



Debunking myths about Shakespeare's language with corpus methods

*Jonathan Culpeper,
Lancaster University, UK*

@ShakespeareLang



Arts & Humanities
Research Council



THE QUEEN'S
ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2015



Acknowledgements

- Jane Demmen
- Sheryl Banas
- And the Encyclopaedia of Shakespeare's Language project folks: <http://wp.lancs.ac.uk/shakespearelang/people/>

Coming up ...

- Setting the scene: Encyclopaedia of Shakespeare's Language project
- What is a language myth?
- Myths about Shakespeare's language:
 1. Shakespeare's language is (wholly) Shakespeare's language (Gary Taylor, authorship attribution and corpus linguistics)
 2. Shakespeare had a larger vocabulary than any other writer (Ward & Valenza, and Craig)
 3. Shakespeare coined more words than any other writer (with a discussion of word counting)
 4. Shakespeare's language transcends time and space (with a discussion of the word 'good')
- A brief coda: Shakespeare -- the creative genius

What the project aims to do

- Produce the first systematic and comprehensive account of Shakespeare's language using methods derived from corpus linguistics – an approach that uses computers in large-scale language analysis.

COLLINS
COBUILD
ENGLISH
LANGUAGE
DICTIONARY

LEARN AND USE ENGLISH
WITH THIS UNIQUE NEW DICTIONARY

LONGMAN
GRAMMAR
of SPOKEN
and WRITTEN
ENGLISH

Douglas Biber
Stig Johansson
Geoffrey Leech
Susan Conrad
Edward Finegan

Copyrighted Material

**Early English
in the Computer Age**

Explorations through the
Helsinki Corpus

Matti Rissanen, Merja Kytö,
Minna Palander-Collin
(Editors)

TiEL

Mouton de Gruyter

Copyrighted Material

Corpus Stylistics

Speech, writing and thought
presentation in a corpus
of English writing

Elena Semino and
Mick Short

What will be in the encyclopedia?

Volume 1 (a kind of dictionary)

Focuses on the use and meanings of each of Shakespeare's words, both in the context of what he wrote and in the context in which he wrote.

Every word is, for example, compared with a 321 million word corpus comprising the work of Shakespeare's contemporaries.

What will be in the encyclopedia?

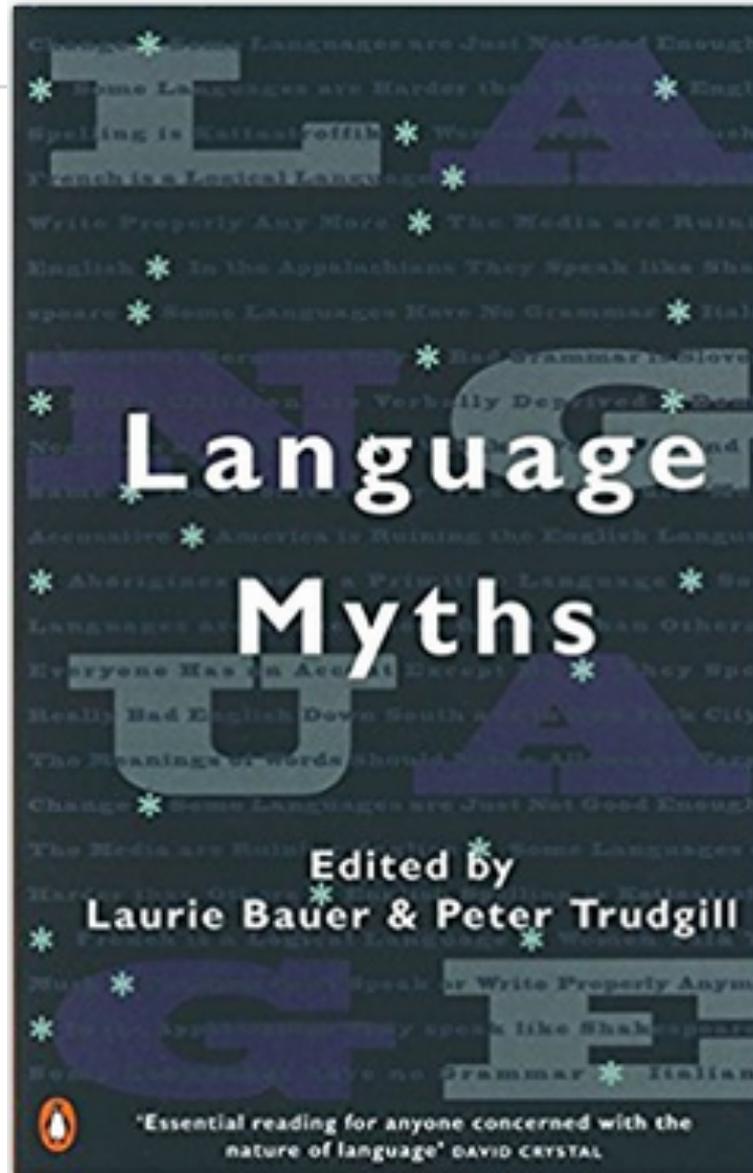
Volume 2 (a compendium of semantic patterns)

Focuses on patterns of words in Shakespeare's writings. It describes how these patterns create the 'linguistic thumbprints' of characters, different genders, themes, plays and dramatic genres. It also considers clusters of words that relate to concepts (e.g. love, death).

