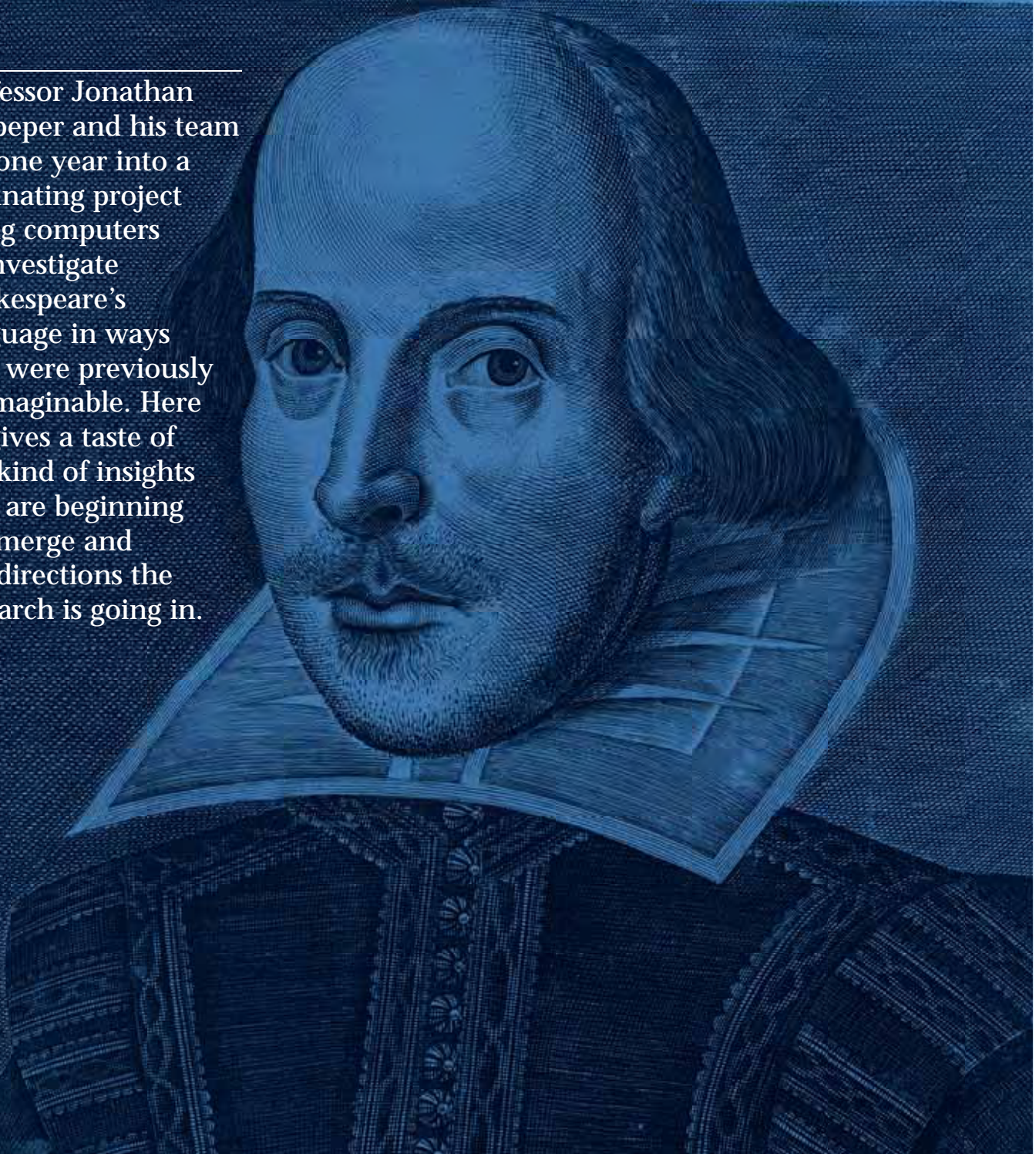# SHAKESPEARE'S
LANGUAGE

Professor Jonathan Culpeper and his team are one year into a fascinating project using computers to investigate Shakespeare's language in ways that were previously unimaginable. Here he gives a taste of the kind of insights that are beginning to emerge and the directions the research is going in.

