Chapter 3
Individuation Is There in All the Different Strata: John Burrows, Hugh Craig and Willard McCarty

Abstract This oral history interview between Willard McCarty (on behalf of Julianne Nyhan), John Burrows and Hugh Craig took place on 4 June 2013 at the University of Newcastle, Australia. Harold Short (Professor of Humanities Computing at King's College London and a Visiting Professor at the University of Western Sydney in the School of Computing, Engineering and Mathematics) was also present for much of the interview. Burrows recounts that his first encounter with computing took place in the late 1970s, via John Lambert, who was then the Director of the University of Newcastle's Computing Service. Burrows had sought Lambert out when the card-indexes of common words that he had been compiling became too difficult and too numerous to manage. Craig's first contact was in the mid-1980s, after Burrows put him in charge of a project that used a Remington word processor. At many points in the interview Burrows and Craig reflect on the substantial amount of time, and, indeed, belief, that they invested not only in the preparation of texts for analysis but also in the learning and development of new processes and techniques (often drawn from disciplines outside English Literature). Much is said about the wider social contexts of such processes: Craig, for example, reflects on the sense of possibility and purposefulness that having Burrows as a colleague helped to create for him. Indeed, he wonders whether he would have had the confidence to invest the time and effort that he did had he been elsewhere. Burrows emphasises the network of formal and informal, national and international expertise that he benefitted from, for example, John Dawson in Cambridge and Susan Hockey in Oxford. So too they reflect on the positive results that the scepticism they sometimes encountered had on their work. As central as computing has been to their research lives they emphasise that their main aim was to study literature and continuing to publish in core literature journals (in addition to DH journals) has been an important aspect of this. Though they used techniques and models that are also used by Linguists and Statisticians their focus has remained on questioning rather than answering.

Biographies

John Burrows was born in Armidale, New South Wales, Australia, in 1928. He was Professor of English at the University of Newcastle, Australia from 1976 to 1989. Following his retirement in 1989 he became Emeritus Professor and Director (1989–2001) of the newly established Centre for Literary and Linguistic Computing (CLLC). As discussed below, his computer-assisted textual analysis research combined two previously separate approaches: counts of common words (often referred to as 'function words') and Principal Component Analysis.[1] His research is seminal and internationally recognised; his contributions are both to theory and methodology. Among his most important publications is the book *Computation into Criticism*: *A Study of Jane Austen's Novels and an Experiment in Method* (1987) and the article "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship" (2002). In 2001 he was awarded the prestigious Busa Award for Outstanding Contributions to the field of Humanities Computing.

Hugh Craig was born in Watford, England, UK in 1952. He was appointed Professor of English at the University of Newcastle, Australia in 2004, and Director of the Centre for Literary and Linguistic Computing (CLLC) in 2001. He has also held posts as Head of the Department of English; Head of the School of Language and Media; Head of the School of Humanities and Social Science; and Deputy Head of the Faculty of Education and Arts. His internationally recognised research is on Computational Stylistics and its applications to Shakespeare and Early modern English drama. His many publications include some of the most authoritative texts on the applications of computing to literary problems, for example, his chapter on 'Stylistic Analysis and Authorship Studies' in the *Companion to Digital Humanities* (Schreibman et al. 2008). The new knowledge he has contributed to Shakespeare Studies is brought out especially in the co-edited *Shakespeare*, *Computers and the mystery of authorship* (Craig and Kinney 2009).

Willard McCarty is FRAI/Professor of Humanities Computing, Department of Digital Humanities, King's College London; Professor, School of Computing, Engineering and Mathematics, University of Western Sydney; Editor, *Interdisciplinary Science Reviews*; and Editor of Humanist. In 2013 he won the Roberto Busa Prize of the Alliance of Digital Humanities Organizations (ADHO). In 2006 he won the Richard W. Lyman Award from the National Humanities Center and the Rockefeller Foundation, U.S. and in 2005 he won the Award for Outstanding

---

[1] This is a statistical technique that allows single points in a dataset to be examined in terms of pattern and variation. Thus it can be used to examine trends and variations in and across large dataset(s). Oxford Index defines it as a '[m]athematical technique for condensing a metabolomic spectrum to a single point on a graph, permitting rapid comparison between different species, experimental and control groups, etc.' See http://oxfordindex.oup.com/search?q=Mathematical+technique+for+condensing+a+metabolomic+spectrum+to+a+single+point+on+a+graph%2C+permitting+rapid+comparison+between+different+species%2C+experimental+and+control+groups%2C+etc.%EF%BF%BD++

Achievement, Computing in the Arts and Humanities from the Society for Digital Humanities/Société pour l'étude des médias interactifs, Canada. His work is centred on computing across the Arts, Humanities and interpretative Social Sciences. His numerous publications include *Humanities Computing* (2005), which made a seminal contribution to the articulation and design of the intellectual foundations of DH.

