CHAPTER 1

Introduction

Hugh Craig and Arthur F. Kinney

One of the earliest champions of language in Shakespeare's time was Thomas Wilson. In *The Arte of Rhetorique* (1553) he declares that

> Suche force hath the tongue, and such is the power of eloquence and reason, that most men are forced euen to yelde in that, whiche most standeth againste their will. And therfore the Poetes do feyne that Hercules being a man of greate wisdome, had all men lincked together by the eares in a chaine, to draw them and leade them euen as he lusted. For his witte was so greate, his tongue so eloquente, & his experience suche, that no one man was able to withstand his reason, but euerye one was rather driuen to do that whiche he woulde, and to wil that whiche he did, agreeing to his aduise both in word & worke, in all that euer they were able.
>
> Neither can I see that menne coulde haue bene broughte by anye other meanes to lyue together in felowshyppe of life, to mayntayne Cities, to deale trulye, and willyngelye to obeye one another, if menne at the firste hadde not by Art and eloquence perswaded that, which they ful oft found out by reason. (sigs. A3v–A4r)

At the time such ideas were not especially original – the works of Aristotle and Cicero in the Tudor grammar schools had made them commonplace – but Wilson's ambition and vision are nevertheless unusual. His manual, however – establishing the art of language in ways all the skilled dramatists of Shakespeare's day would observe – is grounded in a philosophy of mind that is sophisticated even by today's insights. Despite Wilson's high aims – they would reach well into the realm of poetry and drama – his counsel was up to the task. '[E]uery Orator', he teaches his readers,

> should earnestly laboure to file his tongue, that his woordes maie slide with ease, and that in his deliueraunce, he maie haue suche grace, as the sound of a lute, or any suche instrument doeth geue. Then his sentencies must be well framed, and his wordes aptly vsed, throughout the whole discourse of his Oracion. (sig. a2r)

Such ideas are restated nearly three decades later in the more famous *Defence of Poetry* of Sir Philip Sidney. Sidney begins his essay by noting that the word poet comes from the Greek word ποιειν, 'which is, to

make', adding, 'we Englishmen have met with the Greeks in calling him a maker: which name, how high and incomparable a title it is, I had rather were known by marking the scope of other sciences than by any partial allegation'.[1] But the force of poetry for Sidney leads, as it does for Wilson, not only to agreement and pleasure, but to a kind of emotional sharing, even a kind of rapture. The poet, Sidney says,

> beginneth not with obscure definitions, which must blur the margin with interpretations, and load the memory with doubtfulness; but he cometh to you with words set in delightful proportion, either accompanied with, or prepared for, the well enchanting skill of music; and with a tale forsooth he cometh unto you, with a tale which holdeth children from play, and old men from the chimney corner. (p. 92)

We have not substantially bettered these concepts in the twenty-first century, but we have deepened our knowledge of just how such poetic language comes to be; psychology, linguistics, physics, and neuroscience have all come into confluence, showing us how the human brain works – not just ours, but those of Shakespeare and his contemporaries. They have shown us the processes by which we acquire, process, and interpret knowledge; how the human brain processes language; and, most significantly and most amazing of all, how each person's processing of language is individually distinct. Word deployment is individual to a high degree; and understanding this permits us, for the first time, to address and answer, at least provisionally, some basic questions about the lives and works of Shakespeare and his contemporaries. As Harold Love comments, a personal idiolect individualizes the sociolect.[2]

'Every mental process', such as that of creating poetry, and drama, Edward O. Wilson contended in *Consilience: The Unity of Knowledge* in 1998, 'has a physical grounding and is consistent with the natural sciences'.[3] John Carey expands on this premise in *What Good Are the Arts?* (2005):

> [There] are innate operations in the sensory system and the brain. They are laid down by the joint operation of two kinds of evolution, genetic and cultural. Genes prescribe certain regularities of sense-perception or mental development, and culture helps determine which of the prescribing genes survive and multiply. There are primary epigenetic rules and secondary ones. The primary ones

