Shakespeare’s Language: New perspectives on old language

Jonathan Culpeper,
Lancaster University, UK

@ShakespeareLang
Overview

• The project and the corpus-based approach
• Methodological challenges and solutions
• Shakespeare and numbers: A glance at neologisms and survivals
• Words (Vol.1): Shakespeare dictionaries and the treatment of the ‘horrid’ and ‘good’
• Word patterns and ‘themes’ (Vol.2): Character and play profiles
• Some other areas: Multi-word units and the language of emotion
• Conclusions
What the project aims to do ....

• Produce the first systematic account of Shakespeare’s language using methods derived from corpus linguistics – an approach that uses computers in large-scale language analysis.
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. The volume establishes both what is unique about Shakespeare's language and what Shakespeare's language meant to his 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.
Methodological issues

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.

Solution: *Variant Detector* (VARD) program, primarily devised by generations of scholars at Lancaster, but most recently given a significant boost by Alistair Baron.
A glimpse at Shakespeare: The First Folio and spelling variation in English (Baron et al’s 2009)
Methodological issues (contd.)

The comparative corpus

**Problem:** Size matters

- Any pattern is a matter of frequency.
- Linguistics is centrally focussed on patterns in language.
- Historical linguistics work is often hampered by low frequencies, because the historical record is not complete.
- Corpus-based methods and concepts (e.g. collocates) are centrally driven by frequencies and statistical operations.

**Solution:** Various new corpora and electronic texts, but especially *Early English Books Online* (EEBO-TCP) – 1520-1679, and at least 723 million words.
Shakespeare and numbers: Neologisms and survivals

Myths about Shakespeare and the English language:

What can we ‘learn’ from the internet?

• He coined more words than other writers, around 1700 words ...
• or is that 3,000 ...
• or did he invent half the words in the English language ...

N.B. The issues are twofold: neologisms and survivals
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 EEBO-TCP

Preliminary findings:

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

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 English texts before Shakespeare?

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

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

• Examples 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)
Shakespearean ‘dictionaries’ and present-day corpus-based dictionaries

Some typical differences in approach:

• Words for inclusion: ‘hard’ words vs. all words in the corpus
• Word-meanings: etymological meanings and etymological organization vs. meanings based on usage in context and organised according to frequency

Note:
No Shakespearean dictionary has treated Shakespeare’s language as relative, i.e. put Shakespeare’s usage in the context of that of his contemporaries.
Case study: ‘horrid’ today

Examples from the BNC (random):

one day could take over from Morgan. A horrid man.
really glad to be on there to dispense with all those horrid people.
the horrid male instructor drills you as if you're in the Green Berets)
Smith being beaten by spotty, horrid little Nails tickled Nutty's imagination.
the tramp! He's horrid!" Shirley's cheeks had turned pale at the thought
will be giving the editor of New Scientist the full horrid details without delay.
recent research suggests that lead isn't as horrid in its effects as the

Top-40 rank-ordered most frequently occurring nouns within 5 words to the right of ‘horrid’ in the BNC:
things, man, thing, creature, stuff, truth, people, feeling, word, beast, phrase,
teeth, girls, flat, day, child, place, state, time, blighters, imprecations, defilement,
deodorants, cruelties, malady, apparitions, weasels, double-glazing, panoply,
sunflowers, bungling, separateness, puns, premonition, shrieks, jingle, hairstyle,
imaginations, blasphemy
Philological approach:

Oxford English English Dictionary

**horrid** (ˈhɔrɪd), a. (adv.) Also 7 horred, horride.
[ad. L. horrid-us bristling, rough, shaggy; rude, savage, unpolished; terrible, frightful, f. horrere: see horre v. Cf. It. *orrido.*]

A. adj.
1. Bristling, shaggy, rough. (Chiefly poetic.)

1590 Spenser *F.Q.* i. vii. 31 His haughtie Helmet, horrid all with gold.

1621 Burton *Anat. Mel.* i. ii. iii. xiv. (1651) 125 A rugged attire, hirsute head, horrid beard.

---

**Case study: ‘horrid’ (contd.)**
Case study: ‘horrid’ (contd.)

2. Causing horror or aversion; revolting to sight, hearing, or contemplation; terrible, dreadful, frightful; abominable, detestable.

In earlier use nearly synonymous with horrible; in modern use somewhat less strong, and tending to pass into the weakened colloquial sense (3).

1601 Shakes. Twel. N. iii. iv. 220, I wil meditate the while vpon some horrid message for a Challenge.

3. colloq. in weakened sense. Offensive, disagreeable, detested; very bad or objectionable.

Noted in N.E.D. as especially frequent as a feminine term of strong aversion.

