Shakespeare's Language: New Perspectives via Corpus-based Approaches

Jonathan Culpeper,

Lancaster University, UK

@ShakespeareLang

Lancaster

Lancaster

Lancaster

Lancaster

Lancaster University

Why approach Shakespeare from linguistics and why now?

Jack Cade:

"It will be proved to thy face that thou hast men about thee that usually talk of a noun and a verb, and such abominable words as no Christian ear can endure to hear." 2H6 IV.vii

- Linguistics has moved on (it's more applied than ever).
- Linguistics has new technologies by which it can be pursued (e.g. the computer).
- Linguistics is increasingly combined with other disciplines, as it is in this project (+ literary studies, computer science, history, etc.)
- Relatively little has been written on Shakespeare's language by linguists.

The Encyclopedia of Shakespeare's Language Project
(http://wp.lancs.ac.uk/shakespearelang)

Main people: Jonathan Culpeper, Dawn Archer, Alison Findlay, Andrew Hardie, Paul Rayson, Jane Demmen, Sean Murphy, Amelia Joulain-Jay, Mathew Gillings …

Funder: UK's Arts and Humanities Research Council (AHRC)

Timeframe: 2016-2019

Aim: 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.

A distinctive feature: It's comparative.

Preliminaries: What is meant by "Shakespeare's language"?

Shakespeare's language = surviving written texts that constitute, for the most part, the language that we call Shakespeare's works

Which "surviving written texts"?

- We need one stable body of texts – a corpus – at the heart of our project (partly for methodological reasons).

- Modern editions are a mish-mash of writings attributed to Shakespeare, and contain various modernizations.

- We do not have the resources to engage in author attribution work or (much) textual editing.

Preliminaries: What is meant by "Shakespeare's language"?

Solution?

- Use an original spelling diplomatic transcription of the largest near-contemporary body of work attributed to Shakespeare, i.e. the First Folio (1623), plus *Pericles* and *The Two Noble Kinsmen*.

- But don't ignore the Quartos. They constitute a secondary dataset.

Preliminaries: The problem with counting words

Our method relies on counting words in order to identify patterns

<u>What is a word?</u>

Two definitions:

(1) 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)

Well come, in deed, by cause, etc.

Hour glass, hour-glass, hourglass

Preliminaries: The problem with counting words

(2) 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)

Problem solved?

Preliminaries: The problem with counting words

Preliminaries: The problem with counting words

Historical issues:

Well come, in deed, by cause, etc.

Hour glass, hour-glass, hourglass

Preliminaries: The problem with counting words

Words and morphological variants:

- 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**

Preliminaries: The problem with counting words

Words and spelling variants:

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.