[1] Sir Philip Sidney, *A Defence of Poetry*, in *Miscellaneous Prose of Sir Philip Sidney*, ed. K. Duncan-Jones and J. van Dorsten (Oxford: Clarendon Press, 1973), p. 77.
[2] H. Love, *Attributing Authorship: An Introduction* (Cambridge: Cambridge University Press, 2002), p. 222.
[3] Quoted in J. Carey, *What Good Are the Arts?* (London: Faber and Faber, 2005), p. 65.

determine the way our senses apprehend the world – the way, for example, our sight splits the wavelengths of visible light into the distinct units that we call the colour spectrum. The secondary epigenetic rules relate to our thinking and behaviour. They include the neural mechanisms of language … (p. 66)

Although our culture has changed considerably since Shakespeare wrote, our brains have not, as Carey's first sense of epigenetic operations indicates. Our process of cognition, then, is transhistorical, and in applying its operations to the sort of mental processing that is revealed in Shakespeare's language, neuroscience has taught us how to determine, through his individual handling of language, how we (even today) can determine new influences on, and accomplishments of, Shakespeare's plays. For language is not just 'a cultural artifact', as Steven Pinker wrote in *The Language Instinct* (1994), but 'a distinct piece of the biological makeup of our brains',[4] although the complicated process of cognition and subsequent representation of thought is not a conscious process. 'The workings of language', Pinker adds, 'are as far from our awareness as the rationale for egg laying is from the fly's'.[5] We can understand more precisely how language works and poetry is formed – and with it, always, the accessible individual voice – by first understanding the dynamics of the brain. Recent cognitive scientists have studied the brain by dividing it into its components: the thalamus, the hippocampus, and the cortical gyri. Each section is made up of complex networks of cells that relate, in countless possible combinations and networks, to one another, leading to patterns that, once constituted by individually processing data from the external world, will provide those tell-tale individual tics that set each person apart, even when that person undergoes the same experience as another at the same moment (seeing a movie with someone else, for instance). The basic working element of the brain is the neuron, an electrically charged type of body cell that can receive and transmit electrochemical impulses; and by long extensions of its cell bodies, neurons form connections with other neurons through synapses. In operation, neurons receive such impulses from literally thousands of other neurons – it is the huge number and their arrangements that will allow for the formation of individual responses. Some of these impulses (rather than others) will excite a particular neuron and cause it to 'fire' out its own reaction as an impulse to other neurons, while inhibiting and withholding other received impulses. As a consequence, each neuron provides a continuous analysis of the activity of large numbers of other neurons: they form, that is, their own links, never acting finally alone and

[4] S. Pinker, *The Language Instinct* (London: Penguin, 1994), p. 18. [5] *Ibid*., p. 21.

always forging combinations that may be strengthened through repeated activity. Even with a very small number of neurons, such as ten or twelve, the number of possible combinations is surprisingly great; but the whole situation is enhanced by the fact that the human brain contains at least ten billion of these tiny cells (and some neuroscientists have estimated as many as a hundred billion). Moreover, a neuron over time will form synapses with at least a thousand other neurons, so that in time there will be as many as a hundred trillion (100 000 000 000 000) in a human brain. Such a huge number is incomprehensible, but there is no question that such dynamic brain functions happen all the time and so quickly that we ourselves sense no pause between a sensation – hearing a line from Shakespeare, say – and responding to it; or, if Shakespeare, seeing a character (in the external world or in his mind's eye, his 'imagination') and verbally expressing it. But since Shakespeare processes the idea or the object in his own way, the expression will bear some stamp of individuality, too; and a scene or an act will become uniquely identifiable.