1666 J. Davies Hist. Caribby Isls 281 Making horrid complaints that treated them ill.
Case study: ‘horrid’ in Shakespearean dictionaries

Shakespearean dictionaries (in brief):

- *Foster* (1908): “(1) Awful, hideous, horrible. (2) Terrific. (3) Horrified, affrighted”.
- *Onions* (1911): No entry.

Nasty = Foster (1)
Frightening = all other definitions
Case study: ‘horrid’ in Shakespeare

Appeare in formes more horrid) yet my Duty, As doth a Rocke Vp Sword, and know thou a more horrid hent. When he is drunke And cleaue the generall eare with horrid speech: Make mad the guilty heard and seene, Recounts most horrid sights seene by the Watch. shall breake his winde With feare and horrid flight. 1.Sen. Noble, To. I wil meditate the while vpon some horrid message for a Challenge. armes. Macd. Not in the Legions Of horrid Hell, can come a Diuell deformitie seemes not in the Fiend So horrid as in woman. all the sparkes of Nature To quit this horrid acte. Reg. Out treacherous Such sheets of Fire, such bursts of horrid Thunder, Such groanes of Curriors of the Ayre, Shall blow the horrid deed in euery eye, on is Of thy deere Husband. Then that horrid Act Of the diuorce, to themselves Beene deathes most horrid Agents, humaine grace I yeeld to that suggestion, Whose horrid Image doth vnfixe my Heire
Case study: ‘horrid’ in Shakespeare

The beginnings of a contextualised dictionary entry:

**Headword:** HORRID. Adj.

**Sense:** Something that is *horrid* causes fear; typically, it refers to supernatural or unnatural acts, sights and sounds. E.G. ‘Whose horrid Image doth vnfixe my Heire’ (Mac.)

**Contexts:** Horrid has a much closer association with Shakespeare's tragedies than either histories or comedies, and is used slightly more frequently by male characters than female. Shakespeare used it considerably more than his contemporary playwrights did. Generally, it is most characteristic of Early Modern plays and, perhaps surprisingly, scholarly literature.

**Distribution:** All = 16 (1.8); T = 10 (3.9), C = 2 (0.6), H = 4 (1.5); M = 14 (1.9), F = 2 (1.4).

**Comparisons:** Pla = 187 (0.17), Fic = 0, Tr = 0, Ha = 0, Sc = 1 (0.14).

- Frequency limitations
Case study (2): ‘good’

Crystal & Crystal (2004): “(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
Case study (2): ‘good’

Frequency: 2711

Pretend some alteration in **good** will? What's heere? I haue vpon My selfe, and my **good** Cousin Buckingham, Will to your Mother, she is low voic'd. Cleo. That's not so **good**: he cannot like her long. Goodmorrow (**good** Lieutenant) I am sorrie For your displeasure: Father Frier. Duk. And you **good** Brother Father; what offence an enuous emulator of euerie mans **good** parts, a secret & villanous she shall be there. Ro. And stay thou **good** Nurse behind the Abbey wall, Mar. Patience deere Neece, **good** Titus drie thine eyes. Ti. Ah Marcus, Anthonio; that I had a title **good** enough to keepe his name company! the singlenesse. Mer. Come betweene vs **good** Benuolio, my wits faints. Enter Count Rossillion. Par. **Good**, very good, it is so then: good, very nightes meete him. 1.Knight. **Good** morrow to the good Simonides. a troublous world. 1. No, no, by Gods **good** grace, his Son shall reigne. signe of Feare. 1 Cit. The Gods bee **good** to vs: Come Masters let's home,
### Case study (2): ‘good’ (contd.)

<table>
<thead>
<tr>
<th>No.</th>
<th>Word</th>
<th>Total no. in whole corpus</th>
<th>Expected collocate frequency</th>
<th>Observed collocate frequency</th>
<th>In no. of texts</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>morrow</td>
<td>332</td>
<td>8.45</td>
<td>113</td>
<td>28</td>
<td>35.793</td>
</tr>
<tr>
<td>2</td>
<td>Lord</td>
<td>2,591</td>
<td>65.948</td>
<td>287</td>
<td>33</td>
<td>27.159</td>
</tr>
<tr>
<td>3</td>
<td>.</td>
<td>66,858</td>
<td>1701.727</td>
<td>2,703</td>
<td>36</td>
<td>24.26</td>
</tr>
<tr>
<td>4</td>
<td>night</td>
<td>785</td>
<td>19.981</td>
<td>88</td>
<td>24</td>
<td>15.105</td>
</tr>
<tr>
<td>5</td>
<td>good</td>
<td>2,756</td>
<td>70.148</td>
<td>184</td>
<td>30</td>
<td>13.534</td>
</tr>
<tr>
<td>6</td>
<td>cheere</td>
<td>56</td>
<td>1.425</td>
<td>17</td>
<td>14</td>
<td>12.627</td>
</tr>
<tr>
<td>7</td>
<td>my</td>
<td>12,103</td>
<td>308.056</td>
<td>526</td>
<td>36</td>
<td>12.389</td>
</tr>
<tr>
<td>8</td>
<td>Sooth</td>
<td>66</td>
<td>1.68</td>
<td>18</td>
<td>13</td>
<td>12.206</td>
</tr>
<tr>
<td>9</td>
<td>bad</td>
<td>157</td>
<td>3.996</td>
<td>28</td>
<td>18</td>
<td>11.758</td>
</tr>
<tr>
<td>10</td>
<td>newes</td>
<td>278</td>
<td>7.076</td>
<td>38</td>
<td>16</td>
<td>11.437</td>
</tr>
</tbody>
</table>
Word patterns and ‘themes’ (Vol.2): Character profiles

