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

Jonathan Culpeper,
Lancaster University, UK

@ShakespeareLang
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.
What will the project produce?

The key output: The encyclopedia

Two volumes:
(1) a kind of dictionary, and
(2) a compendium of word patterns relating to characters, character groups, plays, play-genres, themes, etc..

To be published by Bloomsbury in paper and electronically.
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.)
• Is it actually just a nonce word rather than neologism? Cf. *dropsied* vs. *domineering*
Shakespeare’s language transcends time and space?
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.”


• No Shakespearean dictionary has treated Shakespeare’s language as relative, i.e. put Shakespeare’s usage in the context of that of his contemporaries.
Lexical items: *Scottish, Irish and Welsh* (and contemporary attitudes towards the Scottish, Irish and Welsh)

*Henry V*: Fluellen, Macmorris and Jamy

- 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* (and contemporary attitudes towards the Scottish, Irish and Welsh)
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

- 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)*

<table>
<thead>
<tr>
<th>Romeo</th>
<th>Juliet</th>
</tr>
</thead>
<tbody>
<tr>
<td>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)</td>
<td>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)</td>
</tr>
</tbody>
</table>
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.)
Concluding thought

Corpus analyses are not automatic or free from humans. A human must decide on the data to be used and the kind of analysis to be performed. Moreover, they must interpret the results.

(More information about the project can be found here: http://wp.lancs.ac.uk/shakespearelang)