The brain's operation is unimaginably complex and sophisticated, though, so that we need a sufficient amount of evidence to essay the habits of a particular mind, just as the complexity of the DNA sample provides its own reassurances. The brain works through a densely cooperative and collaborative system, interconnecting its thousands of neurons instantaneously; yet at the same time, the billions of neurons are poised to admit new electrical charges and to form, or reform, their neural pathways that have established the brain's particularity over time. This allows for what we might see, for instance, in the change of Shakespeare's style that we would call, rather simply and superficially, his 'development'. His new experiences can always lead to new patterns, but they will always relate to older ones as well, since they are being processed by the same individually identifiable mind.

Pinker has written that

> virtually every sentence that a person utters or understands is a brand-new combination of words, appearing for the first time in the history of the universe. Therefore a language cannot be a repertoire of responses; the brain must contain a recipe or program that can build an unlimited set of sentences out of a finite list of words. That program may be called a mental grammar. (p. 22)

'The way language works, then', he continues, 'is that each person's brain contains a lexicon of words and the concepts they stand for (a mental dictionary) and a set of rules that combine the words to convey relationships among concepts (a mental grammar)' (p. 85). In compiling data in the

form of language, we need to pay attention to the lexicon (the number and pattern of common words – common to a culture, common to an individual writer), as well as rare or suddenly new words and patterns of words (pairs, for instance). To define and appreciate an individual's linguistic DNA, so to speak, we need to examine both his mental dictionary and his mental grammar.

The computer now allows us to establish the identifiable, distinguishing use of language of individual Renaissance English playwrights, Shakespeare foremost among them. The results we present in this book demonstrate the consistent style of a single author, showing how the linguistic uniqueness sets it apart from works by other playwrights. The visual displays of the data are helpful in another way, too; they show how parts of individual works by some authors can depart from the main set, inviting questions about the nature of this variation, by means of data that is concrete, specific, and rediscoverable.

Pinker's seminal study of language and language behaviour takes up this interest in special patterns – what he calls 'phrase structure' – that is 'the kind of stuff language is made of' (p. 103):

> The discrete combinatorial system called 'grammar' makes human language infinite (there is no limit to the number of complex words or sentences in a language), digital (this infinity is achieved by rearranging discrete elements in particular orders and combinations, not by varying some signal along a continuum like the mercury in a thermometer), and compositional (each of the infinite combinations has a different meaning predictable from the meanings of its parts and the rules and principles arranging them). (p. 334)

Combinations of words, then, are another way in which brains, and persons, expose and identify themselves, usually unconsciously, whether or not they wish to do so. Once neuroscientists began unscrambling and understanding the areas and processes of the human brain, linguists have been enabled to determine, more or less scientifically, how language is formed, by understanding the neural processes and pathways that are responsible.

Pinker goes further, introducing proteins into the equation that brings the brain activity into alignment with the uniqueness of DNA:

> The molecules that guide, connect, and preserve neurons are proteins. A protein is specified by a gene, and a gene is a sequence of bases in the DNA string found in a chromosome. A gene is turned on by 'transcription factors' and other regulatory molecules – gadgets that latch on to a sequence of bases somewhere on a DNA molecule and unzip a neighboring stretch, allowing that gene to be transcribed into RNA [ribonucleic acid, carrying DNA instructions for chemical synthesis to

the cell], which is then translated into protein. Generally these regulatory factors are themselves proteins, so the process of building an organism is an intricate cascade of DNA making proteins, some of which interact with other DNA to make more proteins, and so on. Small differences in the timing or amount of some protein can have large effects on the organism being built. (p. 321)

The confluence of several areas of research, then, including physics and chemistry, leads to their own kind of 'consilience', by which individuality is established with a kind of insight and a sense of certainty heretofore unavailable.