Interview

Willard McCarty [WMC] I'm going to go through six questions which have been asked of everybody in this project, but unusually, because there are two of you and you've known each other for a long time, some sort of cross talk between the two of you will make this a particularly valuable record of your memories, recollections and thoughts on very long careers in DH. The first question is: what is your earliest memory of encountering computing technology and what did you think of the computing technology you encountered at the time?

John Burrows [JB] [pause … laughter… pause …]. It was 1979. I'd been card-indexing examples of tolerably common words (or frequent words) in Jane Austen's novels. My card indexing system was becoming intolerably overburdened, complicated and difficult to manage, and I went downstairs to the Director of our Computing Service, John Lambert,[2] and asked him if any of this could be computerised. He told me about COCOA (Russell 1965), the software program of the day for text management and preparation.[3] He responded with great interest and enthusiasm and we worked right on from there. So, I had the good fortune to have an early positive response from a highly competent and capable man.

Hugh Craig [HC] A Remington word processor that we had in the faculty was my first contact. There was a special room there where the word processor was sitting. Remember [to JB], you were Dean and you put me in charge of that project. Now when was that? That was in the early 1980s, so that was our first bit of word processing technology.

JB 1983.

---

[2] Lambert was first appointed to a Lectureship in Mathematics and went on to be the 'University's Foundation Director of Computing Services'. See: http://www.newcastle.edu.au/research-and-innovation/centre/education-arts/cllc/people

[3] On Tapor, COCOA I is described as 'a program and markup language for generating word counts and concordances written in FORTRAN, from the University College Cardiff and the Atlas Computer Laboratory. It produced keyword-in-context concordances and word frequency profiles for texts, and was considered to be a powerful, highly flexible program for punched card computers'. See: http://www.tapor.ca/?id=222

HC OK

JB It was known as 'the Dean's white elephant'

HC Yes it did have a few problems. I remember that the daisy wheels kept breaking on the machine.

JB It had an abominable problem. I was lucky enough to be away on sabbatical while it was being experienced and came back just after it was resolved. It had the wrong board fitted (it was a closely related but not identical model). It took about 12 months and many visits from the technicians before it was discovered that all they had to do was insert the right board. Afterwards it worked admirably, by which time nobody was interested.

HC People had spent too long battling its problems.

JB And, of course, about $20,000 of faculty money which, at that time, was a considerable sum. It just predated the vigorous growth of PCs.

HC It seemed like the solution at the time.

JB So, your experience was unluckier than mine – it's a wonder you kept going.

HC Well, that was unrelated to Computational Stylistics. I must have noted your work, particularly authorship attribution, happening here. I was working on a problem on Ben Jonson's additions to the *Spanish Tragedy* and I thought "maybe we could apply this to Computational Stylistics", as I don't think it was called then.

JB No

HC And it wasn't called Stylometrics either – what did you call it at the beginning?

JB Nothing. It was just what I was doing. An American couple, I think their name was Sedelow used the term in a book of theirs in the mid-1970s. I took it up from them about 20 years later, principally because I thought by then we were outgrowing what had always been called Stylometry because we were doing more ambitious and complex things than just one-on-one contests between two candidate authors. We were doing more than Authorship Studies and I thought a new term was needed. The old term still survives but the justification of the new one is pretty obvious.

HC The old terms come back, you see it again and again. I'm not sure why, I think because new players keep coming in and they just pick up on some of the older terminology and it comes back.

JB Yes, once again it's the dearth of the history of the field which we have talked about a number of times

HC I think that it makes a big difference if you have somebody in your own institution, or even down the corridor, doing something. I think it's in terms of the sort of confidence you can have that something can be seen through or that you'll get some payoff for your investment. I'm pretty sure that if I had been elsewhere and had just heard of John's work that I wouldn't have had the confidence to invest a lot of time. I would have thought "OK, I can spend a year doing this and get nothing out of it. I don't think I'll do that!" I had known John for a long time and I think there is something about proximity and the sense that you can observe, almost on a daily basis, that things are working out and things are progressing. It's a bit hard to define but that's why I got into it and persisted.

I know that I did my first comparison in 1988. I had maybe six Shakespeare plays and six Jonson plays and the odd thing I always say is that I prepared Hamlet for that study in 1988 and I've probably used it at least every week since then – that same text over and over again. It's a great advantage if the texts are not just a one-off; they're almost not worth it for a one-off study. There's such a big investment in the preparation of the texts in order to do it properly. I think that's true even today; you might get a database from somewhere yet you nearly always need to add something. So, it really pays off when you keep re-using your material. In my case, I've just been able to keep building it up to, I don't know, 225 plays or whatever. But the core ones are still being recycled – I won't say daily but weekly, almost. And Hamlet is still there.

