

J. M. Coetzee's Work in Stylostatistics

Peter Johnston <petejohnston1979_at_me_dot_com>, Royal Holloway, University of London

Abstract

Though better known for the novels that eventually led to his being awarded the Nobel Prize for Literature in 2003, there is a strong but less widely recognised case for claiming J. M. Coetzee as a significant figure in the early development of digital humanities. In light of the recent renaissance of computer-assisted statistical approaches to literary analysis, the present article charts in detail a formative period (1969-1973) during which this most important of novelists first adopted, then deconstructed, and eventually rejected one of the discipline's earliest incarnations.

1 Introduction

Thanks in no small part to his first two works of “fictionalised autobiography” [Coetzee 1997] [Coetzee 2002] and J. C. Kannemeyer’s recent authorised biography [Kannemeyer 2012] the basic plot-points of J.M. Coetzee’s early years are now widely known. Born in Cape Town in 1940, he spent much of his childhood in the more rural Cape Province town of Worcester. After graduating from the University of Cape Town (UCT) with honours degrees in both English literature and mathematics, he travelled to England, where he put his mathematical skills to work as a computer programmer; his literary ambitions manifested themselves at this time in a master’s dissertation on Ford Madox Ford, and the search within himself for an authentic poetic voice. Notably, Coetzee’s fictionalised account of his younger self reveals in some detail the mathematical inclinations of the adolescent who, to adapt Wordsworth’s dictum, would be father to the novelist: he is “studying mathematics” [Coetzee 2002, 4]; he “assists with first-year tutorials in the mathematics department” [Coetzee 2002, 2]; he sees himself as “a mathematician to be” [Coetzee 2002, 20]; he desires “to study pure mathematics to the exclusion of everything else” [Coetzee 2002, 22], believing that “pure mathematics appears to be the closest approach the academy affords to the realm of the forms” [Coetzee 2002, 22]; he is convinced that literature is not “as noble as mathematics” [Coetzee 2002, 53] and wonders whether he will “become like those scientists whose brains solve problems while they sleep” [Coetzee 2002, 144]. Later, as a genuine self-awareness finally begins to dawn upon him, he begins to worry that the Atlas computer that is by now his most frequent companion, as well as his partner in the production of poetry, might “burn *either-or* paths in the brain of its users and thus lock them irreversibly into its binary logic” [Coetzee 2002, 160]. This prospect, combined with his desire to find “the moment in history when *either-or* is chosen and *and/or* is discarded” [Coetzee 2002, 160] contributes to an epiphany of anti-rationalism that would be voiced in one form or another throughout his body of fiction: “Death to reason, death to talk! All that matters is doing the right thing, whether for the right reason or the wrong reason or for no reason at all” [Coetzee 2002, 164].

While the evidence provided by Coetzee’s fictionalised autobiographies is of course rather compromised by its pointed generic instability, it remains somewhat surprising how little detailed attention the critical discourse surrounding his work has paid to his clear and profound inheritance from his engagement with mathematics and, moreover, computer science. Even so, though some of the more widely cited long-form critical works on Coetzee’s writing make no reference to these subjects at all (see, for instance, [Kosew 1996] and [Attridge 2004]), the discourse has not been entirely silent regarding its place in his intellectual development. As each year passes, a growing majority of critics and interviewers have begun to recognise the value of mentioning Coetzee’s undergraduate studies at Cape Town, often if only as part of

a brief biographical summary that may or may not also reference his short-lived career as a computer programmer and his postgraduate work in America: these include [Vanzanten Gallagher 1991, 10]; [Begam 1992, 419]; [Canepari-Labib 2005, 24]; [Head 1997, 1]; [Silvani 2012, 26]; [Fromm 2000, 337]; and [Phillips 1998, 61]. Moreover, and while certainly still in the minority, there exist some critical responses that treat mathematics and computing not just as mere biographical happenstance, but rather in some sense as responsible for certain aspects of Coetzee's literary work: the most valuable of these are [Scott 1997, 82–102]; [Mulhall 2009, 39–41]; [Lamb 2010, 177–183]; [Jahn 2007, 43, 64–66]; [Attwell 1993, 128]; and [Egerer 1999, 96–101]. Given both Coetzee's contribution to the early development of digital humanities and his pre-eminence as novelist, the lack of a detailed outline of his history within the discipline — either within the digital humanities community or in the wider field of literary studies — is an oversight that clearly requires rectifying.

One reason for the lack of focused attention on Coetzee's engagement with these subjects in his fiction might well be the fact that his work in this area now seems rather dated, a relic of the structuralist fervour of the late 1960s against which he had himself quite violently revolted by the time he came to publish his first novel, *Dusklands*, in 1974. Hemmed in by the technological limitations of the age, Coetzee's initial methodology — as well as his later critique of that very methodology — almost inevitably concentrates on the stylostatistical analysis of lexical features, as was characteristic of the period, rather than the more complex linguistic and paralinguistic phenomena that constitute the data available to critics today. As such, the particular methodologies he discusses seem relatively superficial in retrospect, leaving open the question as to how far his specific criticisms might apply to more recent, more sophisticated approaches. Indeed, as Hoover (2008) neatly summarises, and largely as a consequence of the subsequent explosion in accessibility to digitised texts since Coetzee's time in the field, much has changed for the better; so, while critics such as John Burrows and Hugh Craig might broadly represent a continuation of the discipline as Coetzee conceived of it, their contributions to the field of computational authorial attribution have produced consistently convincing results and reliable, widely-used standard statistical measures such as Burrows' "Delta" [Burrows 2002]. Moreover, recent developments have moved away from the simple word-frequency analysis with which Coetzee's work principally engages, tending to focus instead on multivariate approaches such as principal components analysis [Holmes et al. 2001], cluster analysis [Stewart 2003], discriminant analysis [Forsyth et al. 1999], and correspondence analysis [Paling 2009]. Even if one were ultimately to decide that such work still innately suffers from the malaise Coetzee identifies, Stephen Ramsay's "algorithmic criticism" offers an alternative philosophy altogether: the "algorithmic critic", he argues, ought to "refocus the hermeneutical problem away from the nature and limits of computation (which is mostly a matter of methodology)" [Ramsay 2011, 9], actively imagine "the artifacts of human culture as radically transformed, reordered, disassembled, and reassembled" [Ramsay 2011, 85], and approach computational stylometry having already dispensed with any inappropriately scientific desire to achieve singular, definitive conclusions.

Ultimately, though, and putting the minutiae aside, the present-day significance of Coetzee's contribution depends to a large extent on how successfully the discipline as it stands today negotiates Coetzee's deeper, more philosophical query: if one can largely eradicate the subjective biases of traditional literary criticism by restricting one's statements to the near-neutral categorical definitions of mathematics, then how can one return from the mere statement of numerical values to draw meaningful natural-language conclusions, without allowing those subjective biases to creep right back in? With this in mind, the principal purpose of the present essay is therefore not to directly assess Coetzee's work within and around the field of stylostatistics against today's standards, but rather to provide a narrative account of this work, starting with his first published journal article, "Statistical Indices of 'Difficulty'" [Coetzee 1969a], leading through close analysis of his doctoral thesis [Coetzee 1969b], pivoting upon a review of Wilhelm Fucks' *Nach allen Regeln der Kunst* [Coetzee 1971], and ending with an essay on Samuel Beckett's *Lessness* [Coetzee 1973b], which appeared in the year before the publication of *Dusklands*. While this account makes few direct connections between this early critical work and Coetzee's fiction, then, it is intended as a resource through which such connections might later be drawn.