Preliminaries: The problem with counting words

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

Solution: Our general 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*

➤ Use a spelling that had EModE currency

➤ Prioritize the most frequent spelling in Shakespeare

A glance at the First Folio and spelling variation in English (Baron, Rayson & Archer 2009)
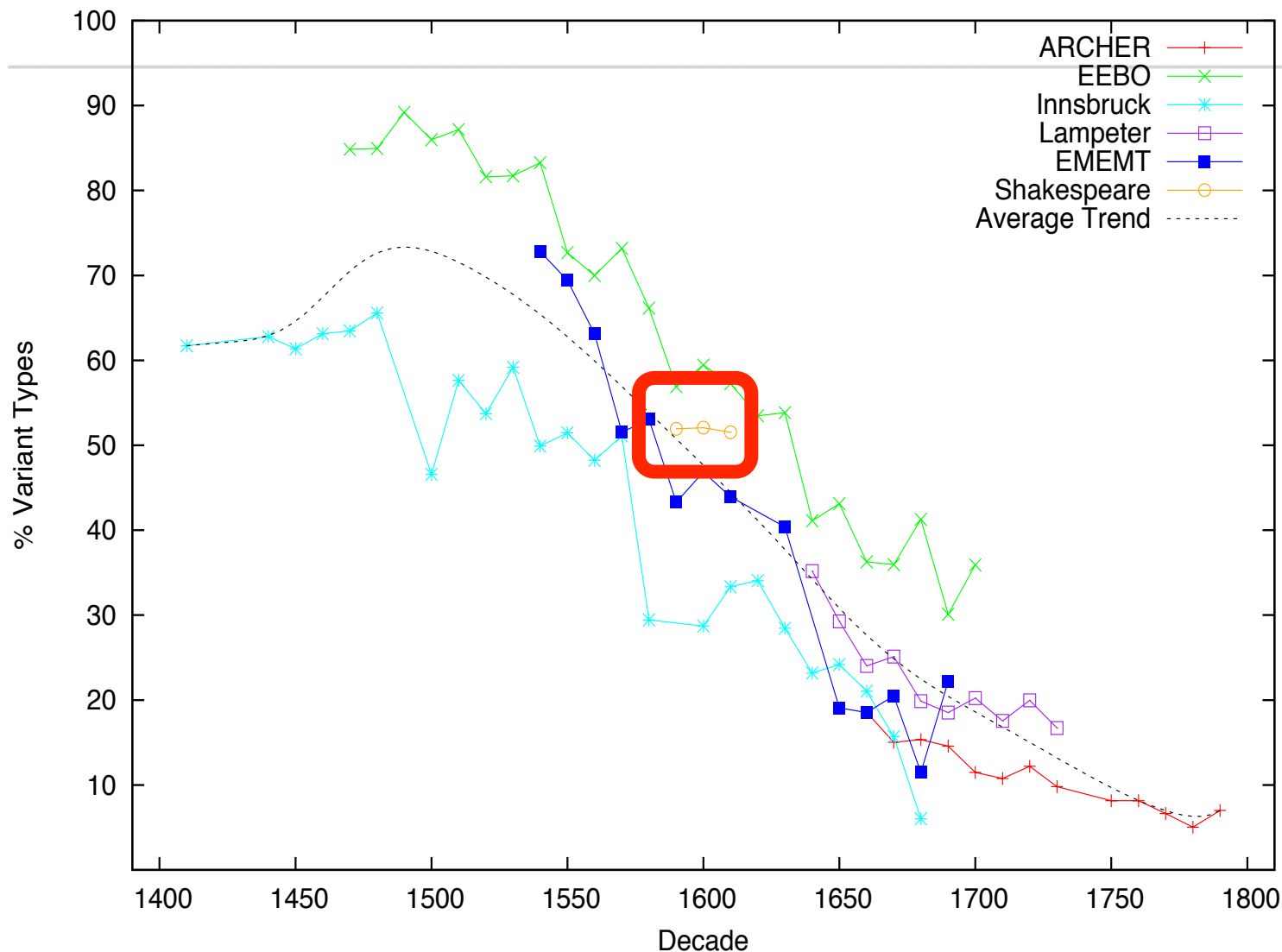
Shakespeare's neologisms?

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"

Shakespeare's neologisms?

Work on neologisms:

- 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.
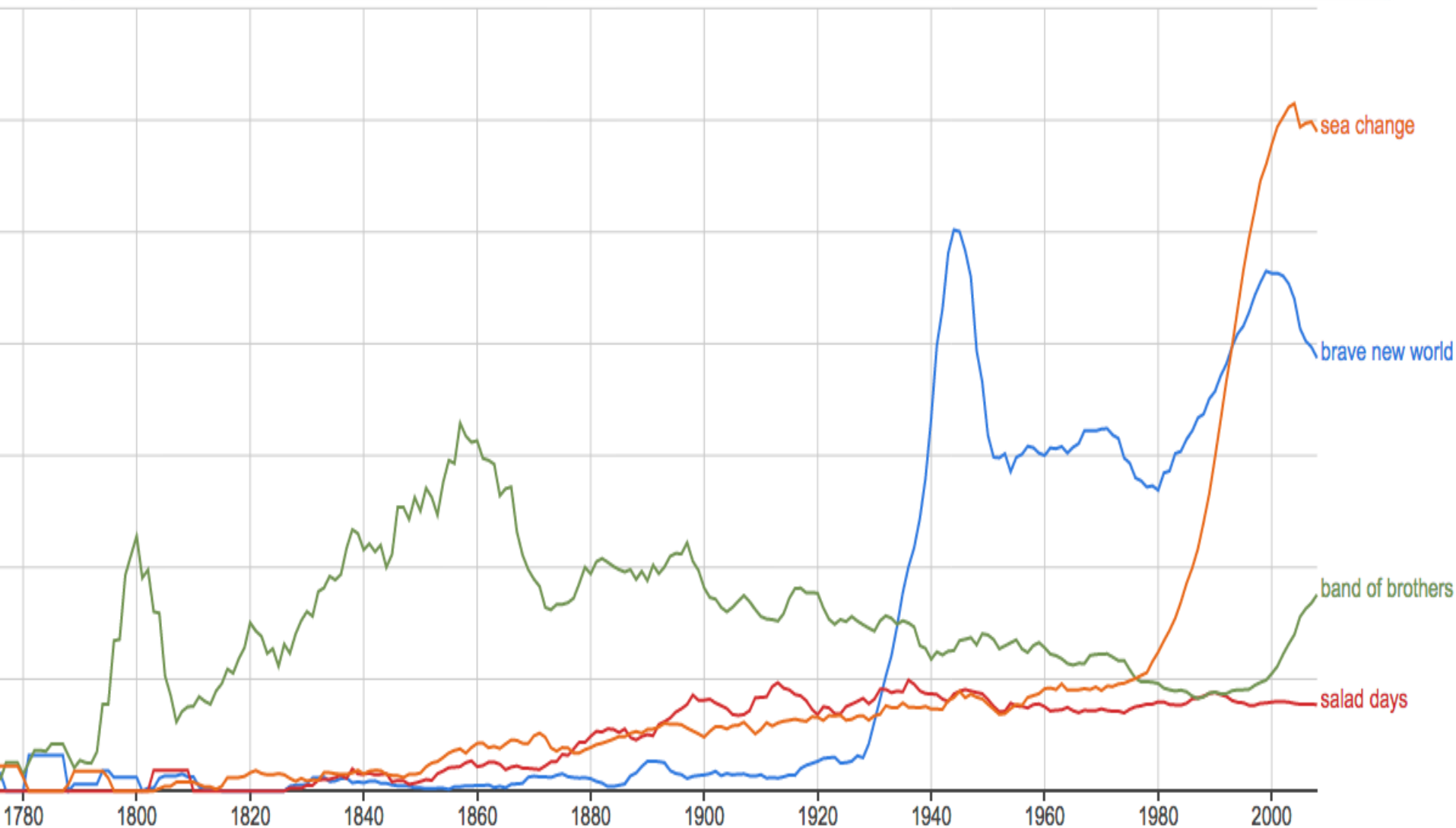
Shakespeare's neologisms?

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 *acerbic,* that appear in mixed Latin-English texts before Shakespeare?

- Is it actually just a nonce word rather than neologism? Cf. *dropsied* vs. *domineering*

Shakespeare's neologisms

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)

Shakespeare's language transcends time and space?