JB Another piece of serendipity in my early days was my first author. As I said a while ago, I'd been doing card indexes looking at Jane Austen's language and she just absorbed this sort of punishment. She always rewarded you with an interesting answer to your question. If I'd tackled some other stylistically-duller author I'd probably have given up long before, but she just kept seducing me, which is something Ms Austen might not have wished to hear. And then, shortly after my conversation with John Lambert in August or September 1979, I went off to Cambridge for a year and had the good fortune to meet John Dawson [the Manager of Cambridge University Literary and Linguistic Computer Centre]. Through John and another man whose name escapes me at the moment, Robinson I think, I was told about Susan Hockey (see Chap. 6) and the work in Oxford University Computing Services. So, between Dawson's center in Cambridge and Hockey's center in Oxford, I was doing a lot of criss-cross travelling in the course of that year and learned a great deal and got a lot of encouragement and support to continue. I think I struck a lot of good luck, in a number of ways, early on when one might have been discouraged.

WMC Yes, stories of good luck are to be expected. The second question is – I'm not sure why it's here really – about formal and informal training. Both of you began when there was no formal training, I know I began when there was no formal training, and you've already answered more or less how you learned. But I'm wondering

if you could comment a bit more on the process of picking up this set of technologies and what that was like, and your relationship to John Lambert, in particular, because you had somebody in the computing center who was sympathetic.

JB The relationship with Lambert was enormously important throughout those years. You know, by 2001 he had been retired from his post as Director of Computing Services for 6 or 8 years. To amuse himself, and to earn a bit of money on the side, he became our programmer in our center. He worked actively with us right up to the time of his death. The prototype software that he designed for us called LILAC (Literature, Language, Computing) I use every day still. It was never refined as he would wish to see it but he left us a good enough working model. Now, the essence of that part of it, I think, is how much support the Humanities person needs from the computing person, Lambert or Dawson or Hockey, while he's finding his feet. It was 15 years before I could do much work by myself, on my own, without referring to somebody else all the time. So the training was a long, long slow process. Admittedly, in my own defence, I was a busy person at the time and doing a lot of other things.

HC It's just as well we weren't Statisticians and it's just as well we weren't Linguists because we would never have started. You know, the Statisticians would have been worried about normal distribution of the data, about not having enough. I think we would have been too inhibited. If we had been Linguists we would have been interested in Chomsky and Universal Grammar and not at all interested in data, as Chomsky wasn't. We never got much buy-in from Linguists. The best buy-in was probably from Statisticians once we had accumulated quite a lot of data. I think that if we'd been trained–[to JB] I don't know what you think about it–Statisticians or Linguists, we would not have thought to do this kind of stuff because it was very exploratory, and no one would have held out much hope of finding any interesting patterns. Let's say, the less training the better.

JB I quite agree. I got a great deal of support from our Professor of Statistics, Annette Dobson, who was sympathetic to my ignorance and stupidity. I had good help from statistically-informed friends, but I agree with you here. On the whole, the more strongly expert people were statistically, the more inclined they were to want us to use methods that yielded definite answers: yes/no answers. Our interest was rather more in exploring to find out what the answers might be, and what questions they might provoke. The finality of a Linear Discriminant Analysis,[4] for example, was never really suitable to our need: it closed the question, but we didn't want it closed. We wanted to go on thinking about why it should look like that.

HC Principal Component Analysis (PCA) was just the key, wasn't it? It was a beautiful way of combining the multivariate (combining all those different variables

---

[4] According to the *Oxford Dictionary of Statistics* (2008), Discriminant Analysis is a 'procedure for the determination of the group to which an individual belongs, based on the characteristics of that individual'.

in an exploratory way) and letting the data speak for itself. PCA does this beauti-
fully, as opposed to Discriminant, which wants a closed answer. It over trains and is
over optimistic and gives rise to all those problems. [To JB] who put you on to
PCA? That was really fortunate.

JB Nick McLaren in the Cambridge Computing Laboratory. Then friends of mine
refined the rough model that McLaren had suggested and taught me how to use it
better.

HC Nobody thought that function words would give you anything because every
one used them at the same rate and they were empty words, or stock words, you
know, classically.