- What language characterizes Romeo and what language Juliet?
- What are their linguistic styles, their idiolects?

Lily James and Richard Madden.

(Photo: Johan Perrson)
Word patterns and ‘themes’ (Vol.2): Character profiles

• Studies of linguistic style normally involve the analyst spotting linguistic features that might be diagnostic of that style, and then offering analysis/discussion of those.

• Limited to human spotting abilities. As John F. Burrows (1987: 1) put it, ‘It is a truth not generally acknowledged that, in most discussions of works of English fiction, we proceed as if a third, two-fifths, a half of our material were not really there.’

• The solution is to identify statistically key words, i.e. get a computer to compare the frequencies of words in one body of data against another and identify which are unusual relative to the other. And then analyze/discuss those.

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>

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)

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.)
Word patterns and ‘themes’ (Vol.2): Play profiles (Cf. Archer, Culpeper & Rayson 2009)

- Thematic profile: Semantic categorization (‘lexical fields’)
- Each word assigned to a semantic category

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>general and abstract terms</td>
<td>B</td>
<td>the body and the individual</td>
</tr>
<tr>
<td>F</td>
<td>food and farming</td>
<td>G</td>
<td>government and public</td>
</tr>
<tr>
<td>K</td>
<td>entertainment, sports and games</td>
<td>L</td>
<td>life and living things</td>
</tr>
<tr>
<td>O</td>
<td>substances, materials, objects and equipment</td>
<td>P</td>
<td>education</td>
</tr>
<tr>
<td>T</td>
<td>Time</td>
<td>W</td>
<td>world and environment</td>
</tr>
<tr>
<td>Z</td>
<td>names and grammar</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lancaster University

[Diagram showing a table with semantic categories]
Love in Shakespeare’s works

In Archer et al (2009), we explored the love theme within:

‘Love’ tragedies:
• Othello, Anthony and Cleopatra and Romeo and Juliet

‘Love’ comedies:
• A Midsummer Night’s Dream, The Two Gentlemen of Verona and As You Like It
LOVE OVERUSED IN “LOVE” COMEDIES
(relative to the three “love” tragedies)

PARTICIPANTS ...

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

“A DIFFERENT GENDER BIAS:
PARTICULARLY IN RELATION TO AGENCY ...?”

“intimate/sexual relationship” material divides into ...

Male agents – kiss, kissing, kissed, kisses
Female patients – seduced, deflowered
Both fall in love, falling in love, fell in love

... PROCESSES
Other “overused” categories in 3 comedies

<table>
<thead>
<tr>
<th>Living creatures</th>
<th>Negative – bears, serpent, snail, monster, adder, snake, claws, chameleon, worm, monkey, ape, weasel, toad, rat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neutral – cattle, horse, goats, creature, capon, nest</td>
</tr>
<tr>
<td></td>
<td>Positive – deer, dove, nightingale</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(Not) sensible</th>
<th>V. interesting connection with love ...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“We that are true lovers run into strange capers; but as all is mortal in nature, so is all nature in love mortal in folly” ... AYLI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensory: taste</th>
<th>Sweet/-er/-est – representative of “sweet talk” used in courtship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bitter/-ness, sour/-est, taste/-s – often relate to the troubles of love (e.g. unrequited love)</td>
</tr>
</tbody>
</table>
“Overused” categories in the three tragedies

<table>
<thead>
<tr>
<th>War</th>
<th>Lack of life</th>
<th>Religion / supernatural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not calm / angry</td>
<td>Lack of power</td>
<td>Movement</td>
</tr>
</tbody>
</table>
Captures violent conflicts characterising tragedies ... esp. in *R&J* and *Othello*. **Rage, fury** - fairly evenly distributed in our 3 “love” plays but ... **revenge** > all but one from *Othello* ... **abused, abuse** > Othello reflecting on Desdemona’s treatment of him ... **whipped** > all but one from *A&C* ... **slew** > all but one relating to deaths of Tybalt or Mercutio in *R&J*

Hierarchies differ in terms of freq. and type:
more ‘domestic’ in comedies – more ‘military’ here.
**Knave, sirrah, minion**, etc. – tend to be used abusively

Military activity in *A&C* and *Othello* ... but also metaphorical usages:

*Othello*  
Are not you a strumpet?

