



PROJECT MUSE®

Periodizing Samuel Beckett's Works: A Stylochronometric Approach

Dirk van Hulle, Mike Kestemont

Style, Volume 50, Number 2, 2016, pp. 172-202 (Article)

Published by Penn State University Press

DOI: <https://doi.org/10.1353/sty.2016.0003>



➔ *For additional information about this article*

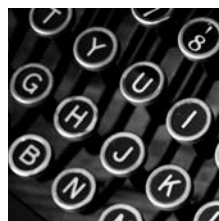
<https://muse.jhu.edu/article/618453>

Periodizing Samuel Beckett's Works

A Stylochronometric Approach

Dirk van Hulle and Mike Kestemont

UNIVERSITY OF ANTWERP



ABSTRACT: We report the first analysis of Samuel Beckett's prose writings using stylometry, or the quantitative study of writing style, focusing on grammatical function words, a linguistic category that has seldom been studied before in Beckett studies. To these function words, we apply methods from computational stylometry and model the stylistic evolution in Beckett's oeuvre. Our analyses reveal a number of discoveries that shed new light on existing periodizations in the secondary literature, which commonly distinguish an "early," "middle," and "late" period in Beckett's oeuvre. We analyze Beckett's prose writings in both English and French, demonstrating notable symmetries and asymmetries between both languages. The analyses nuance the traditional three-part periodization as they show the possibility of stylistic relapses (disturbing the linearity of most periodizations) as well as different turning points depending on the language of the corpus, suggesting that Beckett's English oeuvre is not identical to his French oeuvre in terms of patterns of stylistic development.

KEYWORDS: function words, periodization, Samuel Beckett, stylochronometry, stylometry

INTRODUCTION

Probably best known as the author of *En attendant Godot* [*Waiting for Godot*], Samuel Beckett was not only a bilingual playwright, but also a poet, translator, essayist, and novelist. Notably his prose fiction will be the focus of this article, in which we propose a quantitative method to delineate a periodization of Beckett's oeuvre. In art studies in general, there is a tradition of distinguishing an "early" and "late" period in an artist's work, sometimes with a distinct "middle" period in between. The late Beethoven sonatas are a good example, or the early Rembrandt's "smooth" style versus the

rough paint surfaces of the late Rembrandt. In literature, the “early” novels of Jane Austen (such as *Elinor and Marianne*, *First Impressions*, and *Susan*) are distinguished from their published counterparts (*Sense and Sensibility*, *Pride and Prejudice*, *Northanger Abbey*), and even earlier works are referred to as “youthful writings” by “Young Jane” (Byrne). Shakespeare’s plays are commonly grouped into the “early,” “middle,” and “late plays.” But it is often difficult to determine exactly when an author’s work moves from, say, the “early” to the “middle” stage.

In Beckett studies, we find the same pattern of periodization, ending with the “late style” (Gontarski). But we also observe the difficulty of clearly determining where one period ends and the next one begins. Peter Boxall problematizes the idea of periodizing Beckett’s oeuvre, because the neatness of such a narrativization entails the danger of doing injustice to the singularity of the individual works. But he admits that it is hard not to parcel them into a beginning, a middle, and an end. And his version of this narrative runs “from the Joycean extravagance of his early, mannered work, through the comic agony of frenzied becoming in his middle period, to the bleached impossibility of his later prose” (34). He sketches an outline of these three phases: the early period up to and including the novel *Watt*, written during the Second World War; the rich middle period up to and including *The Unnamable*; and the later, “stunted” and “halting” prose after the close of *The Unnamable* (33).

In “Early Beckett,” the opening essay in *The New Cambridge Companion to Samuel Beckett*, John Pilling suggests a break that is marked by Beckett’s decision to write in French:

Eventually, though it took a long time to emerge, it became expedient for Beckett to insist upon a break between his pre-war writing and his post-war writing, as expressed not only in conversation with Charles Juliet but with many others. By abandoning his native English, the line became easier to draw. (28)

This periodization, however, is more complex than it may seem. During the war, Beckett wrote a novel in English: *Watt* is therefore neither “pre-war” nor “postwar.” Moreover, the (end of the) Second World War did not coincide exactly with Beckett’s decision to start writing in French. He had already written several poems in French in 1937 and 1938. This essay will

concentrate on Beckett's prose fiction, and—with this focus in mind—"early Beckett" can be said to consist of the collection of stories *More Pricks than Kicks* (which Beckett started writing in 1931, published in 1934), *Dream of Fair to Middling Women* (written in 1932, published posthumously in 1992), and the novels *Murphy* (started in 1935, published in 1938) and *Watt* (started during the War in 1941, published in 1953).

The "middle Beckett" is often considered to start with the *Nouvelles* (*La Fin*, *L'Expulsé*, *Premier Amour*, *Le Calmant*) and the novel *Mercier et Camier*. But again, this should be nuanced. Beckett started writing the first of the *Nouvelles* in English. He "drew the line" between English and French by literally drawing a line in the middle of his story *La Fin* (originally called *Suite*). The manuscript held at Boston College (BC MS 1991-001, ref 53, 31r) shows how Beckett first started in English and, after about 28 pages, suddenly drew a line (in March 1946) and continued writing in French. After this story, he wrote the novel *Mercier et Camier* (started in 1946, published in 1970 in French, in 1974 in Beckett's English translation) and the other *Nouvelles*: *L'Expulsé* (1946, published in 1946; *The Expelled*, published in 1962); *Premier Amour* (1946; published in 1970; *First Love*, published in 1973); *Le Calmant* (1946, published in 1955; *The Calmative*, published in 1967).

