Shakespeare’s language: Insights afforded by new developments in corpus linguistics

Jonathan Culpeper,
Lancaster University, UK

@ShakespeareLang
Overview

- The project and the corpus-based approach
- Methodological challenges and solutions
- A glance at current Shakespeare dictionaries

- Case study (1): ‘horrid’
- Case study (2): ‘good’
- Case study (3): ‘ah’

- Case study (4): Multi-word units
- Case study (5): Character profiles
- Case study (6): Play profiles

- 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 generation 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.
Current Shakespearean ‘dictionaries’


Various labels:

- ‘dictionary’
- ‘glossary’
- ‘lexicon’
- ‘word-book’

- ‘concordances’ (Spevack 1968-70; Howard-Hill 1969-72)
Current Shakespearean ‘dictionaries’ (contd.)

With various contents:

**Linguistic** [e.g. Foster (1908), Onions ([1911] 1986), Schmidt ([1902] 1971), Crystal and Crystal (2002), etc.]
- Word
- Part-of-speech
- Brief definition
- Illustrative quotation(s)

**Frequency information** [e.g. Spevack 1968-70; Howard-Hill 1969-72]
- Index of all words (plus textual location)
- Frequency of word-form (absolute + relative)

**Non-linguistic** [e.g. Boyce (1996), Wells (1998)]
- Play summaries (largely plot)
- Character descriptions
- Cultural information
- Biographical information
A corpus-based approach to dictionaries

Integration of linguistic description, frequency information and non-linguistic information.

Starting point research questions:
• How often does X occur?
• How often do the particular meanings of X occur?
• What kind of words does X tend to co-occur with?
• What kinds of register does X co-occur with?
• What kinds of speaker/addressee does X co-occur with?
• Etc., etc.

• Contextualised definitions (cf. Collins COBUILD)
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. Shakespeare’s usage with that of his contemporaries.
Case study (1): ‘horrid’

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 (1): ‘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).


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 (1): ‘horrid’ (contd.)

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
Case study (1): ‘horrid’- A glimpse at Shakespeare

Shakespearean dictionaries (in brief):

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

Nasty = Foster (1)

Horrifying = all other definitions
Case study (1): ‘horrid’- A glimpse at Shakespeare (contd.)

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 cleave 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 (1): ‘horrid’- A glimpse at Shakespeare (contd.)

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, and is associated with roughness, sharpness, and cruelty. 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).

**Contemp. views**: ‘prickie, horrid, sharpe’, ‘rough, horrid, cruell’, ‘gastly, or horrid in lookes’ (Florio, 1611).

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

Dictionaries (in brief):

• *Foster* (1908): “(1) Not bad, worthy of praise; (2) Fit, adapted; (3) Trustworthy, genuine; (4) Kind, benevolent; (5) Proper, right; (6) Substantial, safe, solvent, able to fulfil engagements, (7) Real, serious; (8) Favourable, propitious, (9) Abundant, rich, (10) Skilful, clever, (11) Adequate”.
  + phrases and compounds.

• *Onions* (1911): “(1) Conventional epithet to titles of high rank, (2) comely, (3) Financially sound; (hence) wealthy, substantial. Notes quasi-adverbial usage, e.g. ‘good easy man’.”
  + phrases and compounds
Case study (2): ‘good’ (adj.)