SALAD DAYS
BAND OF BROTHERS
SEA CHANGE

Public domain

Shakespeare's Language?

A trip to any library reveals shelves groaning with the weight of literary critical works on Shakespeare, but only a few volumes on his language. This is odd, if one considers the high esteem in which Shakespeare's language is held. It is precisely to make up for this deficit that I am leading a project investigating Shakespeare's language – the **Encyclopaedia of Shakespeare's Language** project, funded by the Arts and Humanities Research Council (AHRC). One of the things that is rather different about this project is that it deploys computers to examine how Shakespeare's language is used. Not only that, it uses computers to compare his language use with that of his contemporaries. To get such a project off the starting blocks, however, one needs to confront two particular problems.

Can we really talk about 'Shakespeare's language'? These days, authors lock themselves in their studies, garden sheds or whatever, churn out their piece, send it to their publisher, and then, after some editing, it appears in print. The connection between author and work, and all the words in it, is relatively clear. In Shakespeare's case, the most complete body of his work, the *First Folio*, was published in 1623, seven years after his death in 1616. It was cobbled together from various, sometimes less than official, copies of his works. I write 'his works', but even that is not secure. In early modern England, collaborative works were common. The notion of plagiarism was not the same as it is today. Copying another writer could be perceived as complimentary: you showed that their work was worth copying. So, ultimately, when we talk about 'Shakespeare's language', we mean a set of texts which Shakespeare played a leading role in creating, and not that Shakespeare wrote every single word.

Variable Spellings

The second problem is the fog of variability. When Shakespeare was writing, English had only just begun the processes of standardisation. The particular problem concerns spelling. Aside from the fact that whatever spelling was in Shakespeare's original manuscript is not necessarily what got printed, and that typesetters varied spellings even over the course of one play (thus creating further distance between us and Shakespeare), there is the problem of retrieval. Let's imagine that you want to use a computer to look up all the cases when Shakespeare uses the word **sweet**. Searching on that spelling would not be entirely successful. It would miss out cases spelt **sweete**. Or, let's say, you want to look up the word **hourglass**, not just in Shakespeare, but also in the works of his contemporaries, for the purposes of comparison. You would minimally need to check hour **glass**, **houre glass**, **hour glasse**, **houre-glasse**, **hour-glass**, **houre-glass**, **hour-glasse**, **houreglasse**, **hourglass**, **houreglass** and **hourglasse**. In fact, compound nouns like this are an acutely problematic area, varying hugely even from one Shakespeare edition to another.

Computational techniques for analysing language have made some progress with both of these problems. Regarding the first, there are increasingly sophisticated techniques for identifying recurring patterns of words that might characterise Shakespeare's writings. The recent decision of *The New Oxford Shakespeare* to add co-authors, such as Christopher Marlowe, to some of 'Shakespeare's plays' was based on analyses of such patterns (especially, the use of combinations of grammatical words, such as 'and with', that are barely noticeable when you read them but could reveal the unconscious preferences of a writer). Regarding the second, at Lancaster University we have developed a piece of software, VARD (http://ucrel.lancs.ac.uk/vard/about/), which enables us to regularise spelling. It does not delete the original spellings, which could provide useful insights (into pronunciation, for example), but adds some code to associate them with a regular form (thus, **sweete** would be associated with **sweet**).

Shakespeare's Neologisms

Claims about the number of words that Shakespeare supposedly coined are part of the Shakespeare myths factory. On the internet, we find that Shakespeare coined 'more words than other writers, around

1,700 words'. Another authoritative statement gives the number as 3,000. Yet further trawling reveals that he 'invented half the words in the English language'. Some variation in numbers results from what one considers a different word to be. Higher counts are probably caused by taking word-forms such as **walk**, **walks** and **walked** as different words, whereas lower counts would treat them as one (one 'lexeme'). Presumably, that statement about the English language is a dramatic exaggeration. But how much of an exaggeration? The *Oxford English Dictionary* contains almost 200,000 different word entries (lexemes). Shakespeare's total output amounts to around 20,000 different words (lexemes). So it is a massive exaggeration!