To the memory of my beloued,
The AVTHOR
MR. WILLIAM SHAKESPEARE:
AND
what he hath left vs.
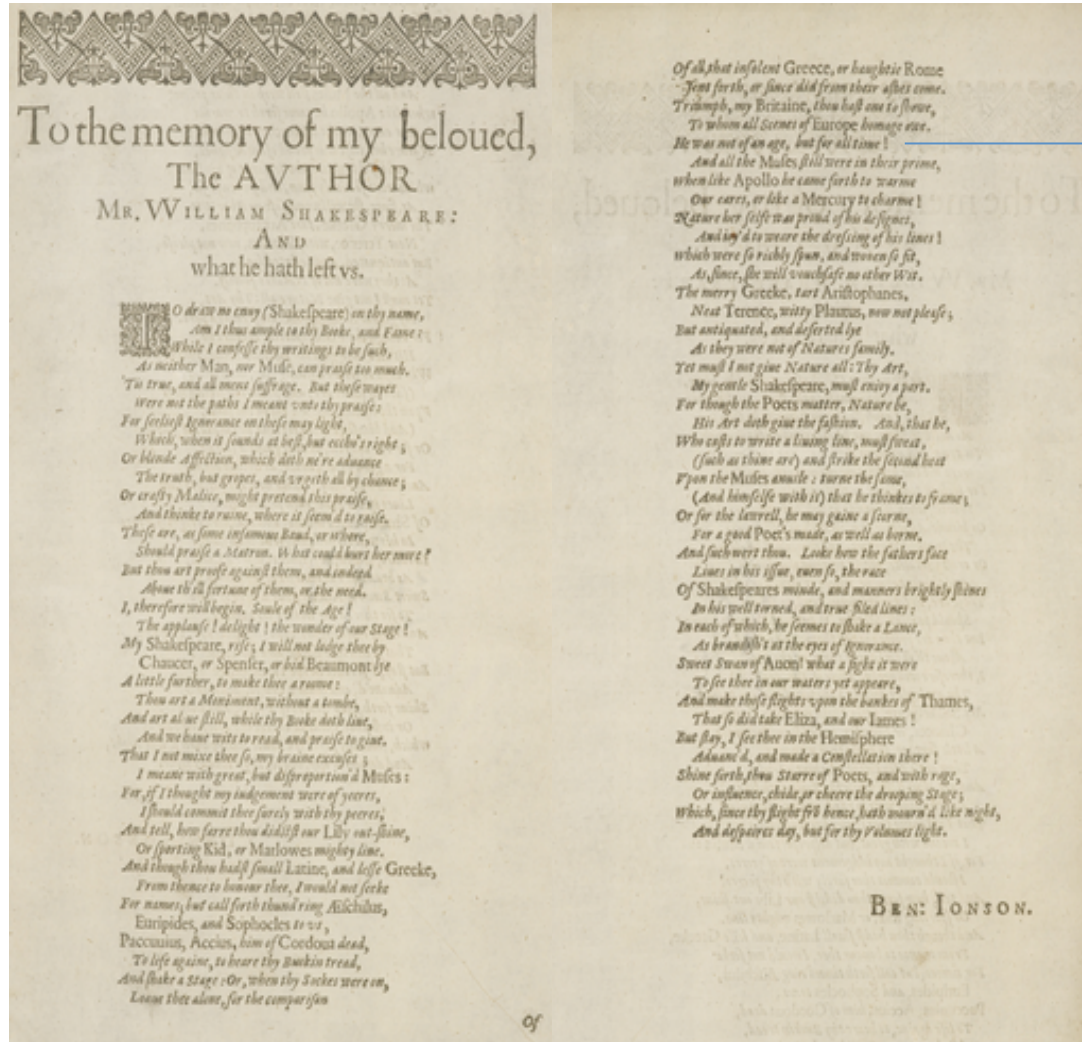
He was not of an age, but for all time !

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.

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

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.

Shakespeare's language transcends time and space - *good*

Encyclopaedia of Shakespeare's Language

Definition preview: definition 104 for entry good_ADJ

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.

good will As now.

good morrow Good morning.

good night As now.

| T | C | H |

Contemporary attitudes towards the Scottish, Irish and Welsh
(with Alison Findlay)

Henry V : Fluellen, Macmorris and Jamy

Globe Education

- How did contemporaries of Shakespeare view the Scots, Welsh and Irish?
- How did people write about them at the time?

Lexical items: *Scottish*, *Irish* and *Welsh*

Focus: words that frequently co-occur with Scottish, Irish or Welsh, i.e. collocates

- **Data**: *Early English Books Online* – approx. 1.2 billion words (?)

- **Period**: 1580-1599 – 82,180,304 words (around *Henry V*)

- **Tool**: CQPweb (Andrew Hardie)

- **Settings**: 5 words left and right of target item (within sentence boundaries). Mutual information. Minimum frequency is 10.

Lexical items: Frequencies

How often do they occur?

Scots/Scottish: **5,407** instances in 282 texts

Irish: **1,160** instances in 144 texts

Welsh: **802** instances in 115 texts

Lexical items: *Scots*/*Scottish*

Thematic groups (amongst the top 50 collocates)

Associated groups (*confederates***,** *ioine***)**: *Picts, Irish, Britains, Frenchmen, Danes,* etc.

Scottish kings/queens and nobles: *Malcolm, Ferguse, Kenneth, Donald, Bruce, Galled,* etc.