- *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
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 ev ery 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 betwenee 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>WORD</th>
<th>TOTAL</th>
<th>LEFT</th>
<th>RIGHT</th>
<th>L5</th>
<th>L4</th>
<th>L3</th>
<th>L2</th>
<th>L1</th>
<th>*</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>600</td>
<td>253</td>
<td>347</td>
<td>63</td>
<td>62</td>
<td>62</td>
<td>56</td>
<td>10</td>
<td>0</td>
<td>7</td>
<td>110</td>
<td>89</td>
<td>74</td>
<td>67</td>
</tr>
<tr>
<td>AND</td>
<td>651</td>
<td>352</td>
<td>299</td>
<td>81</td>
<td>63</td>
<td>66</td>
<td>93</td>
<td>49</td>
<td>0</td>
<td>15</td>
<td>131</td>
<td>49</td>
<td>42</td>
<td>62</td>
</tr>
<tr>
<td>TO</td>
<td>569</td>
<td>276</td>
<td>293</td>
<td>65</td>
<td>76</td>
<td>78</td>
<td>49</td>
<td>8</td>
<td>0</td>
<td>45</td>
<td>94</td>
<td>43</td>
<td>57</td>
<td>54</td>
</tr>
<tr>
<td>THE</td>
<td>577</td>
<td>306</td>
<td>271</td>
<td>80</td>
<td>65</td>
<td>51</td>
<td>27</td>
<td>83</td>
<td>0</td>
<td>4</td>
<td>34</td>
<td>67</td>
<td>90</td>
<td>76</td>
</tr>
<tr>
<td>MY</td>
<td>612</td>
<td>351</td>
<td>261</td>
<td>43</td>
<td>43</td>
<td>28</td>
<td>26</td>
<td>211</td>
<td>0</td>
<td>107</td>
<td>33</td>
<td>38</td>
<td>40</td>
<td>43</td>
</tr>
<tr>
<td>LORD</td>
<td>313</td>
<td>57</td>
<td>256</td>
<td>15</td>
<td>18</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>0</td>
<td>137</td>
<td>82</td>
<td>18</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>YOU</td>
<td>543</td>
<td>289</td>
<td>254</td>
<td>55</td>
<td>62</td>
<td>65</td>
<td>56</td>
<td>51</td>
<td>0</td>
<td>9</td>
<td>24</td>
<td>83</td>
<td>61</td>
<td>77</td>
</tr>
<tr>
<td>A</td>
<td>565</td>
<td>387</td>
<td>178</td>
<td>49</td>
<td>39</td>
<td>38</td>
<td>40</td>
<td>221</td>
<td>0</td>
<td>23</td>
<td>23</td>
<td>39</td>
<td>40</td>
<td>53</td>
</tr>
<tr>
<td>IS</td>
<td>295</td>
<td>159</td>
<td>136</td>
<td>30</td>
<td>18</td>
<td>31</td>
<td>51</td>
<td>29</td>
<td>0</td>
<td>2</td>
<td>18</td>
<td>33</td>
<td>43</td>
<td>40</td>
</tr>
<tr>
<td>OF</td>
<td>357</td>
<td>225</td>
<td>132</td>
<td>40</td>
<td>33</td>
<td>35</td>
<td>60</td>
<td>57</td>
<td>0</td>
<td>6</td>
<td>43</td>
<td>16</td>
<td>35</td>
<td>32</td>
</tr>
<tr>
<td>FOR</td>
<td>252</td>
<td>122</td>
<td>130</td>
<td>22</td>
<td>21</td>
<td>20</td>
<td>48</td>
<td>11</td>
<td>0</td>
<td>17</td>
<td>37</td>
<td>29</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>SIR</td>
<td>170</td>
<td>46</td>
<td>124</td>
<td>11</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>3</td>
<td>0</td>
<td>61</td>
<td>27</td>
<td>9</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>YOUR</td>
<td>308</td>
<td>189</td>
<td>119</td>
<td>30</td>
<td>35</td>
<td>26</td>
<td>29</td>
<td>69</td>
<td>0</td>
<td>3</td>
<td>16</td>
<td>38</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>MORROW</td>
<td>137</td>
<td>25</td>
<td>112</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
## 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>
Case study (3): ‘ah’

Dictionaries (in brief):

- *Foster* (1908): No entry.
- *Onions* (1911): No entry.

Is it a word?
Case study (3): ‘ah’ (contd.)

 depends on your definition of a word.

Orthographic word = ‘a string of uninterrupted non-punctuation characters with white space or punctuation at each end’ (Leech et al. 2001: 13-14)
Does it have meaning?

Halliday (1978, 1985) functional components:
- Ideational (information; logical)
- Textual (informational structure)
- Interpersonal (pragmatic)

- Interactional (discoursal)
Case study (3): ‘ah’ (contd.)

- **Speaker attitude/state: sorrow, emotional distress**

  Des. To whom my Lord? With whom? How am I false?

  Oth. **Ah** Desdemona, away, away, away.

  Des. Alas the heauy day: why do you weepe? Am I the motiue of these teares my Lord?  (OTH)

- **Speaker attitude/state: pity**

  Glou. Canst thou blame him? His Daughters seeke his death: **Ah**, that good Kent, He said it would be thus: poore banish'd man: Thou sayest the King growes mad, Ile tell thee Friend I am almost mad my selfe. (KIL)
• Speaker attitude/state: Surprise, realisation

Enter Adriana and Luciana.

Adr. **Ah** Luciana, did he tempt thee so? (COM)

• Discourse marker: preface to the correction / rejection of the previous speaker’s proposition(s), emotions or actions

*Men.* These three World-sharers, these Competitors Are in thy vessell. Let me cut the Cable, And when we are put off, fall to their throates: All there is thine.

*Pom.* **Ah**, this thou shouldst haue done, And not haue spoke on't. In me 'tis villanie, In thee, 't had bin good seruice: [...] (ANT)
Case study (3): ‘ah’ (contd.)

- Discourse marker: reinforces elicitation

*Leon.* All thy tediousnesse on me, *ah*?

