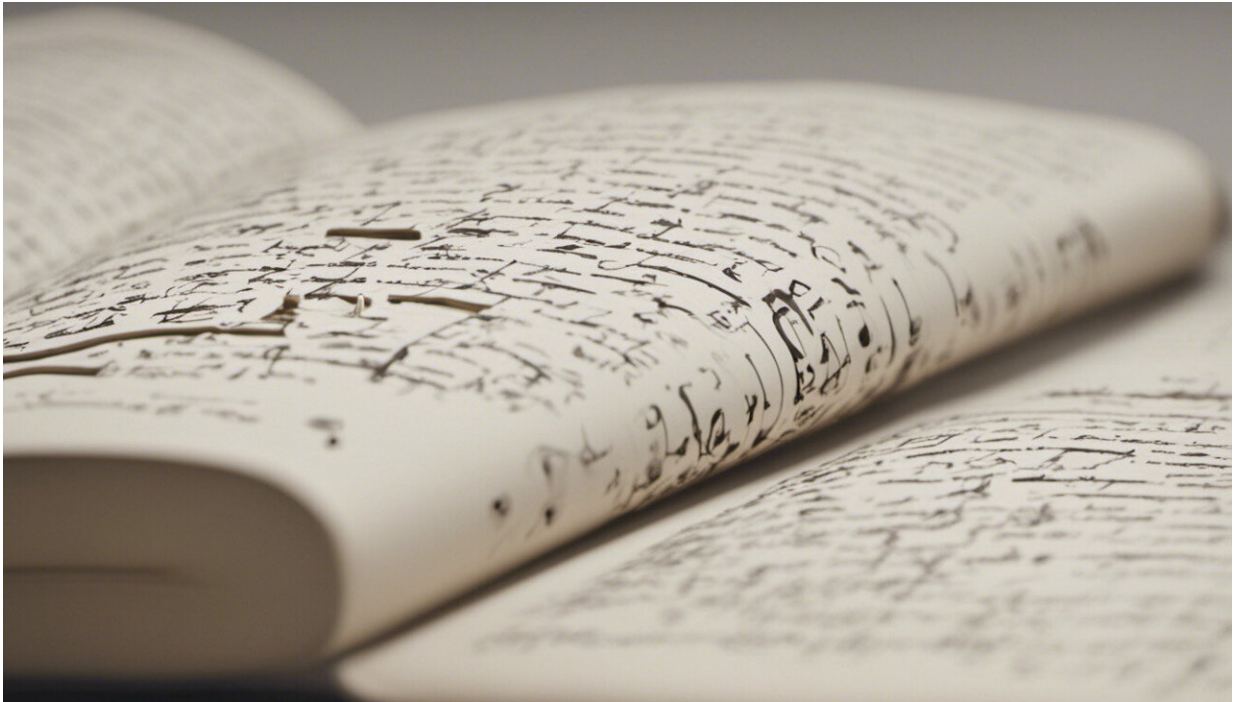# How linguists are unlocking the meanings of Shakespeare's words using numbers

September 7 2023, by Jonathan Culpeper

Credit: AI-generated image (disclaimer)

Today it would seem odd to describe a flower with the word "bastard"—why apply a term of personal abuse to a flower? But in Shakespeare's time, "bastard" was a technical term describing certain plants.

Similarly, associating the word "bad" with success and talking of a "bad success" would be decidedly odd today. But it was not unusual then, when success meant outcome, which could be good or bad.

Corpus linguistics is a branch of linguistics which uses computers to explore the use of words in huge collections of language. It can spot nuances that might be overlooked by linguists working manually, or large patterns that a lifetime of studying may not reveal. And numbers, counts of words and keeping track of where the words are occurring, are key.

In my experience at conferences and the like, talk of numbers is not unanimously well received in the world of literary studies. Numbers are sometimes perceived as being reductive, or inappropriate when discussing creative works, or only accessible to specialists.