If such advanced science is staggering in its determinations and its implications, its outcome, in some ways, is not unexpected at all. How many times, for example, have we said, 'This passage reminds me of *X*', or, 'This cannot be by *X*; it seems nothing at all like him'? Carey finds such visible distinctions between Shakespeare and Marlowe, for instance, by sensing that one uses metaphor, the other tends to employ images directly, rather than comparatively. His sense is corroborated by applying something like Pinker's mental dictionary and mental grammar (although he undoubtedly arrived at his observations by some other means). This is Carey:

It is often said that Shakespeare took up where Marlowe left off, and could not have written his plays without Marlowe's example. But Marlowe is actually a completely different kind of writer, much more wooden and solid and distinct than Shakespeare for all his flamboyance. Shakespeare's superior indistinctness can easily be seen if we compare the way Marlowe's Jew, Barabas, and Shakespeare's Jew, Shylock, talk about their wealth. Here is Barabas:

> Bags of fiery opals, sapphires, amethysts,
> Jacinth, hard topaz, grass-green emeralds,
> Beauteous rubies, sparkling diamonds … [I.i.25–7]

And so on. Pretty good, you will say. Yes, it is. But it is not very indistinct, so the imagination has not much to do. You can easily picture bags of jewels. Of course, even Marlowe's lines are beyond the reach of visual arts like painting or photography. You cannot paint grass-green emeralds, except by some ponderous device like juxtaposing painted grass and painted emeralds, whereas language can merge the two in a flash. Painting cannot manage metaphor, which is the gateway to the subconscious, and that hugely limits it by comparison with literature. True, there is Surrealist painting, but it is static and deliberate, and quite unlike the flickering, inconsequential nature of thought. However, with all due credit to Marlowe's jewels, compare Shakespeare's Shylock when he hears that his daughter (who has run off with her lover, taking some of her father's gold and jewels with her) is living it up in Genoa and has exchanged a ring for a monkey.

> Thou torturest me Tubal, – it was my turquoise, I had it of Leah when I was a bachelor: I would not have given it for a wilderness of monkeys. [*The Merchant of Venice*, III.i.112]

> Marlowe could never have written that. Quite apart from the human depth, the indistinctness is what stamps it as Shakespeare's. 'A wilderness of monkeys', the lightning phrase with which Shylock registers his wit, scorn and outrage, is unforgettable and unimaginable – or, rather, imaginable in an infinite number of ways. How do you imagine it? Are there trees and grass in the wilderness? Or just monkeys? Are they mixed monkeys, or all of one kind? With tails or without? Of what colour? What are they doing? Or are these questions too demanding? Is the impression you get much more fleeting, much less distinguishable from the mere blur of total indistinctness? At all events, compared to 'grass-green emeralds', 'a wilderness of monkeys' is a wilderness of possibilities. We are tempted to say that it is a 'vivid' phrase, and it is understandable that we should want to use that word about it. But 'vivid' is often used to describe clear-cut effects, such as a bright pattern or colour composition, and Shakespeare's phrase is not vivid in that way, rather the opposite. It manages to be at once vivid and nebulous. It is brilliantly and unfathomably indistinct, which is why the imagination is gripped by it and cannot leave it alone. (pp. 216–17)

Carey has an especially well-honed literary sensibility, and he can often observe what many readers would not. But to describe jewels by visually conveying their direct appearance and to let that imply wealth alongside using an incident, even a far-fetched incident, to measure wealth – one jewel misspent is a very different sense of wealth than accumulating it – is a distinction we can easily comprehend. Assembling long lists of examples of Shakespearean uses of language alongside Marlovian uses helps us to question comparisons, too, and guard against false analogies. But the labour-intensive task of compiling such data – not to mention the margin of error – makes such a comparison daunting, if not unlikely. Using a computer to gather the evidence to be analysed by the critic, and its utility and interpretation to be determined, the fundamental process of computational stylistics gives to literary criticism (and its associated concerns such as authorship, development, or influence) the means by which we may substantially advance our knowledge of Shakespeare, his works, and even (despite our cultural differences) his differentiated meanings.

We can look at such compilations of data another way. Ian Lancashire of the University of Toronto has used the computer to assemble the data by which he can compare the language and the language usage of Shakespeare and Chaucer.