*Const.Dog.* Yea, and 'twere a thousand times more than 'tis, for I heare as good exclamation on your Worship as of any man in the Citie, and though I bee but a poore man, I am glad to heare it. (MAN)
Case study (3): ‘ah’

Distribution across play genres (Total = 108; distribution per 100,000; only plays in which they occur are included)

<table>
<thead>
<tr>
<th>Tragedy</th>
<th>History</th>
<th>Comedy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titus (13), Romeo (13), Antony (10), Timon (2), Hamlet (2), Othello (1), Lear (1), Coriolanus (1)</td>
<td>Richard III (17), Richard II (5), Henry IV (5), Henry VI (3), John (3), Henry VIII (3)</td>
<td>Loves Labours (5), Troilus (4), Twelfth (3), Much ado (3), Comedy (3), Missummer (2), Merry Wives (2), Cymbeline (2), Two Noblemen (1), Two Gentlemen (1), Taming (1), As You (1), Alls Well (1)</td>
</tr>
</tbody>
</table>

(43 / 205,970) = 20.9

(36 / 151,780) = 23.7

(29 / 294,492) = 9.8
Case study (4): 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</td>
</tr>
<tr>
<td>I will not</td>
<td>what do you</td>
<td>what do you</td>
</tr>
<tr>
<td>I know not</td>
<td>and I will</td>
<td>I don’t want</td>
</tr>
<tr>
<td>I am a</td>
<td>it is not</td>
<td>do you think</td>
</tr>
<tr>
<td>I am not</td>
<td>I have a</td>
<td>do you want</td>
</tr>
<tr>
<td>my good lord</td>
<td>I will not</td>
<td>I don’t think</td>
</tr>
<tr>
<td>there is no</td>
<td>in the world</td>
<td>to do with</td>
</tr>
<tr>
<td>I would not</td>
<td>I tell you</td>
<td>do you know going to be</td>
</tr>
<tr>
<td>it is a</td>
<td>I know not</td>
<td>don’t want to</td>
</tr>
<tr>
<td>and I will</td>
<td>I warrant you</td>
<td></td>
</tr>
</tbody>
</table>

Three-word lexical bundles in order of frequency (coloured items appear in another column)

Data in 2\textsuperscript{nd} and 3\textsuperscript{rd} columns draw from Culpeper and Kytö (2010)
In the Present-day Play-texts one can discern many fragments of questions; there is only one question fragment in the Early Modern data: WHAT DO YOU.

Present-day questions orientate to beliefs, thoughts, wants and feelings. Note the high frequency of private verbs (know, want, think).

Present-day Play-texts deploy the adjacency pair question as a central mechanism in not only creating the interaction but indirectly revealing information for the audience.
Theatrical context: The stage and staging
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. [...]
Theatrical context: The 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).
Case study (4): 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).
Case study (5): 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)
Case study (5): Character profiles

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>


Case study (5): Character profiles (contd.)

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)
Case study (5): Character profiles (contd.)

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.)
Case study (6): Play profiles
(Cf. Archer, Culpeper & Rayson 2009)

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

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>general and abstract terms</td>
<td>the body and the individual</td>
<td>arts and crafts</td>
<td>emotion</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>food and farming</td>
<td>government and public</td>
<td>architecture, housing and the home</td>
<td>money and commerce in industry</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>L</td>
<td>M</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>entertainment, sports and games</td>
<td>life and living things</td>
<td>movement, location, travel and transport</td>
<td>numbers and measurement</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>P</td>
<td>Q</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>substances, materials, objects and equipment</td>
<td>education</td>
<td>language and communication</td>
<td>social actions, states and processes</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>W</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
</tr>
<tr>
<td>Time</td>
<td>world and environment</td>
<td>psychological actions, states and processes</td>
<td>science and technology</td>
<td>names and grammar</td>
</tr>
</tbody>
</table>
‘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*
### Case study (7): Play profiles (contd.)