Volume 3 (a kind of grammar)

Focuses on grammatical words and patterns.

What is a myth about language?



What is a myth about language?

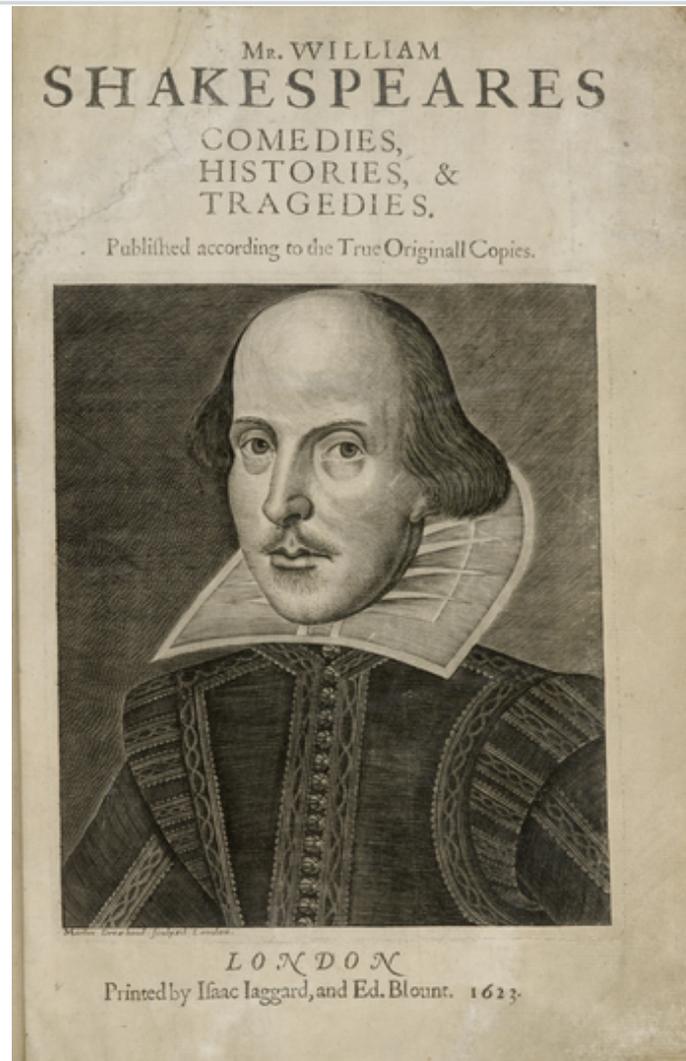
Language myths are:

Beliefs about language that are produced and reproduced within particular communities, and become part of a cultural ideology or “folk wisdom” that is used to evaluate language and account for how it is. They are inconsistent with empirically observed linguistic “facts”.

Myth1: Shakespeare's language is (wholly) Shakespeare's language

- No clear authorial oversight. 36 plays were put together and published in 1623 as the First Folio – Shakespeare had been dead for 7 years.

Myth 1: Shakespeare's language is (wholly) Shakespeare's language



Myth1: Shakespeare's language is (wholly) Shakespeare's language

- No clear authorial oversight. 36 plays were put together and published in 1623 as the First Folio – Shakespeare had been dead for 7 years.
- 18(?) plays had been previously published as Quartos, but some are considered “bad” (perhaps reconstructed from memory)
- An early play-text was a bundle of manuscript fragments written for performance, rather than a unitary whole written for publication

Myth1: Shakespeare's language is (wholly) Shakespeare's language

- Collaboration amongst playwrights is known to have been very common.
- Plagiarism is a modern notion; re-using portions of text from elsewhere could be construed as complimentary.
- So, “Shakespeare's language” = “surviving written texts that purport to represent, for the most part, the language that Shakespeare produced for his ‘literary’ works”

Myth1: Shakespeare's language is (wholly) Shakespeare's language

- The contribution made by other authors to “Shakespeare’s” works? And vice versa.
- Authorship attribution.
- Gary Taylor & Gabriel Egan (2016). *The New Oxford Shakespeare*. Christopher Marlowe credited as co-author of *Henry VI* plays, Thomas Middleton as co-author of *All's Well That Ends Well*; *Arden of Faversham* added to Shakespeare's 'çanon'.
- What was the basis for these decisions?
- N-grams.

Myth1: Shakespeare's language is (wholly) Shakespeare's language

“I will finish this lecture shortly”

I will

will finish

finish this

this lecture

lecture shortly

I will finish

will finish this

finish this lecture

this lecture shortly

Myth1: Shakespeare's language is (wholly) Shakespeare's language

- Does the disputed text have more n-gram types (i.e. different n-grams) in common with Shakespeare's works or Marlowe's?
- Does the disputed text favour the n-grams in the way that Shakespeare's works do or Marlowe's?
- Etc.