JB That was me! That was just one bit of all of this. Poor judgement, good luck,
and personal friends, [laughing] and I mixed teaching with it. Unexpectedly ordi-
nary, boring, empty little words seemed to be doing a lot and that's where the card
index broke down, of course, because one can't index *and* and *of* and *the*. Once I got
it into the computer setup I was able to explore what did happen to *and* and, *of* and
the. Much to my surprise, and everybody else's I think, the result was just as inter-
esting as the result from ostensibly more interesting words of the kind that Stylistics
has been much more focussed on. So, we got through to a layer that could not have
been seriously penetrated without the computer.

HC Yes, that part of language was waiting for the computer to arrive so that it
could become visible. Then PCA somehow went beautifully with function words;
that was John's winning hand, function words plus PCA.

JB I always expected to be completely overtaken and surpassed by some wealthy
American Institute, and it never actually happened. More luck! I seem to be in a
benevolent frame this afternoon.

HC We still come back to function words and PCA. You know, one goes down to
the more interesting words and lots of people find ways of doing that, as we have
ourselves, but then you come back to function words: they're abundant, they've got
a good distribution … they're like the very fabric of language, aren't they?

JB And they not only require a computer, they also require statistics to handle
them.

WMC People are always asking the question you've just answered, which is,
where has the computer made a real difference that no-one could have made by him
or herself – this is a very important point.

JB You can imagine a Victorian Parson mad enough to count up all the *thees* or all the *ofs* but he could never have done multivariate things with them. The first of those two steps is lunacy but the second would have been impossible.

HC And you would have probably just done Shakespeare, and never been comparative, which is the other great thing. That's why I almost challenge John about Austen – if you'd started on Dickens you might have got something of the same. If you're inside that author, you sort of feel that author is the world.

JB Austen is not alone. A comparatively small number of authors have a really strong stylistic gift but I don't think it works for the common run of writers. Nothing in my work would support the idea that it works in the commonest authors.

HC Down to the finest levels of character or progression of characters, yes.

WMC There's another important point there that I picked up on as you were talking. That is your relationship to the other disciplines that touched your intellectual lives, a glancing or peripheral relationship, which, had it been intimate, would have paralysed the work. But, starting from literature you went out and picked up things here and there where they helped the work. That would not have happened with any other kind of relationship.

HC Yes but it is very dangerous because you are working on instinct rather than training, which is risky, certainly with statistics.

WMC But it's a well-educated instinct.

JB I think we're fortunate that we never really wanted to do anything other than study literature; all of the other things were ancillary to that. That central purpose literally questioned the questions of a literary scholar. They might have been the questions of a Linguistics scholar or an Historian, or whatever, but for us they were, have been, and continue to be the questions of literary people

WMC We continue with the question of influence in your career. You spoke of one or two strong influences but who else gripped you, including those at a distance such as people whose books you read?

JB On the whole they were not in computing. As I said, I picked that up *en passant* as time went along. Background influences … let me think. I was enormously impressed by Wolfgang Clemen (1977) on Shakespeare's imagery. I took up detail of the figures and showed how they worked through the plays dramatically. In a way, I think what I am doing is something like what Clemen did except that I am doing it with words rather than with images. And I might say, by the way, looking back to an earlier answer, both Hugh and I laid some emphasis on the function words. A lot of the work was done on them early on and is still, to a very large

extent, about them. But increasingly the other words have come into play. As we've developed a better understanding of what we're doing our vocabulary has spread from the bottom up, rather than the top down. So, we are enriching as we proceed, or so we like to think. So, Clemens was one.

I was also enormously impressed by Erich Auerbach's *Mimesis* (see, for example, Auerbach 2013), partly because of the way in which it was written by a refugee in Istanbul I think it was, during the war, quite without a proper library. He had just a few texts and had to simply write out of his head about what he thought of some of the books which meant most to him. A remarkable study. Those two. Then, afterwards, the New Critics generally

WMC Richards and Ogden for example?

JB Not so much the English ones – the Americans. I didn't ever warm to Richards. I didn't quite find his wavelength or he didn't quite find mine. I can see his importance but he didn't really speak to me. But some of the Americans did. Reuben Arthur Brower's *Fields of Light* (1951), was terribly important to me, you know. All in all, the main influences on me had to do with close reading: the world in a grain of sand.

HC I don't know that there's anyone very close to what we do who has been a big influence. I've lived through deconstruction and post modernism and those sort of eras but in many ways I probably define myself against some of that work. I'm very fascinated by it, it's definitely embedded in my thinking, but a lot of what I've been doing is trying to push back against that sort of work. But I was very influenced by New Historicism in our own area – that is, the renaissance area – people like Stephen Greenblatt and so-called Cultural Poetics, which is a good broadening from close reading. I don't know, a lot of that doesn't relate directly to what we actually do. We had a really nice visit from George Hunter, G. K. Hunter, who did a literary history of renaissance English, you know.