**Most frequent in the comedies**

<table>
<thead>
<tr>
<th>Semantic tag/field</th>
<th>Log</th>
<th>Likelihood</th>
<th>Frequency</th>
<th>%</th>
<th>Frequency</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>S3.2 = intimate/sexual relationship</td>
<td>379</td>
<td>0.64</td>
<td>292</td>
<td>0.36</td>
<td>55.5</td>
<td></td>
</tr>
<tr>
<td>E2+ = liking</td>
<td>325</td>
<td>0.55</td>
<td>321</td>
<td>0.39</td>
<td>17.36</td>
<td></td>
</tr>
<tr>
<td>L2 = living creatures</td>
<td>343</td>
<td>0.58</td>
<td>279</td>
<td>0.34</td>
<td>42.3</td>
<td></td>
</tr>
<tr>
<td>L3 = plants</td>
<td>149</td>
<td>0.25</td>
<td>94</td>
<td>0.12</td>
<td>35.99</td>
<td></td>
</tr>
<tr>
<td>S1.2.6- = (not) sensible</td>
<td>72</td>
<td>0.12</td>
<td>32</td>
<td>0.04</td>
<td>31.02</td>
<td></td>
</tr>
<tr>
<td>X3.1 = sensory: taste</td>
<td>120</td>
<td>0.20</td>
<td>91</td>
<td>0.11</td>
<td>18.41</td>
<td></td>
</tr>
<tr>
<td>T3- = age: young</td>
<td>153</td>
<td>0.26</td>
<td>128</td>
<td>0.16</td>
<td>17.12</td>
<td></td>
</tr>
</tbody>
</table>

As *You Like It* and *Midsummer Night’s Dream* are set in woods. Also, metaphor “sex is agriculture” – crop, reap, etc.

**Participants**

- **Positive**: Sweet, sweetest, sweeter – representative of “sweet talk” used in courtship. Also, metaphor “food is love”:
  - bitter, bitterness, sourest, sour & taste, tastes
  - often relate to the troubles of love (e.g.
  - Unrequited love)
  - Males – *lover, suitor*
  - Females – *virgin, wanton*

**Processes …**

- Transitive with male or female agents - *loves*
- Transitive with male agents – *kiss, kissing, kissed, kisses*
- Transitive with female patients – *seduced, deflowered*
- Intransitive with male or female agents – *fall in love, falling in love, fell in love*
## Case study (7): Play profiles (contd.)
### Most frequent in the tragedies

<table>
<thead>
<tr>
<th>Semantic tag (field)</th>
<th>Comedies</th>
<th>Tragedies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freq %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood Freq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3 = warfare, defence, &amp; the army</td>
<td>425</td>
<td>213.51</td>
</tr>
<tr>
<td>L1- = (lack of) life/living things</td>
<td>490</td>
<td>77.16</td>
</tr>
<tr>
<td>Z2 = geographical names</td>
<td>399</td>
<td>49.56</td>
</tr>
<tr>
<td>E3- = (not) calm/violent/angry</td>
<td>343</td>
<td>33.67</td>
</tr>
<tr>
<td>M4 = movement (by sea/through water)</td>
<td>92</td>
<td>28.51</td>
</tr>
<tr>
<td>S9 = religion and the supernatural</td>
<td>644</td>
<td>21.92</td>
</tr>
<tr>
<td>S7.1- = (lack of) power/organising</td>
<td>193</td>
<td>21.55</td>
</tr>
</tbody>
</table>

Almost all of the most frequent items are from *Anthony & Cleopatra* – e.g. soldier, war(s), army, *baXle*, armour, navy, etc. A few items occur in *Othello*: ‘general’ is a vocative form for Othello; ‘lieutenant’ is a vocative form for Cassio.


Some misnomers – e.g. Moor Indicative of tragedies involving a number of different geographical locations: particularly true of *Anthony* and Cleopatra...

Captures the violent conflicts that characterise the tragedies...

‘quarrel’ – some relate to Abraham Quarrel; however, most refer to squabbles in *Romeo and Juliet* and *Othello*.

‘abused’, ‘abuse’ – Othello reflecting on Desdemona’s treatment of him; some examples also from *Anthony and Cleopatra*.

“rage”, “fury” - fairly evenly distributed in 3 plays.

Tends to tally with military activity in *Anthony & Cleopatra* and *Othello*.

Some metaphorical usages:

- *Othello*:
  - *Are not you a strumpet?*
  - *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.*

- *Desdemona*:
- "Suggests that the tragedies revolve around hierarchical power structures rather more than the comedies."

- *Othello*:
- "’knave’, ‘sirrah’, ‘minion’, etc. – tend to be used abusively."

- *Tybalt*:
- "Heaven", "soul", "devil" – majority from *Othello* or *Romeo & Juliet*.
Conclusions

A corpus approach to Shakespeare’s language means:
• All ‘words’ treated equally (e.g. not just ‘hard’ words).
• Meanings are not restricted to semantic or ideational meaning.
• Meanings based on usage in context (e.g. not etymology).
• The context includes linguistic aspects (e.g. collocations) and non-linguistic aspects (e.g. registers, social properties of the speaker/character).

A corpus approach to literary texts means:
• Makes “distant reading” possible.
• Avoids a narrow focus on what just the parts that the critic notices.

John F. Burrows (1987: 1): ‘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.’
Conclusions (contd.)

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.

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