2 Stylostatistics and "Statistical Indices of 'Difficulty'" (1969)

Appearing in a 1969 edition of *Language and Style*, *Statistical Indices of "Difficulty"* constitutes an extension of the work of the renowned German stylostatistician Wilhelm Fucks, and provides a succinct statement of the twenty-nine-year-old Coetzee's largely optimistic approach to the field of stylostatistics at this time. Published just two years later, though, his

review of Fucks's *Nach allen Regeln der Kunst* demonstrates a stark shift in his thinking: by 1971, Coetzee's various explorations into the ways in which statistical analyses can be used to systematically harness and codify the qualitative in quantitative terms had refocused his thought to such an extent that, far from endorsing such a process, he had now come to recoil against its potential ramifications as a means of manipulation in contexts both social and political.

Put simply, we can understand Coetzee's use of the term "stylostatistics" to refer to the branch of stylistics concerned with those features of a text's style that can be subjected to numerical analysis. The principal aim of the stylostatistician is to strip away the subjectivity implicit in other types of literary criticism, leaving only quantitative propositions that, in Coetzee's words, "will not carry the kind of connotative freight that, for example, the proposition 'A's language is dense' came to carry in Scrutiny criticism" [Coetzee 1969a, 226]. Where the Leavisite tradition to which he alludes here founded its criticism on the notion that "it is upon a very small minority that the discerning appreciation of art and literature depends" [Leavis 1930, 12], the stylostatistician attempts to evade such cultural elitism by constructing formulas that enable her or him to represent certain features of a given text in the form of an objective numerical value that both reveals something meaningful about the text at hand and facilitates direct, quantifiable comparison with other texts that have been subjected to the same analysis. By defining explicitly and with axiomatic precision the processes through which he or she has calculated each numerical value, the stylostatistician can, in theory, provide an 'index' pertaining to a given stylistic feature that is both entirely unambiguous and consistently reliable across a diverse body of texts. It is essential, however, that the stylostatistician should recognise that the explanatory scope of the results of her or his analysis is limited to an extent covariant with the degree to which her or his terms gain meaning from their natural language equivalents. In Coetzee's "Statistical Indices of 'Difficulty'", for instance, the quotation marks surrounding the word "Difficulty" communicate the fact that he is not in the final analysis interested in producing any clear definitions of "difficulty" per se: this term, as it is used in natural language, will ultimately remain at least partially obscure. Given this inevitable impediment, Coetzee's opening paragraph clarifies his conception of the role of stylostatistical analysis:

I take it as universally acknowledged that "difficult" in the proposition "A's style is difficult" is a complex word, and hence that the proposition in fact expresses a number of component conclusions, many of them quantitative in nature and therefore capable of being chiselled into numerical form. We may infer, indeed, that these quantifiable components take their origin in some quantitative, cumulative procedure, however loose, that we follow in our minds as we read, and hence that propositions about "difficulty" and perhaps other so-called qualities of style are most simply and logically formulated with their quantitative and nonquantitative components kept distinct. [Coetzee 1969a, 226]

The stylostatistician, then, limits his or her analysis to those "component" textual features that are "capable of being chiselled into numerical form", considering these as mere indicators of an overall "composite" quality – in this case "difficulty" – that, while it can never be fully accounted for without recourse to inherently subjective "Scrutiny criticism", the critic nevertheless wishes to assess to as great a degree of specificity as possible. By extension, we may deduce, any ordinary language term we might use to describe the style of a given text might conceivably be constructed at least in part from "quantifiable components" that correspond to an either literally or metaphorically quantitative or cumulative concept in which they in some likely subconscious form participate.

Within this methodological framework, "Statistical Indices of 'Difficulty'" has at its core two principal concerns: first, to elaborate upon and refine a stylostatistical index of textual "difficulty" proposed in 1952 by Wilhelm Fucks; and, second, to use this specific elaboration and refinement as a means of approaching more general and enduring issues within the field of stylostatistics. Using William Shakespeare's *Othello*, three works by John Galsworthy, and two by Aldous Huxley as object texts, Fucks proposed an index of "difficulty" based on mean word-length, with the syllable as the unit of measurement.^[1] From this initial affirmation, he constructed an algorithm that enabled him to "chisel" the raw evidence of the distribution of mono- and polysyllabic words throughout each text into a corresponding numerical value, or "trace". Once the "trace" has been evaluated, it is possible to rank the object texts along a scale that might for convenience be called the "Fucks Index": *Othello* was seen to produce the highest trace (107.65), followed by the works of Galsworthy (96.8, 94.06, and 91.07), with those of Huxley falling some distance behind (62.04 and 57.72).

But what, Coetzee asks, does the Fucks Index measure, exactly? And how does the trace it produces correspond with the natural language notion of “difficulty”?

[I]f we propose to describe the style of a text, an index must remain meaningless until we can specify precisely what it measures, i.e. with which phenomena in the text it varies systematically. If the value of the trace is high, as it is for Othello, what features of the language of Othello would this value enable us to predict without referring to the text? If its value is low, as it is for the present essay, what features of the essay are being reflected? Can we specify the features both simply and informatively? [Coetzee 1969a, 228]

The most obvious, and least useful, answer to these questions, Coetzee suggests, is that the “fullest statement of what the index measures is a mathematical restatement of the definition of the index” [Coetzee 1969a, 228]. In other words, what the formula tests is exactly the formula: any attempt to accord ordinary-language meanings to the mathematical terms that constitute the formula, however innocent such translations might seem, can only detract from its precision and increase the vagueness of the conclusions. Nevertheless, as Coetzee points out, this “represents the defeat of any attempt to distinguish between a quality in the text (‘difficulty’ or whatever) and a quantity which measures it”:

What we hope for is presumably a compromise: neither the extreme simplicity but extreme vagueness of words like “difficulty”, nor the tautology of the restatement, but a fairly short, fairly precise set of empiric features of the text, between the index and which there is a fairly steady correlation. [Coetzee 1969a, 228]

So, what Coetzee proposes here is that the index will be meaningful if and only if variation in the trace is consistently accompanied in the object text by a commensurate variation in certain other empirically observable features: if, for instance, the trace is higher in one text than another, then we should reasonably be able to expect that the first text will contain fewer “clusters” of words of more than three syllables – where the term “cluster” is sufficiently defined – and that this increase ought to be commensurate with the increase in the trace value.

Having established this basic goal of stylostatistics as exemplified by the Fucks Index, Coetzee demonstrates by counterexample that the specific formula Fucks uses to calculate his trace will not always yield results that stand up to this test: it is possible, he shows, to deliberately construct texts that have either a high trace and a relatively high number of clusters, or a low trace and a relatively low number of clusters. Moreover, the trace tends to accord disproportionate weight to the values generated by words of higher syllabic length: in other words, a text that has a high trace in respect of its constituent mono-, di-, and trisyllabic words may find its overall trace value affected in an exaggerated way by the occurrence of a couple of highly syllabic words. The second of these two problems he considers as one of categorisation: “if we are prepared to accept a word of three syllables as ‘difficult’ for our descriptive purposes”, he proposes, then we can accept as valid the revised and more “efficient” formula he constructs in the essay so as to negate the biases caused by the inclusion in the analysis of those relatively rare words of four or more syllables. The problem with this solution is again one of unavoidable compromise: the more we impose our natural-language definitions on the axioms of number, the more reliable our results will appear, but the more “connotative freight” the definitions in our conclusions will carry. For this moment at least, Coetzee is prepared to leave this as a methodological dilemma for the stylostatistician: the more troubling implication, however, is that the precision of our quantitative evaluations as such necessarily varies in inverse proportion with their qualitative meaningfulness.