But such authorship attribution studies do not account for **meaning**.

N-grams in Shakespeare and beyond

Shakespeare	EModE Plays	Present-day Plays
I pray you I will not I know not I am a I am not my good lord there is no I would not it is a and I will	it is a what do you and I will it is not I have a I will not in the world I tell you I know not I warrant you	I don't know what do you I don't want do you think do you want I don't think to do with do you know going to be don't want to

Three-word N-grams in order of frequency (coloured items appear in another column)

Data in 2nd and 3rd columns draw from Culpeper and Kytö (2010)

N-grams in Shakespeare and beyond



N-grams in Shakespeare and beyond

Frank What I want to know is what is it that's suddenly led you to this?

Rita What? Comin' here?

Frank Yes.

Rita It's not sudden.

Frank Ah.

Rita I've been realizin' for ages that I was, y' know, slightly out of step. I'm twenty-six. I should have had a baby by now; everyone expects it. I'm sure me husband thinks I'm sterile. [...]

Willy Russell, *Educating Rita*, 1981, p.8

N-grams in Shakespeare and beyond



Purpose-built outdoor theatres:

The Theatre (1576),
The Curtain (1577),
The Rose (1587),
The Swan (1595),
The Globe (1599), and
The Fortune (1600).

Myth 2: Shakespeare had a larger vocabulary than any other writer

An educated adult monolingual today: most studies approx.
9,000 – 18,000 words (Treffers-Daller, J. & Milton, J., 2013, *Applied Linguistics Review*, 4(1): 151-172)

- “[...] astonishing vocabulary of some 25,000 words” (Greenblatt, S. 2008, *The Norton Shakespeare*, p.65)
- “Twice as large as an educated person today ... 30,000” (McCrum, R. et al. (2002, *The Story of English*, p.102)

Myth 2: Shakespeare had a larger vocabulary than any other writer

- Vocabulary of 20,000 words “*was a large vocabulary in its day*” (Crystal, D., 2008, *Think on my Words: Exploring Shakespeare’s Language*, p.2)
- “Of Shakespeare it may be said without fear of exaggeration that his contribution to our phraseology is ten times greater than that of any writer to any language in the history of the world” (Weekley, E. 1952 [1928] *The English Language*, p. 55).

Myth 2: Shakespeare had a larger vocabulary than any other writer

Problem: Counting assumptions

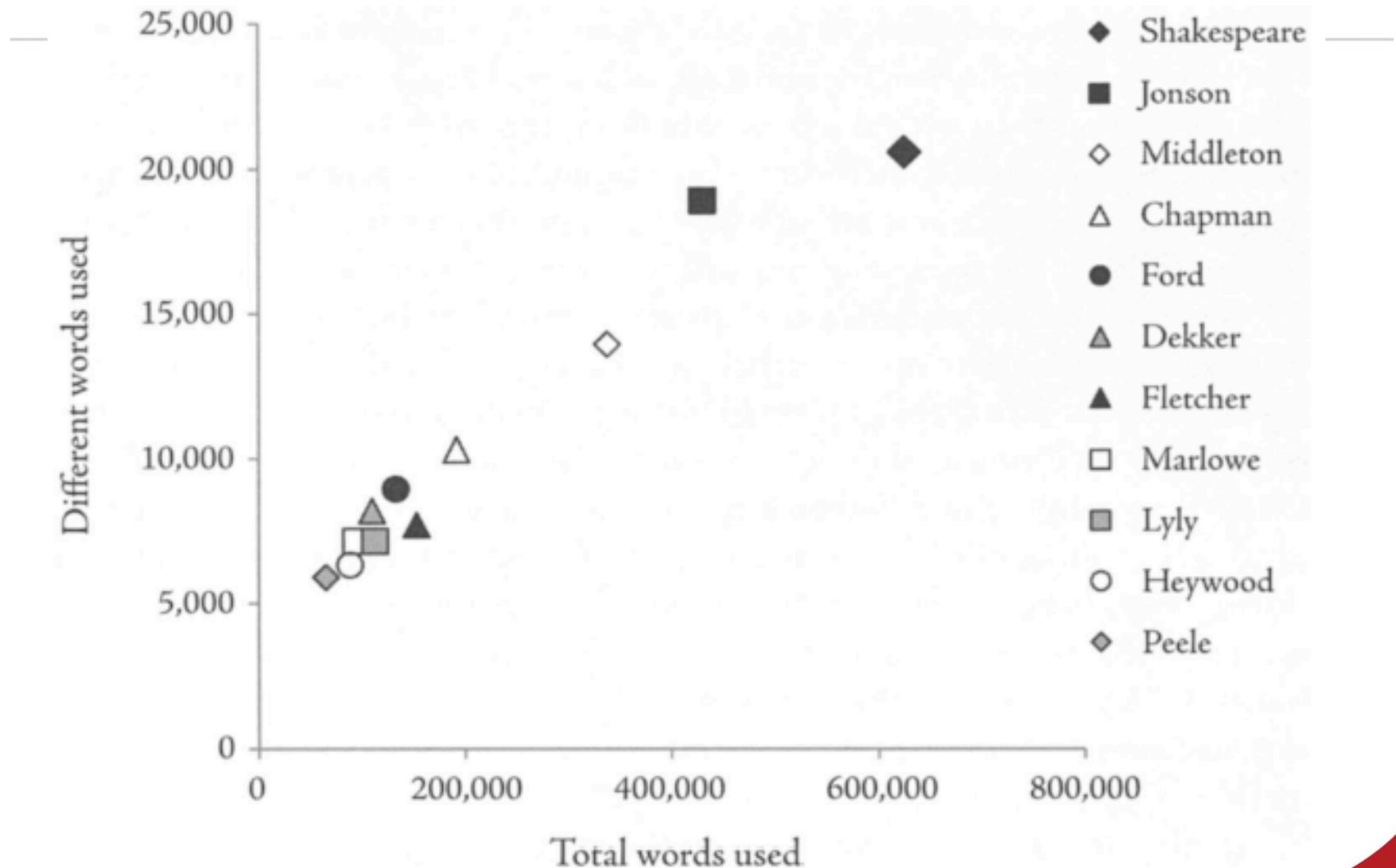
Author	Total different words (approx.)
Shakespeare	20,000