You're always looking for people who have a broader, more conspectus view, because that's, I think, what the computational stuff does well. It's extensive rather than intensive, which people, I think, have struggled with, because we're so used to the intensive. But the real gain is from the wide sweep so one looks for people there. Robert Weimann, a German scholar, latterly does that kind of thing (Weimann and Bruster 2010) and has some broad perspectives. But we're often looking for myths to bust so almost you read these people to find a reasonably categorical statement, preferably slightly quantitative, which can then be tested. So that's a strange form of influence! It's like "give me something I can disprove". I suppose I've a vaguely oppositional perspective on what would otherwise be regarded as influences.

WMC What about other people who were using computers in their research when you got started. Were there any and did you draw anything from them? Do you remember what their views were of what you were doing, or of computing generally?

JB I didn't have a lot of close contact, partly because there wasn't much else going on in Australia at that time. I just had the brief relevant periods in Cambridge, so, on the whole, not. I did learn a lot in the late 1980s and 1990s at conferences with people like Susan Hockey and Paul Fortier.[5] I heard some splendid papers here and there, at the conferences, but on the whole not much in the way of close contact because there was never anyone much close to me until Hugh came along and that made things more interesting because we began talking together and working together.

WMC How about the people here in the computing center? You mentioned Lambert, what did you think about the computing center and the relations for a person like you with the people in it.

JB Well, it was only one-on-one, Lambert and me. I'd go down and talk to him, or his deputy Paul Butler was helpful at times, but on the whole, I didn't have much connection with the center as a center or the service as a computing service. My contact was much more with the Director so that it was a personal affair rather than a departmental one.

WMC What about you, Hugh? What about the other people using computing at the time and your closeness to them or distance from them? I know that in my case I actively disliked most of the people having something to do with computing for many years!

HC I didn't have any strong feelings that way. I guess the center was already providing that sort of ambience and technical support so it was already well in train. I didn't have to do much pioneering there. We had Alexis Antonia already there as a wonderfully patient person, and a Linguist, to help with preparing texts. Certainly no negative experiences; it was fairly restricted really. There weren't a lot of competitors, not a lot of opposition, so…

JB The journals that were important were *Literary and Linguistic Computing* particularly but *Computing in the Humanities* as well. It's the only field in which I've ever worked where people really seem to read each other's articles. In English Studies, I think on the whole, this wouldn't be altogether true. A lot of people write for the standard academic necessity of writing but don't on the whole interchange ideas with each other and they don't much care what the other fellow is writing. That's putting it too strongly, but I feel there's a step difference between the inter-

---

[5] Paul Fortier (1939–2005) was University Distinguished Professor of French at the University of Manitoba, Canada. The European Association for Digital Humanities established the 'Paul Fortier Prize' in recognition of the many contributions he made to Humanities Computing. See: http://adho.org/awards/paul-fortier-prize

relationships in English Studies and those in the general area of DH where people really do seem to know what other people have said in the journals.

WMC I've heard this said before too with respect to the friendliness of the people and in the degree in which they want to relate to each other. I know that was my experience when….

JB And not too much belligerence either of the kind that's so common, for example, in Classics where so many of them hate each other. There have been some notable attacks on generally deserving objects but I don't think that there's much general belligerence at all.

HC The interesting relationship I reckon is with our English colleagues in the English department or discipline or whatever. That's been the most potent one for me, like trying to persuade them that this is a worthwhile activity, and you're actually learning something this way. I don't know if they ever quite got persuaded, but we're keeping on trying.

JB The scepticism is enormously useful!

HC Yes, so we have a number of very, very bright and learned colleagues who we found hard to persuade (but we kept on trying) and that's a very good sort of proving ground. I think some of them are half way there. They're half way to the point that they can see that there is some value in it but they wouldn't want to do it themselves, and I guess it's slightly disappointing. It would be nice to get a few more over the line and for them to say "I can see it's valuable and I'm prepared to spend the next 6 months doing it". I certainly learned a lot about trying to persuade close colleagues that I really respected that this was something worthwhile and still get the reaction that it's an awful lot of trouble to learn so little. Then you have to persuade them that it's little but at least it's something you know, if you know what I mean, whereas you can make a grand statement, as they like to on the whole, which is just worthless.

WMC It's a little bit at a time.

HC Yes, and what there is, is solid. It's not likely to be reversed in a hurry.

JB I think your father rates a mention, doesn't he, as a shrewd questioner and challenger?

HC Yes my father is a good mathematician and so I worked with him doing some PCAs. I don't think I could do it now, but, you know, diagonalising the matrix and so on was good in the early days for making sure you really understood what was happening.