Hostile: *iuadeth, discomfited, borders, invaded, chased*

Scottish histories: *chronicles, writers, yere*

Political power: *nation, nobility, ambassadors,* etc.

Religion: *communion, supper*, etc.

Lexical item: *Irish*

Strongest collocate: *Irish rug*

"Show me a fair scarlet, a vvelch frise, a good Irish rug" (Eliot, 1595)

Lexical item: *Irish*

Thematic groups (top 50 collocates)

Negative connotations:

Uncivilised: *savage, wild*

Hostile: *wars*, *enemies*, *against*

Ungovernable: *rebels*

[**Insignificant????**: *mere** (Holinshed)]

Associated groups: *Scottish, Scots, English [rebels]*

Suppressed: *slue, hundred*

Political power: *nation, lords*

Language: *tongue, language, speak, called*

[MacMorris: "What ish my nation? Ish a villain and a bastard and a knave and a rascal?"]

Lexical item: *Welsh*

Thematic groups (top 50 collocates)

Associated groups (collaborators): *Englishmen*

Suppressed: *number, against, king, Danes*

Welsh language: *English, tongue, y, speaks, call, called, word*

William Allen, *A Conference About the Next Succession* (1595)

"… in the **welsh** also towards the English, who are a different people and of different language, and yet are they governed peaceably by the English, & the English again do account them for their country men …"

Character profiles: *Romeo* and *Juliet*

Lily James and Richard Madden.

(Photo: Johan Perrson)

- What language characterizes Romeo and what language, Juliet?
- What are their linguistic styles, their style markers, their keywords?

Character profiles: *Romeo* and *Juliet*

Rank-ordered keywords for Romeo and Juliet (raw frequencies in brackets)

Romeo	Juliet
beauty (10), love (46), blessed (5), eyes (14), more (26), mine (14), dear (13), rich (7), me (73), yonder (5), farewell (11), sick (6), lips (9), stars (5), fair (15), hand (11), thine (7), banished (9), goose (5), that (84)	if (31), be (59), or (25), I (138), sweet (16), my (92), news (9), thou (71), night (27), would (20), yet (18), that (82), nurse (20), name (11), words (5), Tybalt's (6), send (7), husband (7), swear (5), where (16), again (10)

Character profiles: *Romeo* and *Juliet*

Romeo:

- She hath, and in that sparing makes huge waste; For **beauty**, starv'd with her severity, Cuts **beauty** off from all posterity. She is too **fair**, too wise, wisely too **fair**, To merit bliss by making **me** despair: She hath forsworn to **love**, and in that vow Do I live dead that live to tell it now. (I.i)

- If I profane with our unworthiest **hand** This holy shrine, the gentle sin is this; Our **lips**, two blushing pilgrims, ready stand To smooth that rough touch with a tender kiss. (I.v)

Character profiles: *Romeo* and *Juliet*

Juliet**:**

- **If** he **be** married, / Our grave is like to **be our** wedding-bed (I.v.)

- **If** they do see thee, they will murder thee (II.ii.)

- But **if** thou meanest not well (II.ii.)

- Is thy news good, **or** bad? answer to that; Say either, and I'll stay the circumstance: Let me be satisfied, is 't good **or** bad? (II.ii)

- Tis almost morning; I would have thee gone; And **yet** no further than a wanton's bird […] (II.ii.)

Love and semantic analysis in Shakespeare's works
(with Dawn Archer and Paul Rayson)

- Three "love" tragedies *(Othello, Anthony & Cleopatra, Romeo & Juliet)*
- Three "love" comedies *(A Midsummer Night's Dream, Two Gentlemen of Verona, As You Like It)*

- Words tagged for semantic field
- Statistical comparison of semantic tags

The semantic category most over-used in 'love comedies'

PARTICIPANTS ...

Twosomes – 'couples', 'lovers'
Males – 'lover', 'suitor'
Females – 'virgin', 'wanton'

A DIFFERENT GENDER BIAS: PARTICULARLY IN RELATION TO AGENCY …?