Myth 2: Shakespeare had a larger vocabulary than any other writer

Problem: Counting assumptions

Author	Total different words (approx.)
Shakespeare	20,000
Jonson	18,500
Peele	6,000

Hugh Craig (2011) Shakespeare's Vocabulary: Myth and Reality, *Shakespeare Quarterly*, 62(1), p.61



Hugh Craig (2011) Shakespeare's Vocabulary: Myth and Reality, *Shakespeare Quarterly*, 62(1), p.62

Table 1: Number of different words used in the first 10,000 words of plays¹

Author	Segments	No. of different words		
		Lowest	Highest	Average
Webster	3	1,694	1,957	1,827.0
Dekker	5	1,502	2,011	1,772.0
Peele	4	1,649	1,845	1,749.5
Marlowe	5	1,554	1,924	1,740.8
Jonson	12	1,494	1,964	1,727.1
Greene	4	1,512	1,829	1,675.5
Shakespeare	28	1,426	1,930	1,663.5
Lyly	8	1,497	1,688	1,604.5
Chapman	10	1,281	1,783	1,581.3
Heywood	5	1,389	1,769	1,554.6
Middleton	12	1,308	1,655	1,550.8
Fletcher	7	1,394	1,682	1,525.5
Wilson	3	1,333	1,710	1,499.0

¹ Analysis is by author, arranged by average count, largest value first.

Myth 3: Shakespeare coined more words than any other writer

- How deep and apparently inexhaustible were the wells of his memory and invention, and how marvellous his aptitude for word-coining, are evident from the addition of 302 peculiar words in writing *Troilus and Cressida* (Alfred Hard, 1943, *The Review of English Studies*, 19(75): 254)

Myth 3: Shakespeare coined more words than any other writer (contd.)

What can we 'learn' from the internet?

- “Shakespeare coined more words than other writers, around 1700 words ...”
- “The English language owes a great debt to Shakespeare. He invented over 1700 of our common words ...”
- “Shakespeare introduced nearly 3,000 words ...”
- “Shakespeare is credited by the Oxford English Dictionary with the introduction of nearly 3,000 words into the language”
- “Shakespeare invented a quarter of our language”
- “Shakespeare invented half the words in the English language”
- “Shakespeare *is* our language”

Myth 3: Shakespeare coined more words than any other writer (contd.)

The problem with counting words

- a) Defining a word
- b) Textual starting point
- c) Word-forms and lexemes
- d) Spelling variation

Myth 3: Shakespeare coined more words than any other writer (contd.)

Defining a word

a) A phoneme or phonemes bounded by pauses?

Tybalt: Gentlemen, **good den**, a word with one of you.

Romeo and Juliet, III.1

b) A unit of meaning?

➤ *The plane landed* = 3 words?

➤ *The plane took off* = 3 words? (cf. phrasal verbs)

➤ *He kicked the bucket* = 2 words? (cf. idioms)

Polonius: **God buy you**; fare you well.

Hamlet II.1

Myth 3: Shakespeare coined more words than any other writer (contd.)

Goodbye < god be with ye

Good buy (i.e. redeem, save) *you*

c) A minimal free form?

- Leonard Bloomfield (1887-1949): words = minimal free forms
- minimal free forms = the smallest units that can meaningfully stand on their own (i.e. be a complete utterance)
- *The, of, to, at, by, etc.?*

Present-day *gonna* < *going to* (BNC “gon-na”)

Alarum – French *à l’arme*; Italian *all’arme*

a larum > *alarum* > *larum*

Myth 3: Shakespeare coined more words than any other writer (contd.)

d) A letter or letters bounded by spaces?

orthographic word = 'a string of uninterrupted non-punctuation characters with white space or punctuation at each end' (Leech et al. 2001: 13-14)

A 'natural' division?