10

In “Statistical Indices of ‘Difficulty’”, Coetzee’s major reservation as to the efficacy of stylostatistical analysis is manifest by the ease with which one might construct a subject text designed specifically to violate the propositions according to which a given index might be said to be “meaningful”. In the case of the Fucks Index, he attributes the potential for counterexamples to the fact that stylostatistical analysis, like all statistical analyses of natural phenomena, operates not within a fixed, deterministic space, but rather a probability space within which there will inevitably exist superficially troublesome outliers.^[2]

11

language *en masse* exhibits many of the characteristics of chance phenomena, and [...] the inverse correlation between the value of the trace and the degree of presence of polysyllables and

polysyllable clusters, while not invariable, has a high probability associated with it. [Coetzee 1969a, 231]

This leaves us in a quandary, for, while it “may seem odd at first sight that something which is so largely a matter of design as a literary text should exhibit randomness” [Coetzee 1969a, 231], and, while this “kind of breakdown in the trustworthiness of the Fucks trace [...] will in practice occur very seldom” [Coetzee 1969a, 231], it is the nature of probabilistic distributions to throw up anomalies such as these without deliberate design. So, “[w]e cannot, unfortunately, claim that it will never occur, for then we would have to show that texts like [the counterexamples] are not ‘natural’, and would inevitably be reduced to talk about intention” [Coetzee 1969a, 231]. The thorny problem of intention, and to “texts [that] are constructed with an eye to the code rather than to the message” [Coetzee 1969a, 231], had not only been a driving force behind Coetzee's work as a computer poet, but also occupies a significant position in his contemporaneous work on Beckett: by the time his thesis was complete, what might have initially appeared as mere irritations in the practice of a few esoteric critics operating at the limits of stylistic analysis would in fact shed light on problems with a far greater resonance.

3 “The English Fiction of Samuel Beckett” (1969)

While “Statistical Indices of ‘Difficulty’” might seem rather tentative in its criticism of the orthodoxy of stylostistics as an academic discipline, Coetzee's doctoral thesis, “The English Fiction of Samuel Beckett”, is noticeably more sceptical from the outset. Having by now immersed himself in the discourse of stylostistical analysis for four years, and having examined the relationship between language, numerical methods, aesthetics, and epistemology from the position of both composer and critic, Coetzee now seemed to be reaching something approaching a resolution to the more profound aspects of his doubts. Crucially, the thesis begins with a reference to a quotation from the mathematician and philosopher David Hilbert:

12

In the course of a fusillade against what he calls “the revolt against reason” in present-day humanistic studies, Joshua Whatmough quotes a pronouncement of David Hilbert's from 1918: “Everything that can be an object of scientific thought at all, as soon as it is ripe for the formation of a theory, falls into the lap of the axiomatic method and thereby indirectly of mathematics.” [Coetzee 1969b, 1]

Literary criticism *qua* Whatmough and Hilbert, then, can only be deemed to be “scientific” if its commitment to the demands of the axiomatic method is absolute, such that its terms and procedures are explicitly and rigidly defined and delineated, immutable, and therefore repeatable and comparable across each and every subject that falls under its critical gaze. Coetzee, however, remained unconvinced by the position represented here by Whatmough and Hilbert; the four years he had spent researching and writing his thesis gave him the critical tools required to critique in minute detail the use of axiomatic methods within literary criticism. It was on this basis that he was consequently able to recognise the extent of the complexities involved in any attempt to disentangle from their nested assumptions a workable distinction between natural language and the language of mathematics.

In service of his ambitious and interdisciplinary thesis, then, Coetzee establishes his critical methodology as operating within a theoretical space delimited by “two poles” of thought he discerns within the existing discourse. The first of these “poles” he attributes to Bernard Bloch, according to whose “classic definition [...], style had indeed become an object of scientific thought, was ripe for the formation of a theory, and was falling, not at all indirectly, into the lap of mathematics” [Coetzee 1969b, 1]. Coetzee exemplifies this definition with a quotation from Bloch in which the American linguist renders the “style” of a text as reducible to “the message carried by the frequency distributions and the transitional probabilities of its linguistic features, especially as they differ from those of the same features in the language as a whole” [Coetzee 1969b, 1] [Bloch 1953, 42]. For Bloch, then, the moment that we conceptualise “style” as an “object of scientific thought” we automatically circumscribe it in terms of a propositional content – its “message” – that is fully coextensive, without remainder, with the quantitative description generated by stylostistical analysis.

13

Coetzee's selection of Bloch as the principal representative of his first “pole” situates the terms of debate in an unmistakably pointed fashion: working within the tradition of his mentor at Yale, Leonard Bloomfield, Bloch was among

14

those most responsible for the development of the discipline of structuralist linguistics. By the time Coetzee came to work on “The English Fiction of Samuel Beckett”, this tradition had effectively been superseded in terms of academic prestige by the Standard Theory of generative grammar first outlined by Noam Chomsky in the late 1950s. Indeed, in his notes to a 1969 course on “English and Linguistics”, Coetzee outlines a brief “History of Syntax,” in which he characterises “American Structuralism” by way of a single quotation from Bloomfield, the implicit pessimism of which goes some way to indicating the beleaguered state in which this school of linguistics was considered at the time of Coetzee's thesis: “The statement of meanings is [...] the weak point in language-study, and will remain so until human knowledge advances very far beyond its present state” [Coetzee 1969c]. In the sense that Bloomfield's approach to linguistics was essentially positivist, determinist, and behaviourist – perhaps exemplified at its most extreme in his Laplacian contention that, if only one had sufficient data, then one could predict all future events, including all future speech-acts – it is perhaps unsurprising that the Coetzee of 1969 was interested in but ultimately sceptical of the truth-claims of the discipline Bloomfield's work inaugurated.

While Coetzee's selection of Bloch as the representative of this “pole” establishes one horn of the dialectic as a particular strain of structuralist linguistics, then, and hence indicates his entry-point into the discourse to which the thesis ostensibly contributes, his decision to introduce this perspective in the context of two thinkers concerned with, in Hilbert's case, the first principles of mathematical philosophy and, in Whatmough's, “the revolt against reason” in present-day humanistic studies, indicates from the outset Coetzee's recognition of the potentially widespread ramifications of his own conclusions. The terms through which he expounds his counterpoint to Bloch's position, moreover, delineate the scope of the thesis in accordance not merely within the boundaries of his own work as a stylostatistician, but also in such a way as to facilitate an extended interrogation into certain epistemological issues towards which his prior work had been at most tangentially oriented. He defines the second “pole” by reference to the views of the principal subject of the thesis, Samuel Beckett, whom he characterises as reacting “strongly against any simplification of language [...] and indeed against any abstraction from words as counters in a calculus of thought to words as counters in the less flexible calculus of language” [Coetzee 1969b, 3].

At its most fundamental level, then, Coetzee's thesis originates from a compulsion to explore the sense in which Beckett's rejection of the type of abstraction routinely performed by structuralist linguists such as Bloch and Bloomfield constitutes a further denunciation of the apparent ease with which certain terms and categories from the discourses of statistics and probability theory had begun to migrate into humanistic studies. The dextrous manner in which he introduces his negotiation of these terms merits close examination:

Between the conceptions of style held by Bloch and implied by Beckett there are no doubt similarities: Beckett's “writing without style” could be interpreted as writing with the statistical features of the language as a whole, whatever that may be. But there is a deeper cleavage which gives the two viewpoints a polar and antithetic relation. Underlying Bloch's definition is the idea of the text as a collection of sets of linguistic features (phonemes, morphemes, words, etc.) which can be treated like members of statistical populations; and a statistical population is only a metaphor for a set of points in probabilistic space. To Bloch, a word can be conveniently reduced, for the purposes of study, to a dimensionless and immaterial point. For Beckett, on the other hand, the “terribly arbitrary materiality of the word's surface”^[3] is, we infer, at least in 1937, a burden. [Coetzee 1969b, 3]

Even as he establishes the nominal focus of “The English Fiction of Samuel Beckett” as literary “style”, then, Coetzee indicates that his primary objective is in fact to map the process by which structuralist linguists such as Bloch and Bloomfield sought to transform quality into quantity, texts into statistical populations, and words into dimensionless, immaterial points in probabilistic space. Viewing the resulting map through the lens of Beckett's fiction, he enabled himself to touch upon a variety of issues far beyond the scope of either the traditional form of literary criticism he had undertaken in his MA thesis on “The Work of Ford Madox Ford”, or the by now stabilising orthodoxy of stylistic analysis as exemplified by the work of Wilhelm Fucks. At issue in the thesis were not only questions of the stylistic qualities of Beckett's work and, by extension, literary texts in general; not merely metatheoretical concerns regarding the discourses and methodologies of stylistics and stylostatistics; and not simply the same problems of meaning construction in

mathematics that had exercised Hilbert, Whitehead, and their followers: while “The English Fiction of Samuel Beckett” has ramifications for each of these complex fields of study in isolation, its unique value is to be found in its subtle and delicately-handled assimilation of these issues into a more profound and further-reaching philosophical space that drew both its assumptions and its responses to those assumptions from the curious matrix of ideas through which each of these apparently disparate discourses passes during the various stages of their construction.