"intimate/sexual relationship"

Male agents – 'kiss', 'kisses'
Female patients – 'seduced', 'deflowered'
Both 'fall in love'

... PROCESSES

The quality of 'love' in the Comedies vs. Tragedies

Lancaster University

|	Comedies	Tragedies
Sexual relationship **Intimate relationship** **Liking**	adoration (1), adore (2), adored (1), affection (7), affections (1), amorous (1), applaud (2), applause (1), apple of his eye (1), beloved (8), chastity (3), cherish (1), cherished (2), copulation (1), couples (4), dear ... fa... (7)... (1)... (2)... (3)... lo... lovers (26), loves (26), loving (10), paramour (2), precious (5), prized (1), relish (2), revelling (1), revels (5), savours (3), seduced (1), sensual (1), suitor (2), take to (1), that way (1), virgin (4), wanton (5)	adore (1), adores (1), affection (8), affections (7), affinity (1), amorous (6), applauding (1), applause (1), beloved (5), beloving (1), bewhored (1), carnal (1), chastity (3), cherish (1), cherished (1), cherishing (1), courts (1), cuckold (6)... d... d... e... fa... (1)... k... (1)... lo... (29), loving (12), lust (10), lusts (1), paramour (1), precious (6), prized (1), rate (3), rated (1), relish (1), revel (3), revels (4), sluttish (1), suitor (2), suitors (2), take to (1), that way (1), wantons (1), wooer (1)

apple of his eye, copulation, couples, gone for, seduced, sensual, virgin

affinity, bewhored, carnal, cuckold, darling(s), devotion, lust(s), sluttish, wooer

The language of emotion in Shakespeare's plays (with Alison Findlay, Beth Cortese, Mike Thelwall and Dawn Archer)

"Sentiment analysis"

- Emotion words, whether they are positive or negative (valence), and their strength.
- *SentiStrength* (Thelwall; http://sentistrength.wlv.ac.uk/)
- Lexicon adjusted for EModE and Shakespeare in particular.
- Checked against a human rater.

Overall negative sentiment across Shakespeare's plays (average negative sentiment subtract average positive sentiment)
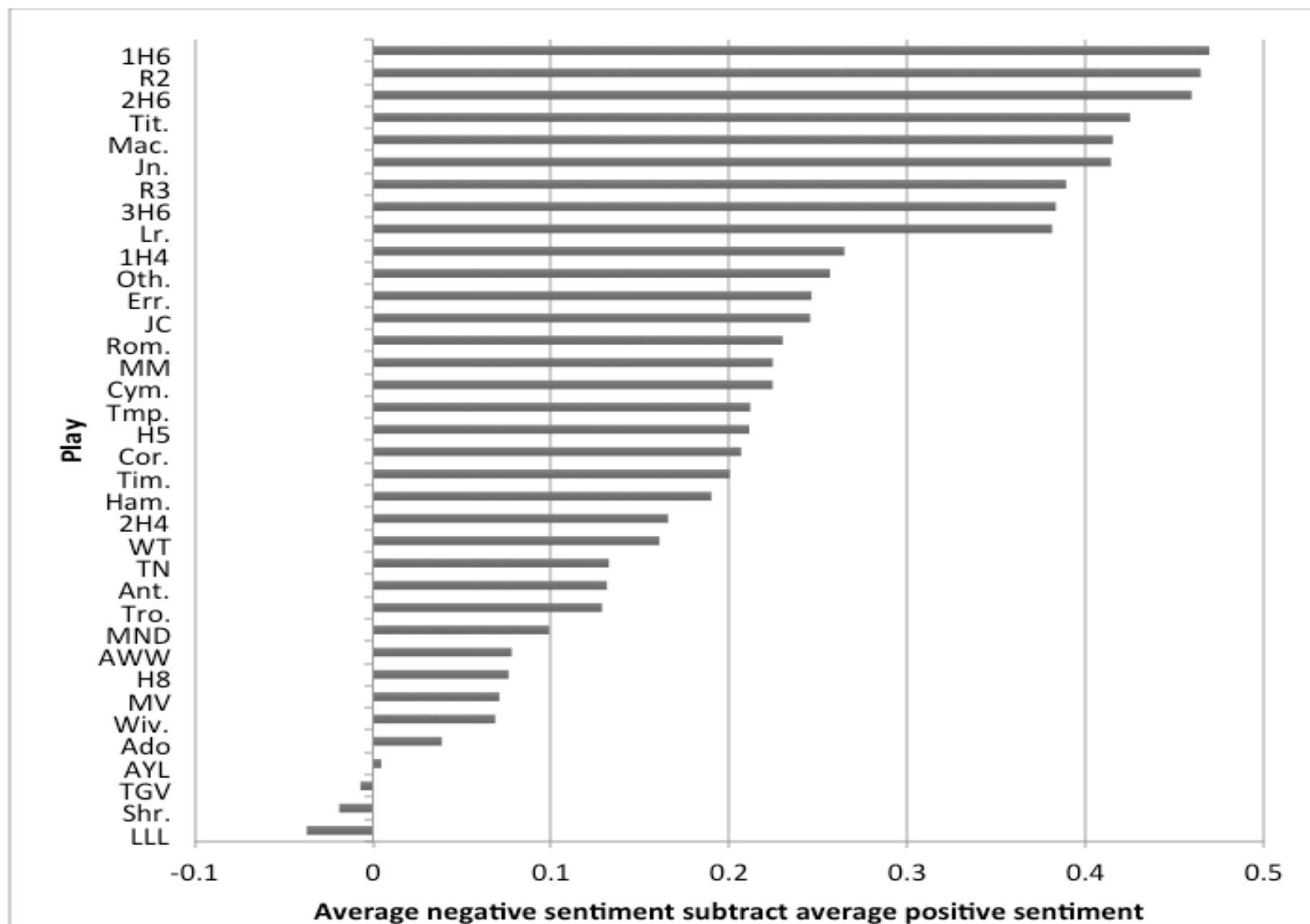
Concluding thought