Myth 3: Shakespeare coined more words than any other writer (contd.)

To mummy dabb

I like you and I love you

Love from

Emily

Myth 3: Shakespeare coined more words than any other writer (contd.)

hourglass

hour-glass

hour glass

Myth 3: Shakespeare coined more words than any other writer (contd.)

Shakespeare: The textual starting point

Problem: Modern editions of Shakespeare are edited loose collations of the Folio and Quartos, mixed with a liberal dose of editorial licence.

Solution: Have as our base the First Folio with original spelling, and, specifically, the 'diplomatic' transcription (i.e. a faithful warts and all transcription) produced by *Shakespeare Internet Editions* (<http://internetshakespeare.uvic.ca/Foyer/plays/>).

Problem: A faithful transcription is more of an ideal than a reality.

Myth 3: Shakespeare coined more words than any other writer (contd.)

Some typical transcription errors

- y^t ("that") (AW, II:i)

1. Lo. G. Oh my sweet Lord y you wil stay behind vs.

`<u norm="1 Lord" label="1. Lo. G"> Oh my sweet Lord CyC you , wil stay behind vs.</u>`

- Space-saving formatting conventions (H8, II:ii)

In the vnparciall iudging of this Businesse. **ted**
Kin. Two equall men : The Queene shall be acquaint-
 Forthwith for what you come. Where's *Gardiner?*

`<u norm="King Henry" label="Kin."> Two equall men: The Queene shall be acquaintForthwith<lb type="inline"/> for what you come. Where's Gardiner? ted<lb/></u>`

Myth 3: Shakespeare coined more words than any other writer (contd.)

Some typical transcription errors

- Mis-tagged text (1) *(Ham IV:v)*

**And of all Christian Soules, I pray God.
God buy ye. *Exeunt Ophelia***

Soule.<lb/><lb/> And of all Christian Soules, I pray God. </u>
<stage>**God buy ye.** Exeunt Ophelia</stage>

"God buy ye" is part of the soldier's dialogue, not stage direction.

Myth 3: Shakespeare coined more words than any other writer (contd.)

Some typical transcription errors

- Mis-tagged text (2) (MM III:i)

Duke. It is a rupture that you may easily heale: and the cure of it not onely saues your brother, but keepes you from dishonor in doing it.
Isab. Shew me how (good Father.)
Duk. This fore-named Maid hath yet in her the con-

<u norm="Duke" label="Isab."> Shew me how (good Father.)</u>

The speaker tag should be “Isabella”, as the Folio label indicates.

Myth 3: Shakespeare coined more words than any other writer (contd.)

Word-forms and lexemes

- Dictionary headword:

do

- Modern (morphological) word-forms:

do, does, doing, did, done

- Early modern (morphological) word-forms:

do, does, do(e)st, doth, doing, did, didst, done

Myth 3: Shakespeare coined more words than any other writer (contd.)

Word-forms and lexemes

Dictionary headword:

do = **1**

Modern (morphological) word-forms:

do, does, doing, did, done = **5**

Early modern (morphological) word-forms:

do, does, do(e)st, doth, doing, did, didst, done = **8**

Myth 3: Shakespeare coined more words than any other writer (contd.)

Spelling variation:

Problem: You decide to study the use of the word *would* in a corpus. You type it into your search program ... and look at the result.

But you miss: *wold, wolde, woolde, wuld, wulde, wud, wald, vvould, vvold*, etc., etc.

One orthographic word today; many in EModE.

Solution: *Variant Detector* (VARD) program, primarily devised by generations of scholars at Lancaster, but most recently given a significant boost by Alistair Baron.

Further problem: What do you regularize the spelling to? There is no standardised regular form in the way that there is today.

Myth 3: Shakespeare coined more words than any other writer (contd.)

Solution: Our policy was to

- Preserve the morphology, e.g. 2nd + 3rd person verb inflections (-(e)st, -(e)th), past tense forms (e.g. *holp*), past participle forms (e.g. *holpen*), plural forms (e.g. *shooen*), non-standard superlatives (e.g. *horrider*), and *you/thou*,
- Only use a spelling that had EModE currency.
- Prioritize the most frequent spelling in Shakespeare

Myth 3: Shakespeare coined more words than any other writer (contd.)

Work on neologisms (with Sheryl Banas):

- 1,502 words recorded in the Oxford English Dictionary as first citations in Shakespeare
- We are checking these in *Early English Books Online* (EEBO-TCP): 125,000 titles of printed material, 1473-1700. About 1.2 billion words.

Preliminary findings:

- If the current pattern continues, less than a quarter of those 1,502 words can reasonably be attributed to Shakespeare.

Myth 3: Shakespeare coined more words than any other writer (contd.)