3.1 Traditional Versus Stylostatistical Criticism

On the basis of his introduction, Coetzee suggests that the “significance of Beckett’s attack on ‘style’ should now be becoming clearer” [Coetzee 1969b, 4]; though its superficial target might well have been the nature of literary language – and especially the nature of any disparities that might obtain between French and English as vehicles of literary expression – Coetzee notes that “Beckett’s description of the ‘materiality of the word’s surface’ pictures language as a wall between objects and their percipients” [Coetzee 1969b, 4]. In the context of his own continuing investigation into the relative validity of attempting to build this wall from either linguistic or mathematical “bricks”, or a combination of the two, it is perhaps useful to note that Coetzee sees the “position on style” he adopts in the thesis as being “plainly closer to Beckett’s than to Bloch’s” [Coetzee 1969b, 6]. He locates this position more specifically as one that remains equally unconvinced by the methodologies of, on the one hand, critics whose use of established literary-critical language defers to imprecisely defined “connotative freight” and, on the other, the prevailing orthodoxies of contemporary stylistics. In the first case, for example, he characterises Hugh Kenner’s principal approach as constituting an attempt “to catch the essence of Beckett’s style in a metaphorical way”:

Thus, for example, of the “unique translucent enumerating style” of *Watt* he writes, “It is an austere prose, not narcissistic, nor baroque. It is not opulent. It moves with the great aim of some computation, doing a thousand things but only necessary ones.” [Coetzee 1969b, 9–10]

Similarly, he explains, Ludovic Janvier’s *Pour Samuel Beckett* (1966) “has some perceptive pages on the ‘dizziness’ (vertige) induced in the reader by his mathematical comedy” [Coetzee 1969b, 9]. Pointing out that both Kenner and Janvier rely in these instances upon “a tradition of literary criticism in which terms like ‘austere’ have an agreed meaning, and in which insight into the nature of a style is a partly intuitive act” [Coetzee 1969b, 10], Coetzee recognises that though it may neither define its terms with the specificity demanded by the stylostatistician nor proceed from “statements which can be verified by quantitative analysis” [Coetzee 1969b, 10], this type of traditional criticism nevertheless evades the “general positivism” [Coetzee 1969b, 17] to which he concludes stylistics – particularly in the structuralist tradition represented by Bloomfield and Bloch – had by then become excessively beholden. Stylistic analysis, he continues, is often predicated on ultimately arbitrary processes of division that fail to take sufficient account of the “artistic whole” [Coetzee 1969b, 17] and therefore systematically neglect the fundamental truth that the “experience of a work of literature is not necessarily linear in time,” and instead tacitly defer to an “analogy of reader to decoding device” that he considers “misleading” [Coetzee 1969b, 18]. Coetzee consequently devotes much of “The English Fiction of Samuel Beckett” to a systematic reconstitution of certain arcane, technical aspects of stylostatistical practice that need not be rehearsed here. More significant for the purposes of the present study are those instances in the course of the thesis in which his more broadly methodological allegiances begin to make themselves known; to this end, Coetzee’s analysis of Beckett’s *Watt* is highly instructive.

It is probably not too controversial to state baldly that *Watt*, begun in February 1941 and eventually published, following extensive revisions, in 1953, is generally considered to be among Beckett’s most “mathematical” novels. In this sense, Coetzee’s comment that the novel’s eponymous protagonist is “like Leibniz’s automaton with a spark of life” [Coetzee 1969b, 31] and, “[s]tanding Bergson on his head, [...] something living encrusted on the mechanical” [Coetzee 1969b, 32] represent additions to an existing consensus rather than anything more revolutionary. To clarify Coetzee’s precise conception of Watt’s condition, though, one might first note that he considers it to be “characteristic of Watt that he believes that an empirical question can be solved by logical analysis”:

No empirical data are introduced into his chains of speculation. The multiplication of these chains depends on a maneuver in four stages: statement of a question, proposal of a hypothesis,

breakdown of the hypothesis into components, and analysis of the implications of the hypothesis and its components. [...] The third stage typically breaks the chain into two or more branches. The only qualification Watt demands of a hypothesis is that it answer the question: his criterion is one of logic rather than of simplicity. [Coetzee 1969b, 81]

On the one hand, then, Watt's consciousness represents the very model of the supposedly perfectly closed logical system of mathematics; on the other, his access to the sensory world beyond this closed system introduces experiential data that consistently evade its processes of assimilation and hence are habitually disregarded. This disregard, Coetzee continues, is in fact a "disregard for simplicity" and is the

foundation of [Watt's] logical comedy, for simplicity is the only criterion that can put a stop to an endless proliferation of logical speculation. In *Watt* we regularly, with a sinking feeling, find ourselves at the beginning of infinite series. [Coetzee 1969b, 81]

Such is the finite nature of a text – and, indeed, a consciousness – however, that, whatever Watt's predilections, "the infinite series which automatically spring up must somehow be terminated" [Coetzee 1969b, 81], with inevitably absurd consequences. One such example, Coetzee reminds us, "terminates in the solipsism that is one of Watt's answers to the infinities of logic: fish that need to rise and fall exist because my naming of them brings them into existence" [Coetzee 1969b, 81]. Without the incursion of external experiential data, Watt's case informs us, a closed system of logic shall produce no meaning other than that embedded in its logical categories; without a logical system predicated on experience to guide its selections, however, the process through which such experiential data are collected is as likely to cause the regression to terminate in absurdity as in rationality; and without a referential framework in which to compare our findings, moreover, we must conclude that we shall inevitably have no means of telling the difference.

Watt's consciousness, then, is analogous to the type of deterministic formal axiomatic system of which the modern computer is perhaps the most familiar model. Built from a series of axioms or rules for behaviour, the system is set into motion by the intrusion of an essentially arbitrary piece of empirical data, which consequently acts as its originary affirmation. As Coetzee explains, with every passing instance in which Watt initiates an exhaustive combinatorial analysis in response to a particular set of circumstances, the reader gains a cumulative sense of the inextricability of his condition; the "attempt to understand the nature of the simplest sensory perceptions" [Coetzee 1969b, 35], he elaborates, leads to a paradox born of the complex, self-referential nature of the concept of infinity.