Another great question came from Anne Barton (who was a very good Ben Jonson scholar) in Cambridge. Those were the very early days when I was trying to persuade her of a certain thing. She said "yes, this sounds fun but I just don't know how much faith to have in your results", which was a brilliant! "I can see technically it might be ok, but how much faith should I have in it when as a reader I might think something differently?" We've all sort of lived by that, you know.

WMC   John, you used a phrase that I really liked about the mounting evidence that this multitude of weak markers is something secure, that they add up to a view of literature which is probabilistic and, well, in my words, the ground is getting more solid. The mounting evidence and the patience over time in advancing step by step (and I think it is advancing) was brought to mind by your comment about the little things versus the grand statement.

The last of the required questions is about conferences. You've mentioned a bit about conferences, I suppose that the size of this country and its distance from where most of the conference activity and literature goes on meant that there weren't a lot of them. When did the conference engagement with this kind of work begin, and what was it like?

JB   Well, I gave a paper to a conference in Adelaide, the Australian branch of the MLA, AULLA in 1974, and I just talked about some word counts in Jane Austen. Someone said "have you tested this at all with anything like the chi-squared test". I said "no, I don't know anything about that, I just count on my fingers [laughing]. I used a simple word counter and here are the comparative results".

It wasn't until afterwards in Cambridge that I began to understand a bit about chi-squared and a few other things, 5 or 6 years later. Overseas conferences for me, in this field I mean, began in San Francisco in what must have been 1981 at the big ACH/ ALLC conference of that year. From then on I went to it around every second year for a dozen years or so. After I retired I eased out of them. I found them well worth doing, I enjoyed the people and the papers. It was very arduous – I was Head of Department a lot of that time, and then Dean. I'd be away for only a week in Australia or America or Europe and return to a desk full of work. Pretty sore, but it was worth it. Any particularly memorable one? Yes, Columbia, South Carolina, [to WMC] you were there with me at the time. Georgetown in 1993? You and Harold were both there.

WMC   Christ Church, Oxford, in 199[2] – that's where I met you.

JB   Oh yes, that's a good one too. Christ Church, that's right. We were both together at Columbia, Willard and I, but we really met in 199[2].

Harold Short [HS]   New York, 2001 was memorable for lots of the rest of us, John

WMC   Well, Christ Church in 199[2] was the first time I'd ever heard John talk and I went up to him afterwards and I effused in my typical fashion. We've been friends ever since [laughing].

HC It's a good beginning …

JB How not? [laughing]

HC For me, conferences in Renaissance literature or Shakespeare, or whatever, have been just as important. And that's where I feel the work really has to be done. It's very good to learn about what other people are doing technically and so on at the DH sub-conferences. But one of the things that I think makes us distinctive is this: we made a resolution at some point that we would always try to get articles in good journals in our discipline and that those are the people we really wanted to persuade. It's still the quest! But I think we've been distinctive in always trying to keep that link with the discipline and keep persuading our colleagues. Perhaps to no avail, but …

JB We've had a victory or two, but not a huge number, the mainstream journals are still very hard to persuade

HC Yes, but increasingly they are more open – definitely the best ones are.

JB It's beginning to be said in America that DH is the next big thing – be nice if that were true [laughing]!

WMC The last question is my own and off-piste. If you look back on what you've done and what has happened in your field since you got started, what has happened that you think is really important? Can you use that to pick out a trajectory for the field, or more than one trajectory for the field, into the future? Not in terms of pre-dicting the future but in terms of recognising the possibilities that are now before us? What about the past really comes out to you as important, and in using that, what do you see for the future?

JB I'm not dodging it, I'm letting you go first. It is the future

HC Is it what one's self has done?

WMC Yes, start there with what you've done and what you think is important for the future of the field. Something you're proud of; something you are ashamed of [laughter]. That kind of autobiographical sorting of the past to try to pull from it something that we've learned, that makes our choices in the future more like another step in a trajectory.

HC I don't know if I can respond to that question!

WMC I'll think up another question!

JB I think that [pause] it's all empirical at present, and to my mind, that is generally speaking a very good thing. We've learned a great amount about the details of the

ways in which language works. As Hugh was saying a while ago, it's becoming increasingly possible to reach out to larger and larger corpora as the capacity and speed of computers improves, so that we're able to do more with less and we know much more than we did about the intricacies of stylistic patterns.

It's never-the-less true that as Argamon (2012) says – I don't agree with his derogatory way of saying it, he says that the field is a mess – that there have been some major achievements. Now, the field is a mess, he thinks, because no-one has a deep understanding of the patterns at work and what they really mean. What that deep understanding might be, I don't really begin to understand, so for me that's a very good question for the future. I'd like to meet the person who is going to offer answers that speak to me. I don't know what form that will take.