Issues

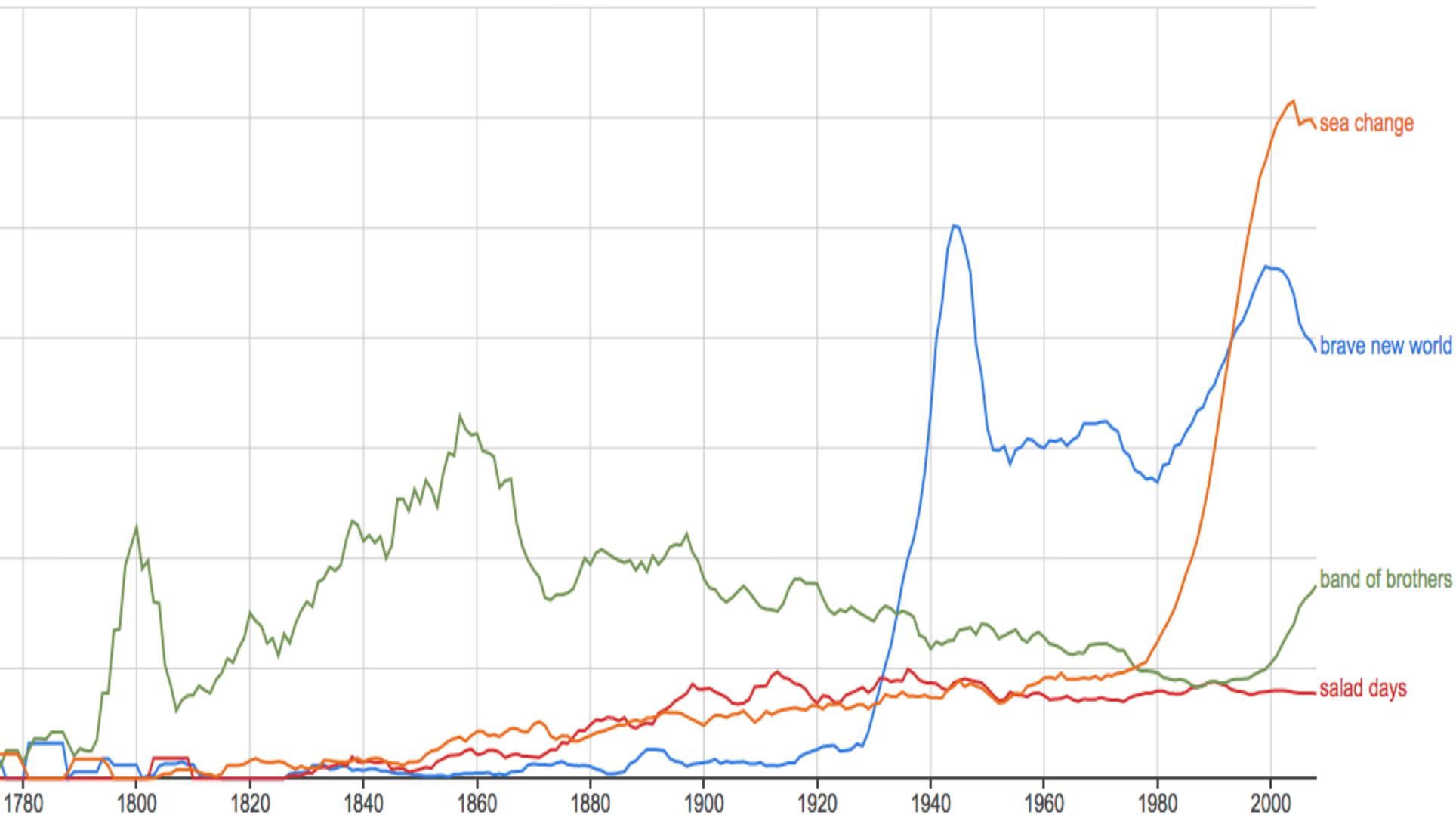
- How do we know that Shakespeare coined it as opposed to recorded it? Cf. *down staires* vs. *incarna[r]dine* (v.)
- What about borrowings, such as Latin *acerb[ic]*, that appear in mixed Latin-English texts before Shakespeare?
- Is it actually just a nonce word rather than neologism? Cf. *dropsied* vs. *domineering*

Myth 3: Shakespeare coined more words than any other writer (contd.)