> I studied two passages from Shakespeare's works, *Hamlet* III.1 (the so-called 'nunnery scene'), and *Troilus and Cressida*, I.3.1–29 (Agamemnon's first speech), and two parts of Chaucer's *Canterbury Tales*, the General Prologue and the Manciple's prologue and tale, both in the context of the complete *Canterbury Tales*. The principal repeated vocabulary unit of both authors was the word-combination. In *The Canterbury Tales*, Chaucer used 12 000 word-forms but 22 000 repeating

fixed phrases. Over two periods, 1589–94 and 1597–1604, Shakespeare's different fixed phrases at least doubled his word types. The vocabulary of both poets consisted, not of single words, but of little networks, a fact consistent with associative long-term memory. The sizes of these networks were well within what working memory could accommodate. The 464 phrasal repetends appearing in both Chaucer's General Prologue and the rest of his *Canterbury Tales* averaged 2.45 words. They fell into 177 networks. Repeating fixed phrases in Shakespeare's texts in both periods averaged 2.5 words. Chaucer's largest repeating combination in the General Prologue (853 lines) had nine words. Shakespeare's largest in *Hamlet* III.l, under 200 lines long, had five words. A second Shakespeare analysis, of Agamemnon's speech in *Troilus and Cressida*, I.3.1–29, found 107 phrasal repetends (repeating elsewhere in Shakespeare's works) in a passage that has only 159 different word-forms. Most combinations are two words in length, and the maximum has four. It is possible that the constraints of working memory affected the quantitative profile of the verbal networks employed by both men.[6]

Such a comparison in linguistic profiles for Lancashire can characterize an author, but, by the same token, it can also help to identify him. Repetends of word-combinations (what he called fixed phrases) are another key linguistic figure made visible through computational stylistics. Lancashire also reminds us that memory plays a major role in cognitive theory and practice. Stimuli processed by the brain that are similar or even repetitive strengthen the memory of them in the brain, but the hippocampus also retains unique memories until later stimuli merge or modify them, or finally displace them as they fade over time. Such data are another legacy that advanced science, especially in the anatomy of the brain, provides for a study of the humanities at the beginning of the twenty-first century.

Historically, most of the work of computational stylistics has gone into authorship studies. Language is a shared system – it must be, if we are to use it as our fundamental means of communication – but each person uses language in a special and individual way. Literary language is only an extreme form of this self-expression. Writers, in fact, often seek to use language in new ways to express their own sensibility, their own particular vision and interpretation. This is especially helpful, then, because the data will show those particularities and can establish individual profiles of literary writers more quickly. Furthermore, when a work is anonymous, such stylistic choices and practices will help to identify authors. It can therefore be no surprise that attribution studies have been especially prevalent in computational analyses over the past several decades. But such a concern

[6] I. Lancashire, 'Cognitive Stylistics and the Literary Imagination', in *A Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, and J. Unsworth (Oxford: Blackwell, 2004), pp. 397–414 (p. 408).

is not new. Modern attribution studies actually began in the Renaissance itself when texts were abundant enough to make such comparisons possible. In the fifteenth century Lorenzo Valla showed that the Donation of Constantine was a forgery through what were then new disciplines, developed by the humanists, of philology and history.[7] From then onwards, the loose collections of writing that survived of the Bible, Homer, and even playwrights like Shakespeare, were no longer accepted uncritically. Since then, much effort has gone into determining what is canonical and what, on the other hand, is apocryphal. External evidence, such as manuscripts, licensing records, and commercial sales, which could sometimes be used to fix authorship, could thus be used in conjunction with internal evidence of a writer's style.