For me, there's never been any surprise in the idea that authors should be identifiable by their style, or patterns in their language, any more than if you and Hugh and two or three people come along a corridor towards me, I don't have to stop and think, is that Willard? Is that Hugh? Is that Harold? Everything about you speaks to me: the way you move, the way you dress, the way you speak and the way you eat. We're like that. We have so much in common, we humans, but we are certainly different in so many ways. It is not in the least degree remarkable to me but people seem surprised and surprised and surprised that our own individuality should speak through and beyond and out of our community. That's the sort of big understanding that I would understand, but I don't think that's what Argamon (2012) means. I would like to know what this other deep understanding of what it is all about might look like. I don't know if that's an answer…

HC   That's exactly what I think. I remember when you were working on that article and commented on how there is different individuality, that your own individuation is there in all the different strata. You said something that crystallised that whole issue of language individuation, which sort of is the answer to the idea that the author is dead, and all the rest of it! It's the empirical answer that people do in fact make their own language, or idiolect, out of languages. That gave the underpinning for a whole lot of work, not only on authorship but individuation in general. But I think we've worked through that; I mean, it's so obvious once you do it that the battle is almost won.

JB   Except that people don't believe it.

HC   No, I'm satisfied. I think everyone sensible is satisfied – it just makes obvious truth. I think that's a real contribution that Computational Stylistics has made: to have that broader idea and then work it through in a whole mass of different studies which show that authors can be distinguished. Linguists are still not very interested in the individuation of language; that's not what they do. They like much more general things about languages or even about sociolects or whatever. I feel we've probably done the individuation work and I don't know what the next phase is beyond that. Some people feel that the work of Computational Stylistics is to endlessly prove that authors are different and that Computational Stylistics can show that, but

I think we're sort of bored with that. That has been demonstrated, it's as rock solid as anything can be, it's no longer the mission of Computational Stylistics and it would be good if it was disassociated from it because we've got our answer.

JB Now, there's double spin here, isn't there? On the one hand, what you say in principle is absolutely true. On the other hand, for me at least, the particular problems of authorship remain fascinating, because so many of them are unresolved. So, I no longer feel that there's any need to demonstrate that it can be demonstrated, but I still passionately believe that the real interest, and the real challenge, lies in the particular problems themselves. However, it's not just authorship attribution, individuation is larger and more interesting than that. My own work will probably continue to be mostly in authorship attribution and individuation – I think the larger issues are fascinating and maybe there's room for a lot more work there

HC There is always to and fro between attribution, which is the bread and butter of Computational Stylistics and continues to ground or authenticate or validate its work. And then there's always the temptation or interest in something beyond authorship. So, I guess one continues to go back and forth between those two. And the great thing is that authorship is a very good testing ground because people are really interested in the answer and you can't muck about. You can't do too much hand-waving about very general concepts. And it's one area where people will actually go back and check your sums, people like Jackson (2002). I think that is quite unusual, certainly in our area, because most people will accept tables and numbers. That's the good thing about authorship attribution, it gets peoples' interest to a very profound degree. But, it's not the whole of the possibilities of the field.

WMC What strikes me is that it really doesn't make any sense at all. In one of his books Ian Hacking remarks that the great achievement of twentieth century physics is the realisation that nature is probabilistic. The fact that you two have shown that literary language is probabilistic means that we, as authors, are operating in the natural world as the natural world operates. And that's more of a question than an answer; I think that's a really interesting question.

JB It's a very elegant way of saying what I was fumbling with there, about the character of individuality.

WMC The fact is that we are an intimate part of the natural world and have been pretending to be separate from it for a very long time. We are an intimate part of it down to the most elusive of aspects of artistic expression – style. As you say, it's instantly recognisable when you're walking down a hallway and you don't have to pause to know who it is …

HC Yes, that concept of style, however elusive, is the other key or one of the keys, definitely.

WMC How far can we take this? How far can we take that probabilistic bond that we have with the natural world? Sociologists have been puzzling over this in large crowds of people and such for a long time. But there's a continuum here that seems to me to be a really interesting question. That is what I say is the significance of your work John, when I'm asked or when I can say it.

JB Think of those people with good musical memories – they can recognise something in a phrase or a couple of bars. Perhaps that is what style is like? But, you do it in tennis, you do it in cricket – it's circumambient. We're part of the natural world.

HS One of the things that I think of is your most significant achievement, actually you touched on it early in the interview, is that the purpose is questions and not answers. In an authorship study you are trying to establish *an* answer. The method and the results are actually about the questions. In a world that often gets far too fixated on the quantitative as a way to answer questions this keeps this work rooted in the Humanities. I think that has been incredibly valuable for the field and continues to be.