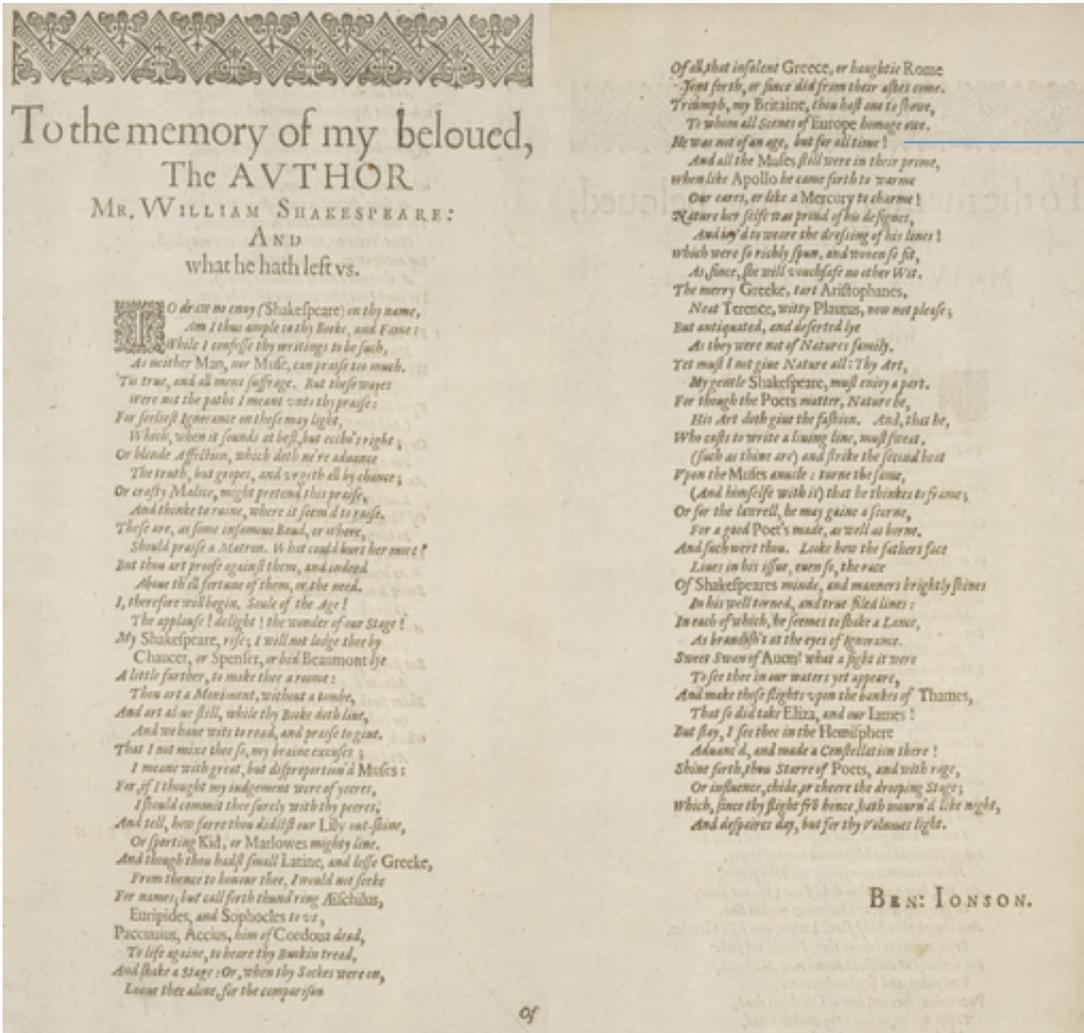
Do Shakespeare's coinages survive into today's English?

- A glimpse of phrases first recorded in Shakespeare and their more recent life.

Four phrases first recorded in Shakespeare and their use in printed material over the last 200 years (Google's N-Gram Viewer)



Myth 4: Shakespeare's language transcends time and space



He was not of an age, but for all time!

Myth 4: Shakespeare's language transcends time and space

- Universal characters, emotions, themes language??

“Shakespeare has given us a universal language medium in which are crystalized the battle hymns, the intellectual conceptions and the spiritual aspirations of the Anglo-Saxons.”

(Rutherford, N.J. and Bennett, E.H., 1918-1922, *English Speaking World*, Vol.2
(8): 14)

- No Shakespearean dictionary has treated Shakespeare's language as relative, i.e. put Shakespeare's usage in the context of that of his contemporaries.

Myth 4: Shakespeare's language transcends time and space - *good*

Crystal & Crystal (2004:201-202):

(1) [intensifying use] real, genuine ('love no man in good earnest'). (2) kind, benevolent, generous. (3) kind, friendly, sympathetic. (4) amenable, tractable, manageable. (5) honest, virtuous, honourable. (6) seasonable, appropriate, proper. (7) just, right, commendable. (8) intended, right, proper. (9) high-ranking, highborn, distinguished. (10) rich, wealthy, substantial.

+ phrases and compounds

good (*n.*) 1 goodwill, goodness HB IV.ii.60 [Griffith to Katherine, of the fall of Walsley's foundation at Ipswich] Unwilling to survive the good that did it
2 good fellow Per II.Chorus.9 [Gower alone, of Pericles] The good in conversation ... / Is still at Tarsus; Tem I.i.3, 15; Tim II.ii.75
3 interest, advantage, benefit Cor II.ii.234 [Sicinius to Brutus, of Coriolanus] It shall be to live them as our good wills, / A sure distraction

good, do prosper, succeed, triumph 2H6 IV.iii.14 [Dick to Cade] If we mean to thrive and do good, break open the gaols

good, do one 1 make prosper, enable to succeed MA I.i.269 [Claudio to Don Pedro] your highness now may do me good; 2H6 I.ii.77; 3H6 III.ii.38
2 be of use to, provide assistance to 2H6 III.ii.219 [King to all, of Gloucester] myself ... / Look after him, and cannot do him good