20

Without the means to make qualitative value distinctions beyond the basic logical tools with which he is endowed, then, Watt is radically unable to determine the limit-point at which his analysis of each given set of circumstances might be said to approximate truth to an extent sufficient to justify action. Indeed, the very idea of cause and effect becomes more and more undermined as Watt's experience becomes progressively "inverted":

21

The explosion of logic, epistemology, and ontology takes Watt into another zone (the asylum) in which he lives a progressively inverse life. Decline and inversion are reflected in Watt's language, as reported by the narrator Sam. Decline and inversion constitute what I call the shape of the telos. What is still lacking is the causal element. For certain reasons a certain kind of man experiences a call to a certain kind of situation, and the result is decline and inversion: we see the results but not the causes, unless we take the step of calling Watt's whole universe absurd. [Coetzee 1969b, 35–36]

It is in a similar context in the essay "Samuel Beckett and the Temptations of Style" [Coetzee 1973a] – the last to be published of the three journal articles he adapted from "The English Fiction of Samuel Beckett" – that Coetzee introduces Richard Dedekind's hypothesis to the effect that "[i]f we can justify an initial segmentation of a set into classes X and not-X [...], the whole structure of mathematics will follow as a gigantic footnote" [Coetzee 1973a, 43]. Given that this constitutes one among very few additions he made in the process of distilling sections from his thesis into forms suitable for publication as journal articles, one might suggest that this aspect of his study had taken on a greater significance in his thinking during the four years between the completion of the thesis and the publication of the article. In Coetzee's characterisation of them, both Dedekind and Beckett are "mathematician enough to appreciate"

that, on the basis of merely one “single sure affirmation,” a “whole contingent world [...] can, with a little patience, a little diligence, be deduced” [Coetzee 1973a, 43].

3.2 Stylostatistics as a Constructivist Discipline

Through an analogy with the analysis of *Watt* that he undertakes in “The English Fiction of Samuel Beckett”, then, one can clearly observe Coetzee’s recognition that even a representational framework so seemingly free from interference from the world of unnegated affirmation as stylostatistical analysis serves as a model for rejecting the “reality” of truths developed within closed meaning systems, in favour of picturing them as merely constructivist:

22

On the other hand, the smallest amplifications of meaning, particularly those which were probably not under the conscious control of the author - - for example, the frequencies of the words in the text - - show, when quantified, what looks suspiciously like system, i.e. they act like well-behaved mathematical functions. Turning the syllogism upside down, we infer that well-behaved mathematical functions defined on the quantified components of the text define components that belong to the smaller amplifications of meaning. [Coetzee 1969b, 40]

In other words, certain initial – and often subconscious – affirmations are ultimately responsible for determining the nature of both the component features we discern as constitutive of a given text and the types of mathematical function that appear to describe or even govern their behaviour. Once a reader makes these affirmations, both the nominal and the functional aspects of the text become ontologically linked in a manner that has astonishingly little to do with the ontological status of the text prior to those affirmations. Extrapolating this observation to the use of natural language alongside quantitative evaluations, one can see the potential for such frameworks to engender obfuscation rather than the desired objective clarity. In Coetzee’s example, for instance, one might question the effect of introducing an index for a term such as “elevation in diction” as a descriptor for certain textual features, other than to provide a misleading “connotative freight”:

By the time sufficiently many literary works have been described in terms of the same measures, the measures themselves may come to have associative values with different texts. We may find, for example, that a high noun-to-adjective ratio is common to Pliny and Thomas à Kempis, a low ratio to Virgil and Tacitus. The ratio may then become associated with a quality we may call elevation in diction. But ultimately elevation will have to be defined in terms of the noun-to-adjective ratio and other measures. There is no escape from the absolute measure of quantification here. [Coetzee 1969b, 44]

More troublingly, perhaps, it is not just in descriptive terms of this nature that we encounter such a problem: the origins of even the most apparently basic linguistic terminology are equally as precarious:

23

a little computation shows us that, whatever definitions of noun and verb we adopt, their effect on the noun-to-verb ratio, while greater than the effect introduced by the uncertainties created by implicit nouns and verbs, is considerably less than the effect that could be introduced by uncertainties in the classifications “noun” and “verb” [...]. It does imply that the potential for disastrous error is high when we depend on figures not derived from identical and therefore exhaustive definitions of noun and verb for the purpose of comparing the “nominalism” of different texts and authors. [Coetzee 1969b, 46]

Coetzee draws attention here, then, to the fatal circularity of any analysis that fails from the outset to recognise the uncertainties inherent in categorisations even as seemingly fundamental as “noun” or “verb”. Generally, the comparison of the works of any two authors requires strict definition of the terms of that comparison: the result of this comparison, however, is destined ultimately to become primarily a comment on the act of definition that has taken place, rather than on any inherent quality of the texts or authors themselves. To Coetzee’s mind, the only conceivable solution to this problem that might help to “square intuition with mathematics” [Coetzee 1969b, 49] would be to refine the precision of our terminological definitions: “our only recourse”, he explains, “is therefore to assign different numerical weights to

different nouns and verbs, based on such criteria as their rarity, their degree of compoundness, etc” [Coetzee 1969b, 49]. Just like Watt, however, we soon find ourselves at the beginning of an infinite regression:

But now we have opened the floodgates. For we are not concerned, for example, with absolute rarity (whatever that is) but with rarity in a context. The position becomes untenable, for no generalization is possible, and the reason for computing to ratio in the first place is to have a measure of nominalism in the text, i.e. to have a generalization about a certain aggregate of particulars. [Coetzee 1969b, 49]

Ultimately, we are left to conclude that the use of the same index on two separate occasions is logically counter-intuitive: whereas two words could previously become “equal by being used with the same frequency”, Coetzee explains, “the notion of equality in meaning is tenuous” [Coetzee 1969b, 50]. The consequences for stylostatistics, as the following quotation suggests, are effectively fatal:

We are faced, then, with a story in which statistical analysis of the distribution of vocabulary, classification of the less neutral diction, and analysis *en masse* of sentence structure, seem at best only to confirm our understanding of the structure of the work and at worst to remain trapped in their own terminology. [Coetzee 1969b, 54]

As such, then, these “amplifications of meaning” – affirmations for which the critic is solely responsible, ranging from the grouping of verbs under some grammatical concept such as transitivity (“hold” with “throw” and “reveal” for example) to the less robustly delineable association of nouns on a semantic basis (“building” with “edifice” and “construction”) – are necessarily echoed in the “function” that analysis of them reveals. In other words, in stylostatistics, as in mathematics generally, our observable and delineable data and our modes of observation and analysis are irrevocably bound up with one another and, in a sense, offer little more than tautology.

Perhaps the most critical of all Coetzee's observations in “The English Fiction of Samuel Beckett”, though, corresponds to a brief, aborted train of thought explored in an endnote that, though it is not followed through to its completion in the thesis, ramifies throughout his contemporaneous work. How is it, he wonders, that certain linguistic phenomena, such as the “inverse relation [...] between rank and frequency” [Coetzee 1969b, 240] of lexical items in an object text are “describable in mathematically simple terms?” [Coetzee 1969b, 240]

Is it coincidence, or is it one instance of isomorphism between the structure of language and the structure of mathematics? In the first case the Zipf-Mandelbrot law^[4] is a useful descriptive fact, loosely a “law.” In the second case it is indeed tautologous, but the consequences are too immense to bear contemplation. [Coetzee 1969b, 240–241]

4 Review of Wilhelm Fucks’ *Nach allen Regeln der Kunst* (1971)

Coetzee, then, was unwilling to contemplate in the context of his doctoral thesis the “immense” consequences attendant on the possibility that the structure of natural language and the structure of mathematics might be isomorphic. However, on the basis of Coetzee's second published discussion of the work of Wilhelm Fucks – a 1971 review of the German linguist's 1968 work, *Nach allen Regeln der Kunst* – one gets a clear sense of his attitude towards the philosophical machinery of statistical analysis such as it emerged from his immersion in that discourse during the late 1960s. Nowhere is his ambivalence towards the wider potential consequences of a positivism founded on the migration of mathematical structures into ostensibly non-mathematical concerns better encapsulated than in his review's vivid opening description, in which he depicts Fucks as either a far-sighted visionary or a reductivist brute: “Depending on how you view him,” he begins,

Wilhelm Fucks is a polymath of refreshing synoptic vision or another of those muscle-men of statistics (Yule, Herdan et al.) to whom a ward of kwashiorkor^[5] victims or a page of print is first of all a set of quantifiable phenomena and only secondarily people or literature. [Coetzee 1971, 92]

Given that Coetzee was already by this stage developing a dramatisation of an extremely similar process in *Dusklands*, it seems especially significant that he chose in this review to extrapolate from Fucks' seemingly harmless literary exercises to an apparently genuine fear that the possible emergence of a "speakeable formalized language" [Coetzee 1971, 94], developed "as a universal language for the technocratic elite" [Coetzee 1971, 94], might "tie succeeding generations into a twentieth century positivist mythology more tightly than natural languages tie us into the mythologies of the past" [Coetzee 1971, 94].