JB On Authorship again, here's an idea that has always meant a lot since it first came to me from Hugh 20 years ago. When you make a proper attribution of a poem, you're fitting it into an interpretative nexus where it makes more sense than it would if you had tried to force it into some alien nexus, and that's when it gets interesting. You get it into its proper home and then you see that the shape of the home has changed a little bit, and so you go on again. So it's not just "yes, this poem is Rochester's", it's what that means to Rochester.

HC Well, there's some connection there with computing power and speed, which is that in an older method you had to construct a sort of a test and a hypothesis and then you could painfully run that through, get an answer one way or another and then maybe try again. That was a very rigid structure. Computational power means that you can do that exploratory data analysis, change a parameter, re-do it, and then it becomes open in the way you're describing. And I think that's what people perhaps don't realise. They say "can your program tell you, or could your program tell you who wrote this book?" That shows no understanding whatsoever of the processes involved in doing a complicated authorship problem.

JB Our friend, Harold Love, used to say that after you finish the computation and calculations, and all the rest of it, that's when the brainwork begins.

HC Yeah, but then you can re-do them all.

JB One chap at one of those symposia was very distressed by the way you're talking right now. He said "how you can call it an experiment when you change your minds a dozen times in the course of a morning, and come at it from so many different angles and different ways? That's not an experiment!" [Laughter].

WMC We know now from really good work in the History of Science like David Gooding's on experiment (Gooding et al. 1989) that that's exactly how experiments are worked. That's the second thing that I think is really important about this work: it exemplifies the experimental method which is brand new to the Humanities. When I started, you planned your computer program out really, really well and did a flow chart and all that. Then you took your deck of cards down to the computing center and if you were really important like a Nobel Prize physicist, you could get your answers back in a couple of hours. Otherwise it was 2 days or a week, only to learn that you'd made some keypunch error [laughter]. It was only the hackers at MIT who had talked about the hands-on imperative who understood. They were sitting at the console, playing with the computer from midnight until eight in the morning. They understood this experimental method, which now we have because now you have these small machines. But, your point about the idea of experiment is really important.

WMC Well, thank you two very much for the interview.

References

Argamon, S. (2012). [Review of] scalability issues in authorship attribution [by] Kim Luyckx [sic]. *Literary and Linguistic Computing, 27*(1), 95–97.

Auerbach, E. (2013). Mimesis: The representation of reality in western literature (1st Princeton classics ed.). Princeton: Princeton University Press.

Brower, R. A. (1951). *The fields of light: An experiment in critical reading*. New York: Oxford University Press.

Burrows, J. (1987). *Computation into criticism: Study of Jane Austen's novels and an experiment in method*. Oxford/New York: Clarendon.

Burrows, J. (2002). "Delta": A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing, 17*(3), 267–287.

Clemen, W. (1977). *Development of Shakespeare's imagery* (2nd Revised ed.). London: Routledge.

Craig, H., & Kinney, A. F. (Eds.). (2009). *Shakespeare, computers, and the mystery of authorship*. New York: Cambridge University Press.

Gooding, D., Pinch, T., & Schaffer, S. (Eds.). (1989). *The uses of experiment: Studies in the natural sciences*. New York: Cambridge University Press.

Jackson, M. D. P. (2002). Pause patterns in Shakespeare's verse: Canon and chronology. *Literary and Linguistic Computing, 17*(1), 37–46.

McCarty, W. (2005). *Humanities computing*. Basingstoke/New York: Palgrave Macmillan.

Russell, D. (1965). A word-count and concordance generator. *Atlas Computer Laboratory, Chilton: 1961–75*. Available at: http://www.chilton-computing.org.uk/acl/applications/cocoa/p001.htm. Accessed 8 June 2015.

Schreibman, S., Siemens, R., & Unsworth, J. (Eds.). (2008). *A companion to digital humanities. Blackwell companions to literature and culture*. Oxford: Blackwell Publishing Professional.

Short, H., et al. (2012). Collaboration must be fundamental or it's not going to work: An oral history conversation between Harold Short and Julianne Nyhan. *Digital Humanities Quarterly, 6*, 3. http://www.digitalhumanities.org/dhq/vol/6/3/000133/000133.html

Upton, G. J. G., & Cook, I. (2008). *A dictionary of statistics* (2nd rev ed.). Oxford/New York: Oxford University Press.
Weimann, R., & Bruster, D. (2010). *Shakespeare and the power of performance: Stage and page in the Elizabethan Theatre* (Reissue ed.). Cambridge: Cambridge University Press.