good (*adj.*) 1 [intensifying use] real, genuine Ar I.ii.26 [Celia to Rosalind] love no man in good earnest; Tim II.ii.232
2 kind, benevolent, generous Tim I.ii.242 [Timon to Apemantus] if thou wert not stollen, I would be good to thee; Tim III.ii.24
3 kind, friendly, sympathetic 2LV.ii.274 [Rosaline to all] The King was wont to rise for a good word
4 amenable, tractable, manageable TS I.i.127 [Hortensio to Gremio, of a husband for Katherine] there he good follows in the world, as a man could fight on them
5 honest, virtuous, honourable Cor II.ii.58 [Menenius to Brutus and Sicinius, of their supporters] they lie deadlier than tell you have good faces
6 seasonable, appropriate, proper CE II.ii.70 [Antipholus of Syracuse to Dromio of Syracuse] learn to jest in good time
7 just, right, commendable 2H6 II.ii.28 [Suffolk to Gloucester, of malice] so sure than well because / So good a quarrel
8 intended, right, proper R3 I.iv.46 [Mercutio to Romeo] Take our good meaning
9 high-ranking, highborn, distinguished R2 I.i.40 [Bolingbroke to Mowbray] They are ... a resolute, / Too good to be so ~~so~~ better (*adj.*)
10 rich, wealthy, substantial Cor I.ii.15 [First Citizen to all] We are accosted poor citizens, the partricks good; 2H4 IV.ii.15; MW I.ii.12

good, 'tis very well Th I.ii.12 [Saturninus to Bassianus] 'Tis good, sir. You are very short with us

good (*adv.*) [intensifying use] really, genuinely 2H4 III.ii.44 [Falstaff to Bardolph] the sack that thou hast drunk we would have bought we might as good cheap at the dearest chandler's in Europe

good-conceited (*adj.*) cleverly devised, ingeniously composed Cym II.ii.16 [Cloten to musicians, of their playing to Imogen] a very excellent good-conceited thing

good deed (*adv.*) in truth, in reality WT I.ii.42 [Hermione to Leontes] yet, good deed, Leontes, / I love thee not a jar o'th' clock behind / What lady she her lord ~~so~~ DISCOURSE MARKERS, p.127

good-den / e'en / even / morrow ~~so~~ GREETINGS, p.206; FAREWELLS, p.170

good-faced (*adj.*) smooth-faced, pretty WT IV.ii.112 [Autolycus to Clown] good-faced sir ... sweet sir

good life comfortable position, respectable way of life MW III.ii.112 [Mistress Page to Mistress Ford] defend your reputation, or bid farewell to your good life for ever ~~so~~ song of good life

Myth 4: Shakespeare's language transcends time and space - *good*

Crystal & Crystal (2004:201-202):

- (1) [intensifying use] real, genuine ('love no man in good earnest').
- (2) **kind**, benevolent, generous.
- (3) **kind**, friendly, sympathetic.
- (4) amenable, tractable, manageable.
- (5) honest, virtuous, honourable.
- (6) seasonable, appropriate, **proper**.
- (7) just, **right**, commendable.
- (8) intended, **right**, **proper**.
- (9) high-ranking, highborn, distinguished.
- (10) rich, wealthy, substantial.

Myth 4: Shakespeare's language transcends time and space - *good*

good**** adj. (good, better, best):

1. A polite address: '(my) good Lord/friend/Sir/Master/Lady/Madam/etc.'. Typically used when meeting or parting, thanking or making suggestions. *But (good my Lord) do it so cunningly* TGV, III. 1.
2. Honest, truthful, principled; of high moral standards. (This sense also shapes the discourse markers '(in) good faith/sooth/troth', which mean truly or honestly). *a man of good repute, carriage, bearing, & estimation* LLL, I. 1.
3. Positive rather than negative. Typically, contrasted with 'bad'. *Is thy news good or bad?* ROM, II. 5.
4. In one's favour, especially favourable wishes or blessings. *The Gods be good to us* COR, V. 4.
5. A welcoming, cheerful manner. *Therefore for Gods sake entertain good comfort, And cheer his Grace with quick and merry eyes* R3, I. 3.

A brief coda: Shakespeare -- the creative genius

Antony: Let Rome in Tiber melt, [...]

(Antony and Cleopatra I.i)

Concluding reflections

Shakespeare and the myths. Why?

“Shakespeare has given us a universal language medium in which are crystalized the battle hymns, the intellectual conceptions and the spiritual aspirations of the Anglo-Saxons.” (Rutherford, N.J. and Bennett, E.H., 1918-1922, *English Speaking World*, Vol.2 (8): 14)

Concluding reflections

Beyond the myths:

- Texts and their production (including transcription)
- Spelling variation
- Words and lexemes
- N-grams and collocations
- Counting assumptions
- The nature of corpus linguistics
- Corpus linguistics and historical linguistics
- Corpus linguistics and lexicography