The essential purpose of *Nach allen Regeln der Kunst*, Coetzee explains, is to reiterate "the theme that the artist, like any other organism, exhibits regularities of behaviour, which can be exposed by statistical analysis" [Coetzee 1971, 92], thereby revealing "the elegantly formulable mathematical distributions underlying such phenomena as the lengths of sentences in a text and the pitches of note-pairs in a concerto" [Coetzee 1971, 92]. Given the philosophical reservations upon which he had predicated the ambivalence at the heart of his doctoral thesis, then, it comes as no surprise to find that Coetzee's response to such a project as Fucks's is at best cautious. Inasmuch as it in principle welcomes the advent of a general introduction to a discourse in which he was himself at this point still relatively heavily intellectually invested, Coetzee's review initially praises Fucks for his capacity to "explain so patiently and with such lavish visual aids his basic procedures" [Coetzee 1971, 92] and hence render "seductive [...] a field which many think of as rather arid" [Coetzee 1971, 92]. Equally, while the book is "emphatically not a handbook" [Coetzee 1971, 92], and does not constitute "a compendium of investigations into intrinsically interesting stylistic topics" [Coetzee 1971, 92], it nevertheless contains examples of such investigations that may indeed "seduce" the non-specialist reader, not least by indicating the value of stylostatistical procedures beyond the academy; in a chapter pointedly entitled "Literary Criminology", for instance, Fucks demonstrates "inter alia that the Gospel of St. John and Apocalypse are probably not from the same hand" [Coetzee 1971, 92], thereby intimating not only the literary value of such practice, but also alluding to its potentially crucial application in the courtroom.

Coetzee's praise is tempered, however, by his enduring conviction that the means through which stylostatistical analysis enacts its negotiation between the qualitative and the quantitative is intrinsically flawed. While any "reasonable man must be convinced that regularities of all kinds, regularities of stress, of syntax, of word choice, and so forth, run through literary compositions, [and] that the set of these patterns comprises a great deal of what we call style" [Coetzee 1971, 93], it nevertheless remains the case that the "overwhelming proportion of [stylostatistical indices] either have no critical application or represent quantitative restatements of qualitative propositions ('A's verse is more varied than B's')" [Coetzee 1971, 93]. As he had earlier expressed at greater length in "The English Fiction of Samuel Beckett", Coetzee points out that this problem is predicated in the first instance on the fact that the "kind of datum that the statisticians, Fucks included, feel at home with is extremely elementary: word length, sentence length, ictus, grammatical class, depth of subordination" [Coetzee 1971, 93], and that he or she will only escape rather prosaic conclusions by producing "a whole new typology of structures" [Coetzee 1971, 93] and programming his or her computer with procedures that are able to "classify and count in a much more complex way" [Coetzee 1971, 94] than those that the discourse of the time seemed happy to accept.

While the review is essentially concerned with commenting upon Fucks's contribution to the discipline of stylistics, then, Coetzee's more considered conclusions refer not strictly to issues of literary criticism, but instead to an epistemological model that stylostatistical analysis covertly advocates. By suggesting that "it would be fairest to take this book as a work of propaganda, a work intended to convince the uninitiated first that there are regularities they had never suspected underlying behaviour" [Coetzee 1971, 92–93], Coetzee draws attention to the surreptitiously political nature of any attempt to specify and formalise these "regularities". Fucks, he explains, "has a distaste for the 'swarms of associations and emotions' that accompany reading and for the 'whole layers of primitive taboos and antiquated mythology' concealed in natural languages" [Coetzee 1971, 94] and hence prefers to reduce linguistic behaviour to those "formal phenomena of the printed text" [Coetzee 1971, 94] that happen, by virtue of their accordance with conveniently quantifiable structures, to be amenable to inclusion within an "objective descriptive aesthetics" [Coetzee 1971, 94] that is unified, comprehensive, and hence logically closed.

By the end of the review, then, one is left with the clear message that, whereas Fucks' "propaganda" is aimed towards assuring his readers that "a literary science of exact numerical description is a good thing" [Coetzee 1971, 92–93],

Coetzee is concerned here, as elsewhere throughout his work as a stylostatistician, to highlight the ramifications of believing, as Fucks does, in the merits of an “objective analysis” [Coetzee 1971, 94], even where this necessitates that we “omit a great deal” [Coetzee 1971, 94] in our description of the phenomenon under observation. It is in the final paragraph of the piece, indeed, that Coetzee briefly discusses Fucks' contemplation of the potential development of “a ‘speakeable formalized language’ as a universal language for the technocratic elite” [Coetzee 1971, 94]. Though Fucks is aware of “Whorf's thesis that languages have built-in epistemological biases” [Coetzee 1971, 94], Coetzee notes in pointed fashion the book's failure to consider the possible ramifications for a future society in which the language developed by the “linguistic engineers” has enshrined in its users a positivist mythology with an even greater delimiting power than ‘natural language’ has had on the cultures of the past and present.

5 “Samuel Beckett's *Lessness*: An Exercise in Decomposition” (1973)

In some regards, Coetzee's route from the review of *Nach allen Regeln der Kunst* to *Dusklands* is not difficult to retrace. In the most explicit thematic sense, for instance, the essay prepares the context for the rationale that the two protagonists of that novel share: a relentless positivist rationality, designed to locate and exploit regularities underlying the thought and behaviour of others. In the final piece of stylostatistical work he was to publish, however, there emerge other, more subtle connections in his contemporaneous thought. Appearing in its English version in 1970 – having originally been published in French, as *Sans*, in 1969 – the subject text of Coetzee's “Exercise in Decomposition” displays, in his words, “features not often encountered in connected discourse” [Coetzee 1973b, 195]. The “most notable” of these, Coetzee elaborates, is its “finiteness”: in the sense that the text of *Lessness* is divided into two halves, each consisting of the same sixty sentences, only in a different order, the novella's linguistic resources are limited to just 166 lexical items, a “finite subset” of the natural language, English, from whose theoretically infinite resources it is ultimately drawn. “It is this fact,” Coetzee states, “which suggests a mathematical approach to the text, an approach not only via the mathematics of indeterminacy, namely probability theory [...] but also via combinatorial mathematics” [Coetzee 1973b, 195].

31

From this starting point, Coetzee first establishes by means of Spearman's rank correlation coefficient^[6] that one cannot dismiss “with any acceptable degree of certainty” [Coetzee 1973b, 195] the hypothesis that the re-ordering of the sentences is effectively random. From here, he next determines that the “unit of combination in *Lessness* is not the word but the phrase of one or more words” [Coetzee 1973b, 196], and that, by using a specified algorithm, we are enabled to “obtain an unambiguous segmentation of the text into 106 different phrases varying in length from 1 to 12 words and occurring, on an average, 5.7 times each” [Coetzee 1973b, 197]. He goes on to demonstrate by means of methods drawn from statistics and probability theory that there are “no closed subsets of phrases” [Coetzee 1973b, 197], that “there is no statistical reason for rejecting the hypothesis that phrases are distributed randomly over paragraphs” [Coetzee 1973b, 197], and that the occurrence of any “clusters” of phrases that do happen to form throughout the text “do not fall into any [...] elementary patterns” [Coetzee 1973b, 197]. *Lessness*, in short, exhibits randomness at practically every conceivable textual level, actively evading capture within mathematically expressible system at every turn.