I have been exploring, with Sheryl Banas, which words Shakespeare might have invented. There are 1,502 words recorded in the *Oxford English Dictionary* as first cited in Shakespeare. We have been checking whether Shakespeare was the first to produce these, using a fully searchable version of Early English Books Online, a collection of historical printed work amounting to about 1.2 billion words of English covering the early modern period. We have not finished this work. However, it seems that less than a quarter of those 1,502 words can reasonably be attributed to Shakespeare.

'Reasonably be attributed' might sound like a very vague way of putting it, but calculated guesses are the best we can do. How do we know that Shakespeare coined it as opposed to recorded it? We can but guess. Shakespeare seems to have been one of the first people, if not the first, to have used **down staires**. This is probably a regular development of the phrase 'down the stairs', and likely used by many in speech at that time. It just so happened that the earliest written record of it is in Shakespeare's works. In contrast, the word **incarna[r]dine**, which existed as an adjective but not a verb, may well have been coined by Shakespeare because he needed a verb for a creative purpose – to make the lines

The multitudinous seas incarnadine,
Making the green one red

(here, **incarnadine** has the sense of making red or pink). How do we know that it is not simply a nonce word – a one-off word made-up for a specific purpose and dropped thereafter? **Dropsied** (to have the flu-like condition of dropsy) seems to be a

nonce word: it first appears in Shakespeare, then disappears. In contrast, **domineering** first appears in Shakespeare, and then spreads to other writers. For words to be neologisms, they need to be recognised and used by communities, as has been the case with **domineering**.

Shakespeare's Survivals

The claim that Shakespeare 'invented half the words in the English language' seems to suppose that Shakespeare's words hung around. Did they? I have not yet systematically examined this issue, but I have looked at some examples of phrases first recorded in Shakespeare's works and their more recent life. Figure 1 above is generated by Google Book Ngram Viewer (http://books.google.com/ngrams). It displays the frequencies of **brave new world** (*The Tempest*), **sea change** (*The Tempest*), **band of brothers** (*Henry V*) and **salad days** (*Antony and Cleopatra*) in books published between 1780 and 2008 (constituting over 350 billion words of English). **Salad days** continually bumps along the bottom of the graph; it is used infrequently. **Band of brothers** has some interesting spikes. The spike around 1800 may be due to the use of the phrase by Admiral Nelson at the Battle of Copenhagen in 1801, and its reporting in the press. Similarly, the spike just before 1860 is close to the Crimean War, 1853-1856. The most recent spike also reflects war – the fictional war of the fictional TV series *Band of Brothers* (2001). **Sea change** is somewhat like **salad days**, until the 1980s when it is picked up and used as a term to report dramatic shifts in financial markets. **Brave new world** is used less frequently than **sea change**, until it is catapulted into the limelight by Aldous Huxley's 1931 novel of that name.

One might wonder whether people are actually aware that these phrases probably emanate from Shakespeare. A Nottingham University PhD thesis written

by Sarah Grandage established that most people are not.

Looking Ahead

I have been scratching at the surface of a myriad of interesting things about Shakespeare's language. I focused on some popular issues, mostly involving numbers. However, I tried to impress that numbers alone are meaningless. We need to know what we are counting, and examine what the counts mean in context. During the Encyclopaedia of Shakespeare's Language project we will examine each and every one of Shakespeare's words, and compare them with those of his contemporaries. We will also examine linguistic patterns in his works – the patterns that constitute and distinguish characters or particular kinds of characters (e.g. male versus female), a comedy from a tragedy or history, or particular themes (e.g. love, death). We are only one year into the project, and so the bulk of our findings will appear over the next two to three years. Check out our website, for updates, resources and so on: **http://wp.lancs.ac.uk/shakespearelang/**

Jonathan Culpeper (Lancaster University, UK; Principal investigator of the AHRC-funded project, Encyclopaedia of Shakespeare's Language.

@ **emag web archive**

- Professor Crystal's series on 'Shakespeare's False Friends'
- Professor Crystal: Shakespeare: New Words, *emagazine* 5, September 1999

emagClips

- Professor Elena Semino on stylistics