Yet, describing any pattern involves numbers. In the first paragraph above, I used the words "normal," "odd" and "unusual" as soft ways of describing frequencies—the numbers of occurrences (think also of, for example, "unique," "rare," "common").

Even talking about "associations" involves numbers. Often associations evolve from an unusually high number of encounters among two or more things. And numbers help us to see things.

Changing meanings

Along with my team at Lancaster University, I have used computers to examine some 20,000 words gleaned from a million-word corpus (a collection of written texts) of Shakespeare's plays, resulting in a new kind of dictionary.

People have created Shakespeare dictionaries before, but this is the first to use the full armory of corpus techniques and the first to be

comparative. It not only looks at words inside Shakespeare's plays, but also compares them with a matching million-word corpus of contemporary early modern plays, along with huge corpus of 320 million words of various writings of the period.

Of course, words in early modern England had lives outside Shakespeare. "Bastard" was generally a term for a hybrid plant, occurring in technical texts on horticulture.

It could be, and very occasionally was, used for personal abuse, as in King Lear, where Edmund is referred to as a "bastard." But this is no general term of abuse, let alone banter, as you might see it used today. It is a pointed attack on him being of illegitimate parentage, genetically hybrid, suspect at his core.

The word "bad" is not now associated with the word "success," yet 400 years ago it was, as were other negative words, including "disastrous," "unfortunate," "ill," "unhappy" and "unlucky."

We can tap into a word's associations by examining its collocates, that is, words with which it tends to occur (rather like we make judgements about people partly on the basis of the company they keep). In this way we can see that the meaning of "success" was "outcome" and that outcome, given its collocates, could be good or bad.

Highly frequent words

We can use intuition to guess some word patterns. It's no surprise that in early modern English, the word "wicked" occurred very frequently in religious texts of the time. But less intuitively, so did "ourselves," a word associated with sermons and plays, both of which have in common a habit of making statements about people on earth.

Highly frequent words, so often excluded by historical dictionaries and reference works, are often short words that seem insignificant. They have a wood-for-trees problem.

Yet corpus techniques highlight the interesting patterns. It turns out that a frequent sense of the humble preposition "by" is religious: to reinforce the sincerity of a statement by invoking the divine (for example, "by God").

Numbers can also reveal what is happening inside Shakespeare's works. Frequent words such as "alas" or "ah" are revealed to be heavily used by Shakespeare's female characters, showing that they do the emotional work of lamentation in the plays, especially his histories.

Infrequent words

What of the infrequent? Words that occur only once in Shakespeare—so-called hapax legomena—are nuggets of interest. The single case of "bone-ache" in Troilus and Cressida evokes the horrifying torture that syphilis, which it applies to, would have been. In contrast, "ear-kissing" in King Lear is Shakespeare's rather more pleasant and creative metaphor for whispering (interestingly, other writers used it for the notion of flattering).

Another group of interesting infrequent words concerns words that seem to have their earliest occurrence in Shakespeare. Corpus techniques allowed us to navigate the troubled waters of spelling variation. Before spelling standardization, searching for the word "sweet," for instance, would miss cases spelt "sweete," "swete" or "svveet."

In this way, we can better establish whether a word written by a writer really is the earliest instance. Shakespearean firsts include the rather boring "branchless" (Antony and Cleopatra), a word probably not coined

by Shakespeare but merely first recorded in his text. But there is also the more creative "ear-piercing" (Othello) and the distinctly modern-sounding "self-harming" (The Comedy of Errors and Richard II).

Why are these advances in historical corpus linguistics happening now? Much of the technology to produce these findings was not in place until relatively recently.

Programs to deal with spelling variation (such as Vard) or to analyze vast collections of electronic texts in sophisticated ways (such as CQPweb), to say nothing of the vast quantities of computer-readable early modern language data (such as EEBO-TCP), have only been widely used in the last 10 or so years. We are therefore on the cusp of a significant increase in our understanding and appreciation of major writers such as Shakespeare.

This article is republished from The Conversation under a Creative Commons license. Read the original article.

Provided by The Conversation