32

In the sense that it generates a relatively mechanistic and conventional form of analysis, and as such is fairly typical of the discourse of computer-assisted literary criticism as it existed in 1973, “Samuel Beckett's *Lessness*: An Exercise in Decomposition” might to this point seem at best unexceptional, and at worst academically indulgent. Where the essay offers an especial insight into Coetzee's intellectual development, however, is in its philosophical interpretation of Beckett's revolt against system. Coetzee begins this interpretation, then, by establishing that “Beckett's most recent fictions, the *Residua*, of which *Lessness* is one, portray an existence whose conditions are stripped further and further down” [Coetzee 1973b, 198]. This “stripping-down” is constituted of three “levels,” which Coetzee characterises as follows:

33

The first level of this consciousness contains a past womb-existence, a set of figments. The second level contains the figments of the new fiction *Lessness* that the consciousness now inhabits: ruin, sand, body, etc. The third level contains only the pair dawn-dusk, each of which eventually cancels both the other and the figments for which the other is responsible. [Coetzee 1973b, 198]

In “Samuel Beckett and the Temptations of Style” [Coetzee 1973a], moreover, he explains that in *Lessness* “an infinite series of nested consciousnesses, each dismissing the figments of its predecessor, is presented in the paradigm of a two-component switching mechanism,” each of which ultimately “annihilates the figments of the other” [Coetzee 1973a, 46]. As a writer whose novels often play with multi-level stagings of competing editorial effacement – most prominently, perhaps, in *Dusklands*, *In the Heart of the Country*, *Foe*, and *Slow Man* – Coetzee’s location of this process in *Lessness* presents one point of methodological kinship. Most interesting, however, is his attribution to Beckett of a particular conception of consciousness, its representation in an apparently linear text, and the value of using a mathematical analysis to determine a means for exploring issues without succumbing to commitment or belief in any component met along the way. In his essay, Coetzee demonstrates that “there are no determinate principles of ordering among phrases, sentences or paragraphs, yet that all are interdependent and connected” and that, consequently, there is “no principle of hierarchy or priority among the components of the work” [Coetzee 1973b, 198]. The upshot of this lack of “hierarchy or priority” is that any of the millions of alternative re-orderings Beckett may have chosen to publish would be as “valid as fiction” [Coetzee 1973b, 198] as that which was, in fact, published. Similarly, the final, linear ordering of the fragments is less expressive of the fundamental meaning of the text than is the process through which the text came to be:

Since any fragment can combine with any other fragment, and since the 106 phrasal components are not only formal elements but also pretty irreducible elements of meaning, composition is a combinatorial game played with creations of what I have called the second level of the imagining consciousness – a level whose creations are dismissed as figments – and the upshot of the game is nothing more than what Sam, in *Watt*, called “a pillow of old words.” [Coetzee 1973b, 198]

Ultimately, then, Coetzee proposes that one ought not to take too seriously any cumulative effects resultant from the essentially arbitrary route taken by the consciousness enacted through Beckett’s fiction, but rather to the ephemeral, non-linear motions through which it passes within the working-out of its finite process:

The residue of the fiction is not then the final disposition of the fragments but the motions of the consciousness that disposes them according to the rules we have traced, and no doubt others we have failed to trace. [Coetzee 1973b, 198]

The “subject of *Lessness*”, he ultimately concludes, “is the plight of consciousness in a void, compelled to reflect on itself, capable of doing so only by splitting itself and recombining the fragments in wholes which are never greater than the sums of their parts” [Coetzee 1973b, 198]. Reflecting back upon Coetzee’s own novels, one may note that it is far from coincidental that many of them – *Dusklands*, *In the Heart of the Country*, *Foe*, *Disgrace*, *Elizabeth Costello*, and *Diary of a Bad Year* in particular – are not unlike *Lessness*, in the sense that they represent the motions of a consciousness through an apparently disordered maze of assertions, appearing to enact a cumulative process as the consciousness experiences, affirms, and effaces various propositions, often paired in binary oppositions, before seemingly arriving at fixed conclusions by the novel’s end.

6 Conclusions

While the force of Coetzee’s conclusions in relation to the discipline as it stands today remains very much up for debate, his work during the period under observation here offers a unique perspective not only on the early years of the field of digital humanities, but also on the intellectual development of one of the most significant novelists of the late twentieth and early twenty-first centuries: one rarely emerges from the revolutionary battles of one’s youth unmarked, and it might be said that in order to locate those marks in the war stories of a veteran, one need not only know that the war happened, but also the detail of each particular battle in which the storyteller fought. As such, the consequences of the present paper are threefold: firstly, critics of Coetzee’s writing ought to approach the thematisation of quantification – particularly in such works as *Dusklands* (1974), *In the Heart of the Country* (1977), *Waiting for the Barbarians* (1980), *Diary of a Bad Year* (2007), and *The Childhood of Jesus* (2013) — with a greater sensitivity to the questions raised here; secondly, the contemporary manifestations of those areas of digital humanities with which Coetzee’s work intersects might now need to reflect upon how to negotiate the paradoxes associated with quantification in meaning

construction, such as he has diagnosed them; and finally, we must all now recognise the value of the contribution made by this most enquiring and incisive of literary minds to the field of digital humanities as a whole.

Notes

[1] Fucks's decision to use the syllable as a measure of "difficulty" is self-evidently problematic. As Coetzee points out, "the phenomenon in which we really ought to be interested is not the syllable (as Fucks assumes) but the morpheme, since we can give a more precise meaning to the definition of a word of many morphemes as 'difficult' than a word of many syllables (consider 'Oopsidaisy')" [Coetzee 1969a, 232]. Here and elsewhere, one should note, while Coetzee criticises the detail of Fucks's methodology, he registers no objection to the underlying principles of the venture.

[2] A "space" is a structured set of points with fixed definitions regarding the behaviour of the space and the relationship between the points; perhaps the most familiar example is the Euclidean plane. A "probability space" is a finite space with an associated probability measure that assigns a value between 0 and 1 to the space as a whole. An 'outlier' is an observation that is considered not to conform to the general pattern of a given data set.

[3] Coetzee's quotation of the "terribly arbitrary materiality of the word's surface" is a translation from Beckett's original German of the phrase "fürchtliche willkürliche Materialität der Wortfläche". See "German Letter of 1937," in [Beckett 1984, 53].

[4] As Coetzee defines it in his essay on Beckett's *Lessness*, the Zipf-Mandelbrot Law describes the phenomenon such that "in normal discourse each extension of the length of the text adds, though more and more slowly, to the number of different lexical items called on." See [Coetzee 1973b, 195].

[5] Common in areas experiencing drought and famine, and characterised most visibly by the distension of the sufferer's abdomen, kwashiorkor is a form of malnutrition that results from insufficient intake of protein.

[6] Spearman's rank correlation coefficient is a measure of the correlation between two variables in a bivariate data set. For example, if one wanted to measure the degree to which the height and mass of each individual in a given set of n students are correlative, one would first rank the students according to each of the two variables (height and mass), calculate the difference in ranks for each student (d) and then calculate the coefficient using the formula:

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2 - 1)}$$

Figure 1.

The resulting coefficient (r_s) ranges from 1 (perfect positive correlation), through 0 (no correlation), to -1 (perfect negative correlation).