*Desdemona*  
No, as I am a Christian; If to preserve this **vessel** for my lord From any other foul unlawful touch Be not to be a strumpet, I am non.
## Multi-word units

<table>
<thead>
<tr>
<th>Shakespeare</th>
<th>EModE Plays</th>
<th>Present-day Plays</th>
</tr>
</thead>
<tbody>
<tr>
<td>I pray you</td>
<td>it is a</td>
<td>I don’t know what do you</td>
</tr>
<tr>
<td>I will not</td>
<td>what do you and I will</td>
<td>I don’t want do you think</td>
</tr>
<tr>
<td>I know not</td>
<td>it is not</td>
<td>I don’t want do you</td>
</tr>
<tr>
<td>I am a</td>
<td>I have a</td>
<td>I don’t think do you</td>
</tr>
<tr>
<td>I am not</td>
<td>I will not</td>
<td>I don’t think to do with</td>
</tr>
<tr>
<td>my good lord</td>
<td>in the world</td>
<td>do you know</td>
</tr>
<tr>
<td>there is no</td>
<td>I tell you</td>
<td>going to be</td>
</tr>
<tr>
<td>I would not</td>
<td>I know not</td>
<td>don’t want to</td>
</tr>
<tr>
<td>it is a and I will</td>
<td>I warrant you</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Three-word lexical bundles in</td>
</tr>
<tr>
<td></td>
<td></td>
<td>order of frequency (coloured items</td>
</tr>
<tr>
<td></td>
<td></td>
<td>appear in another column)</td>
</tr>
</tbody>
</table>

Data in 2\textsuperscript{nd} and 3\textsuperscript{rd} columns draw from Culpeper and Kytö (2010)
Theatrical context: Stage and staging today
The adjacency pair in present-day drama

Willy Russell, Educating Rita, 1981, p.8

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
Theatrical context: EModE stage and staging

Purpose-built outdoor theatres:
The Theatre (1576),
The Curtain (1577),
The Rose (1587),
The Swan (1595),
The Globe (1599), and
The Fortune (1600).
Multi-word units

• A trend in the Early Modern data is for the lexical bundle to begin with a first person pronoun
• Especially notable trend for Shakespeare, where it combines with verbs relating to states, desires and knowledge. *I pray you* is most distinctive.
• Perhaps reflects a tendency for characters to present themselves (and others) relatively directly (including via soliloquies and asides).
The language of emotion in Shakespeare’s plays

+ Alison Findlay, Beth Cortese and Mike Thelwall

- “Sentiment analysis” and commercial goals
- What is it analysing? Emotion words and whether they are positive or negative (valence). Sometimes their strength too.
- SentiStrength (Thelwall; http://sentistrength.wlv.ac.uk/)
- Lexicon adjusted for EModE and Shakespeare in particular.
Overall negative sentiment across Shakespeare’s plays (average negative sentiment subtract average positive sentiment)

Play:
1H6, R2, 2H6, Tit., Mac., Jn., R3, 3H6, Lr., 1H4, Oth., Err., JC, Rom., MM, Cym., Tmp., H5, Cor., Tim., Ham., 2H4, WT, TN, Ant., Tro., MND, AWW, H8, MV, Wiv., Ado, AYL, TGV, Shr., LLL

Average negative sentiment subtract average positive sentiment
Conclusions

A corpus approach to Shakespeare’s language means:

• All ‘words’ treated equally (e.g. not just ‘hard’ words).
• Meanings based on usage in context (e.g. not etymology, not narrowly-defined semantic meaning).
• The context includes linguistic aspects (e.g. collocations) and non-linguistic aspects (e.g. registers, social properties of the speaker/character).

A corpus/computational approach to literary texts means:

• Makes a kind of “distant reading” possible through the identification of linguistic patterns.
Problems and limitations

• The methodology is not suitable for items below a certain frequency.
• Grammatical and semantic annotation need further development (manual correction), if they are to be deployed.
• It is never automatic – the human is needed to (1) devise/train the software, (2) select the data and prepare it; and (3) interpret the results.