No man is an *Iland*, intire of it selfe; every man is a peece of the *Continent*, a part of the *maine*; if a *Clod* bee washed away by the *Sea, Europe* is the lesse […]

(John Donne 1624)

References

Baron, A., Rayson, P. and Archer, D. (2009). Word frequency and key word statistics in historical corpus linguistics. In Ahrens, R. and Antor, H. (eds.) *Anglistik: International Journal of English Studies*, 20 (1), pp. 41-67.

Culpeper, J. (2009). Keyness: words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics*, 14(1), 29-59.

Culpeper, J. and Demmen, J. (2015). Keywords. In Biber, D. and Reppen, R. (eds.) *The Cambridge Handbook of English Corpus Linguistics.* Cambridge: Cambridge University Press, pp. 90-105.

Culpeper, J. and Findlay, A. (2016). Contemporary understandings of Welsh, Scottish and Irish identities: Celtic characters in Shakespeare. Paper given at the 19th International Conference on English Historical Linguistics, University of Duisburg-Essen, Germany, 22-26 August 2016. http://wp.lancs.ac.uk/shakespearelang/files/2016/12/Celtic-characters-in-Shakespeare.pdf.

Culpeper, J. Findlay, A., Cortese, B. and Thelwall, M. (under review). 'What means this passionate discourse?' (Henry VI, Part II, 1.1.101): Measuring Emotional Temperatures in Shakespeare's Drama. *English Text Construction*.

Crystal, D. and Crystal, B. (2004). *Shakespeare's Words: A Glossary and Language Companion.* Penguin.

EEBO-TCP http://www.textcreationpartnership.org/tcp-eebo/

Hardie, A. (2012). CQPweb — combining power, flexibility and usability in a corpus analysis tool. International Journal of Corpus Linguistics 17(3): 380-409. http://www.lancaster.ac.uk/staff/hardiea/cqpweb-paper.pdf.

Leech, G., Rayson, P. and Wilson, A. (2001). *Word Frequencies in Written and Spoken English Based on the British National Corpus.* Harrow: Pearson Education.

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

VAriant Detector (VARD 2) software http://ucrel.lancs.ac.uk/vard/about/.

Tools

http://ucrel.lancs.ac.uk/**vard**/about/

https://**cqpweb**.lancs.ac.uk/

http://ucrel.lancs.ac.uk/**wmatrix**/