Works Cited

- Attridge 2004** Attridge, D. 2004. *J. M. Coetzee and the Ethics of Reading*. Chicago: University of Chicago Press.
- Attwell 1993** Attwell, D. 1993. *J. M. Coetzee: South Africa and the Politics of Writing*. Berkeley: University of California Press.
- Beckett 1959** Beckett, S. 1959. *Watt*. New York: Grove Press.
- Beckett 1970** Beckett, S. 1970. *Lessness*. London: Calder & Boyars.
- Beckett 1984** Beckett, S. and Cohn, R. 1984. *Disjecta*. New York: Grove Press.
- Begam 1992** Begam, R. 1992. An interview with JM Coetzee. *Contemporary Literature*, 33 (3), pp. 419-431.
- Bloch 1953** Bloch, B. 1953. Linguistic Structure and Linguistic Analysis. In: Hill, A. eds. 1953. Washington DC: Georgetown University, pp. 40-44.
- Burrows 2002** Burrows, J. 2002. "Delta": A Measure of Stylistic Difference and a Guide to Likely Authorship *Literary & Linguistic Computing*, 17 (3): 267-287.
- Canepari-Labib 2005** Canepari-Labib, M. 2005. *Old Myths - Modern Empires*. Oxford: Peter Lang.

- Coetzee 1963a** Coetzee, J. M. 1963. *The Work of Ford Madox Ford*. MA. University of Cape Town.
- Coetzee 1963b** Coetzee, J. M. 1963. Computer Poem. *The Lion and the Impala*, 2 (1), pp. 12-13.
- Coetzee 1969a** Coetzee, J. M. 1969. Statistical Indices of "Difficulty", *Language and Style* 2 (3), pp. 226-232.
- Coetzee 1969b** Coetzee, J. M. 1969. *The English Fiction of Samuel Beckett: An Essay in Stylistic Analysis*. Ph.D. University of Texas at Austin.
- Coetzee 1969c** Coetzee, J. M. 1969. *English/Linguistics 211*. [Lecture Notes]. National English Literary Museum, Grahamstown.
- Coetzee 1971** Coetzee, J. M. 1971. Review of Wilhelm Fucks, *Nach allen Regeln der Kunst*. *Style*, 5 (1), pp. 92-94.
- Coetzee 1973a** Coetzee, J. M. 1973. Samuel Beckett and the Temptations of Style. In: Coetzee, J. M. and Attwell, D. eds. 1992. Cambridge, MA: Harvard University Press, pp. 43-49.
- Coetzee 1973b** Coetzee, J. M. 1973. Samuel Beckett's Lessness: An Exercise in Decomposition. *Computers and the Humanities*, 7 (4), pp. 195--198.
- Coetzee 1974** Coetzee, J. M. 1974. *Dusklands*. London: Secker & Warburg.
- Coetzee 1977** Coetzee, J. M. 1977. *In the Heart of the Country*. London: Secker & Warburg.
- Coetzee 1978** Coetzee, J. M. 1978. Hero and Bad Mother in Epic, a Poem. *Staffrider*, March p. 36.
- Coetzee 1980** Coetzee, J. M. 1980. *Waiting for the Barbarians*. London: Secker & Warburg.
- Coetzee 1986** Coetzee, J. M. 1986. *Foe*. London: Secker & Warburg.
- Coetzee 1997** Coetzee, J. M. 1997. *Boyhood: Scenes From Provincial Life*. London: Secker & Warburg.
- Coetzee 1999** Coetzee, J. M. 1999. *Disgrace*. London: Secker & Warburg.
- Coetzee 2002** Coetzee, J. M. 2002. *Youth*. London: Secker & Warburg.
- Coetzee 2003** Coetzee, J. M. 2003. *Elizabeth Costello*. London: Secker & Warburg.
- Coetzee 2005** Coetzee, J. M. 2005. *Slow Man*. London: Secker & Warburg.
- Coetzee 2007** Coetzee, J. M. 2007. *Diary of a Bad Year*. London: Harvill Secker.
- Coetzee 2013** Coetzee, J. M. 2007. *The Childhood of Jesus*. London: Jonathan Cape.
- Coetzee and Attwell 1992** Coetzee, J. M. and Attwell, D. 1992. *Doubling the Point*. Cambridge, Mass.: Harvard University Press.
- Craig 1999** Craig, H. Authorial Attribution and Computational Stylistics: If You Can Tell Authors Apart, Have You Learned Anything About Them? *Literary & Linguistic Computing*, 14 (1): 103-113.
- Egerer 1999** Egerer, C. 1999. Hybridizing the Zero: Exploring Alternative Strategies of Empowerment in J. M. Coetzee's *Foe*. *Joensuu: Studia Carelica Humanistica*, 14 pp. 96-101.
- Forsyth et al. 1999** Forsyth, R. S., D. Holmes, and E. Tse. 1999. "Cicero, Sigonio, and Burrows: Investigating the Authenticity of the *Consolatio*." *Literary & Linguistic Computing*, 14: 375-400.
- Fromm 2000** Fromm, H. 2000. Coetzee's Postmodern Animals. *The Hudson Review*, 53 (2), pp. 336-344.
- Head 1997** Head, D. 1997. *J.M. Coetzee*. Cambridge: Cambridge University Press.
- Holmes et al. 2001** Holmes, D. I., M. Robertson, and R. Paez. Stephen Crane and the New York Tribune: A Case Study in Traditional and Non-Traditional Authorship Attribution. *Computers and the Humanities* 35.3: 315-31.
- Jahn 2007** Jahn, C. 2007. *J.M. Coetzee : Novelist and Moralist*. Norderstedt, Germany: GRIN Verlag.
- Kannemeyer 2012** Kannemeyer, J. C. 2012. *J.M. Coetzee: A Life in Writing*. London: Scribe.
- Kossew 1996** Kossew, S. 1996. *Pen and Power*. Amsterdam: Atlanta.

- Lamb 2010** Lamb, J. 2010. "The true words at last from the mind in ruins": J. M. Coetzee and Realism. In: Bradshaw, G. and Neill, M. eds. 2014. *J. M. Coetzee's Austerities*. London: Ashgate, pp. 177-183.
- Leavis 1930** Leavis, F. R. 1930. Mass Civilisation and Minority Culture. In: Storey, J. eds. 2006. *Cultural Theory and Popular Culture: A Reader*. Harlow: Prentice Hall.
- Mulhall 2009** Mulhall, S. 2009. *The Wounded Animal*. Princeton: Princeton University Press.
- Paling 2009** Paling, S. 2009. Mapping Techno-Literary Spaces: Adapting Multiple Correspondence Analysis for Literature and Art Informatics. *Information Research*, 14(2) paper 401.
- Phillips 1998** Phillips, C. 1998. Life and Times of John C. *English in Africa*, 25 (1), pp. 61-70.
- Ramsay 2011** Ramsay, Stephen. 2011. *Reading Machines: Toward an Algorithmic Criticism*. Chicago: University of Illinois Press.
- Scott 1997** Scott, Joanna. 1997. "Voice and Trajectory: An Interview with J. M. Coetzee." *Salmagundi: A Quarterly of the Humanities and Social Sciences*, 114-115, (Spring-Summer 1997): 82-102.
- Silvani 2012** Silvani, R. 2012. *Political Bodies and the Body Politic in J.M. Coetzee's Novels*. Zürich: LIT.
- Stewart 2003** Stewart, Larry L. 2003. Charles Brockden Brown: Quantitative Analysis and Literary Interpretation *Literary & Linguistic Computing*, 18 (2): 129-138.
- Vanzanten Gallagher 1991** Vanzanten Gallagher, S. 1991. *A Story of South Africa: J. M. Coetzee's Fiction in Context*. Cambridge: Harvard U.P.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.