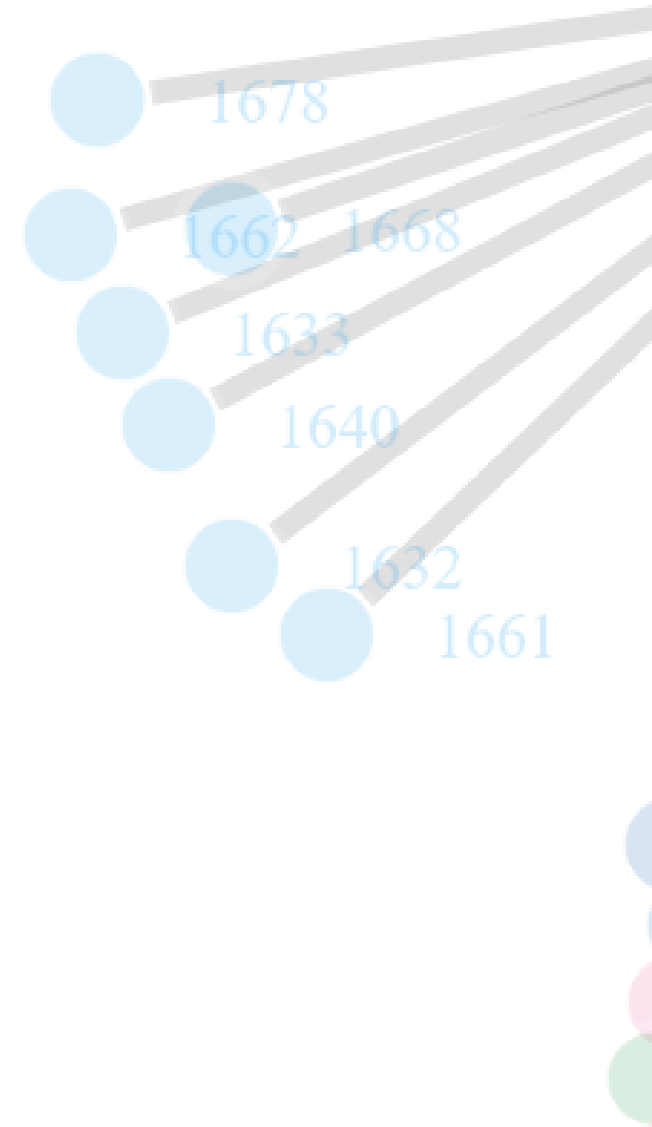
Let's explore some in our corpus

- 'thou'
- 'you'
- 'devil'
- Are there any interesting collocations you spot?
- Any patterns?
- Use default Log Ratio (filtered) as a starting point (finds more exclusive words associated with our node) – feel free to try changing the stats/span and see how this changes your results!



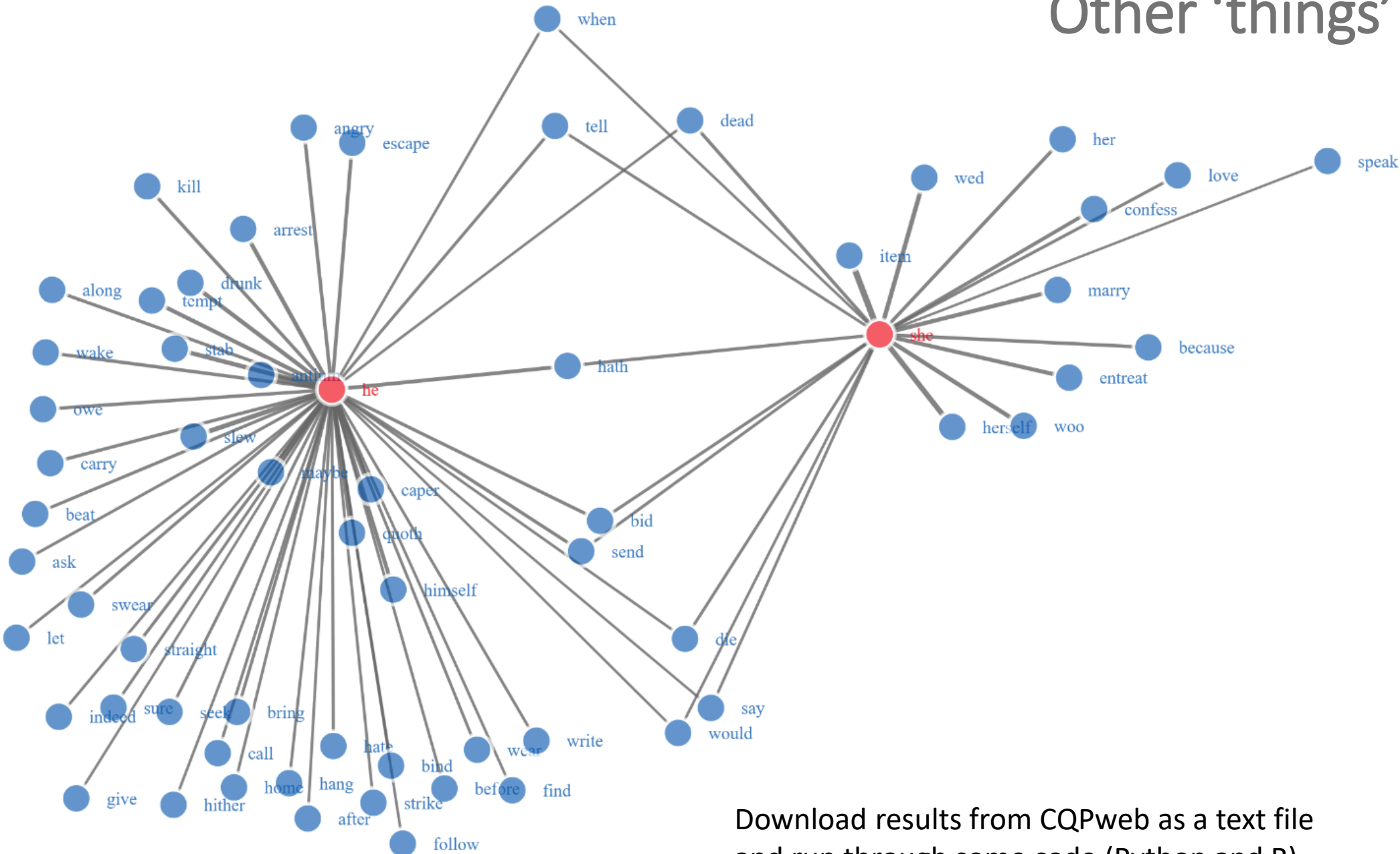
Restricted query

- Collocations between different social classes for 'you'
 - Some things we might have observed:
 - 'Rogue' vs. 'whoreson'
 - Politeness markers – pray/beseech/please



Other 'things' we can do:

Statistic	Log Ratio (filtered)
Statistic cut-off value	1
L and R span	L3-R3
Minimum collocate freq (C)	5
Minimum collocate freq (NC)	5
Notes/filter	Lemma for both node and collocate
Query	{she} {he}



Conclusion

- Corpus linguistics provides other perspectives on texts
- Important to ensure the corpus and queries match the research concerns/question
- Ensure that all stages of the process are included in outputs for replication purposes (where possible – data)
- These are just **some** of the methods in corpus linguistics for more, see the MOOCs at Lancaster University (Shakespeare's Language and Corpus Linguistics)