There are difficulties, of course. An author can limit his style, vary it, imitate someone else to pose as that person, or write a parody so dependent on the original and so different from his own style that the actual authorship of the parody, rather than its target text, is more difficult to discern. Ben Jonson, for example, is now known to have deliberately redrafted his earlier writings to align them with later material so as to give a sense of unity and cohesion to his great folio of 1616.[8] (This was not the case with King James, whose own folio appeared in the same year.) The issue of imitation is an equally thorny one. Alexander Pope, writing in the early eighteenth century, thought that it was folly to try to attribute a work to an author on the grounds of style alone; Joseph Spence thought that Pope must have had in mind the ease with which writers could adopt styles.[9] Samuel Johnson, on the other hand, told Boswell that he thought that 'every man whatever has a peculiar style, which may be discovered by nice examination and comparison with others: but a man must write a great deal', he added, 'to make his style obviously discernible'.[10] In addition, style changes over time (as Jonson had observed): it is well known, for instance, that early Henry James and late Henry James sound quite different.[11] But what computational stylistics has shown is that, even so, tests of common words, rare words, and word pairings, especially when used in conjunction, can detect the similarities that continue to ride as

[7] Love, *Attributing Authorship*, pp. 18–19.
[8] The revision of one play in the Jonson Folio is studied in detail in H. Craig, '"An Image of the Times": Ben Jonson's Revision of *Every Man in His Humour*', *English Studies*, 82 (2001), 14–33.
[9] Joseph Spence, *Observations, Anecdotes, and Characters of Books and Men*, ed. J. M. Osborn (Oxford: Clarendon Press, 1966), 2 vols., Vol. I, pp. 171–2.
[10] Love, *Attributing Authorship*, p. 7.
[11] For a computational-stylistics analysis of the change in James's style, see D. L. Hoover, 'Corpus Stylistics, Stylometry, and the Styles of Henry James', *Style*, 41 (2007), 160–89.

a foundation underneath such easily perceived changes. While even the most perceptive, informed, and experienced readers may be challenged to find the basic (and telling) consistencies, the benefit of computational stylistics is that it can, in fact, do so.

Computational stylistics, that is, can measure change in an author's body of work by balancing such changes against basic consistencies. Such studies can show how variation is 'nested' – how early and late James differ but are still more alike than, say, a contemporary such as Thomas Hardy. Shakespeare's characters, given different lexicons as a means of characterization, can also be detected as the work of the same author, since such a range of word choices and uses remains in a kind of 'envelope' of style that is demonstrably Shakespeare's. Such an 'envelope' still distinguishes Shakespeare from Jonson – or, for that matter, from Marlowe, Middleton, Webster, and Fletcher. Unexpectedly, the study of computational stylistics, and its practices, is perhaps best summarized by someone who is not known for his writing at all, but for his painting: George Braque's dictum was that 'One's style is one's inability to do otherwise.'[12] It is because of the genetic basis, as well as the cultural perceptions of each individual writer that together establish the cognitive processes, that computational stylistics can help us make new discoveries and verify or deny previous ones.

The methods allow us to put intuitions about the distinctiveness of an individual's style on an objective basis, and to estimate systematically the criss-crossing relations of authorial style and other commonalities. They may have some general value in contributing to an overall reassessment of authorship, in an era in literary studies when the individual agency of the author has been subject to intense scrutiny. Indeed, Andrew Bennett in *The Author* (2005), says that the 'crisis of authorship' (Bakhtin's term) has become central to critical debate and interpretation.[13] But now the stylistic dimension to authorship brings a new aspect to this discussion. Once such questions could be approached only through a sophisticated reader's impressions, and seemed to belong to connoisseurship rather than to literary studies proper. They could be overlooked, as they are in Bennett's book. The strength of the effects recorded in the graphs in our chapters changes the picture.

The first decision to make in any computational-stylistics study is what features to count. It is desirable that the features be unambiguous. (In doing the counting any two researchers should come to the same tallies, or

[12] Cited in H. Garner, 'I', in *The Best Australian Essays 2002*, ed. P. Craven (Melbourne: Black Inc., 2002), p. 152.
[13] A. Bennett, *The Author* (London: Routledge, 2005), p. 113.