The remarkably creative five-year period after the War was referred to (by Beckett himself) as "the siege in the room" (Knowlson; Bair). In this period, he also wrote the three novels *Molloy* (started in 1947, published in 1951; English version published in 1955), *Malone meurt* (started in 1948; published in 1951; *Malone Dies*, published in 1956), and *L'Innommable* (started in 1949, published in 1953; *The Unnamable*, published in 1958). These three novels have been published as a so-called "trilogy."¹

Still, that unity did not prevent critics from categorizing the novels in separate periods. In Brian McHale's periodization of Beckett's works, the border runs between *Molloy* and *Malone meurt* [*Malone Dies*]. In *Constructing Postmodernism*, McHale indicated a difference between modernism and postmodernism that was marked by "the distinction between the cognitivist and the postcognitivist Beckett" (34). McHale thus implied a distinction between an early Beckett and a later Beckett, that is, "the Beckett who is still preoccupied with modernist issues of reliability and unreliability of narrators, radical subjectivity, and multiplicity of perspectives, as in *Watt* and *Molloy*" (34) and "the Beckett who focuses instead on the status of fictional worlds, the power (and impotence) of language to make and unmake worlds,

and the relationship between fictional being and elusive “real” being, as in *Malone Dies*, *The Unnamable*, and many of the later short texts” (34).

H. Porter Abbott situates the border after the *The Unnamable*: “Watt, *Mercier et Camier*, the *Nouvelles*, and the ‘trilogy,’ all conform to the quest structure, despite the manifold incompetence of the questers” (*Beckett Writing Beckett* 89). But the next work, *Textes pour rien* (started in 1951, published in 1955; *Texts for Nothing*, published in 1967), “marks a pause in the story of the œuvre” (89).² According to this periodization, it is the “willful shredding of narrative linearity within the *Texts*” (90) that marks the end of a period, which coincides with the end of the so-called “great creative period” (Federman and Fletcher 63). The *Texts for Nothing* were followed by *Foirades* (most of which were written in the 1960s; published in 1973; *Fizzles* published in 1976) and *From an Abandoned Work* (started in 1955, published in 1956; *D’un ouvrage abandonné*, published in 1967).

Comment c’est (started in 1958, published in 1961; *How It Is*, published in 1964) is considered another pivotal work. The title has been interpreted as a “pun on beginnings (*comment c’est* is *commencer*)” (Abbott, *Beckett Writing Beckett* 102), and would in that sense mark the beginning of the late period. Again, the notion of a (failed) beginning was emphasized in the very short *Faux départs* (1964, published in 1965). The subsequent prose pieces were usually so short that—simply in quantitative terms—most of them can barely be used for stylometric analysis: *All Strange Away* (started in 1964, published in 1976); *Imagination morte imaginez* (written and published in 1965; *Imagination Dead Imagine*, published in 1965); *Assez* (written in 1965, published in 1966; *Enough*, published in 1967); *Bing* (written and published in 1966; *Ping*, published in 1967); *Sans* (written and published in 1969; *Lessness*, published in 1970); *Le Dépeupleur* (written and published in 1970; *The Lost Ones*, published in 1972); *Abandonné* (started in 1970, published in 1972); *Still* (started in 1972, published in 1975); *Sounds* (started in 1972, published in 1978); *Still 3* (started in 1973, published in 1978); *As the Story Was Told* (written and published in 1973); *La Falaise* (started in 1974, published in 1975); *Un Soir* (started in 1979, published in 1980; *One Evening*, published in 1980); *The Way* (written and published in 1981); *Ceiling* (written in 1981, published in 1985; *Plafond*, published in 1985). The slightly longer prose texts are sometimes referred to as the “second trilogy” (although Beckett did not like this term) and were published both separately and together under the title *Nohow On: Company* (started in

1976, published in 1980; *Compagnie*, published in 1980); *Mal vu mal dit* (started in 1980, published in 1981; *Ill Seen Ill Said*, published in 1981) and *Worstward Ho* (started in 1981, published in 1983), followed by the short text *Stirrings Still* (started in 1983, published in 1989; *Soubresauts*, published in 1989).

The last text, *Comment dire* [*what is the word*], is again very short and usually treated as a poem, although it might just as well be regarded as a prose text—a piece about a sentence that does not manage to get written. It ends with the words “comment dire,” or “what is the word” in Beckett’s translation. Beckett thus made his oeuvre end in the middle of a sentence—a poetical statement indicating his view on his work in terms of continuous incompleteness. The Joycean notion of a “work in progress” is not entirely suitable to denote this continuing incompleteness as it implies the notion of improvement (“progress”), whereas Beckett preferred to think in terms of “mere gress” because of its “purity from destination and hence from schedule” (*Letters* 186).

In spite—and because—of Beckett’s aversion to “schedule” and “destination,” scholars have tried to at least determine stages in his aesthetic “gress.” John Bolin, for instance, focuses on the voices of the professor and the poet (the voice of ironic commentary versus a poetic voice that is open to the contingent) arguing that the latter is given greater prominence for the first time in the novel *Watt*. From this perspective, Beckett’s last novel in English before he switched to writing fiction in French thus becomes a pivotal work that closes the early period, separating it from his mature work, more or less in accordance with Pilling’s periodization. But other critics already discern aspects of the later work in the writings of the 1930s. For instance, Fintan O’Toole finds “the seed of Beckett’s mature work” in passages from the unpublished story “Echo’s Bones” (written in the Fall of 1933, published posthumously in 2014 in an annotated edition by Mark Nixon).

Ann Banfield suggests a four-phase evolution toward “tattered syntaxes,” reflected in four types of title: (a) the clever inventiveness of *More Pricks than Kicks* and *Dream of Fair to Middling Women*; (b) the proper names such as *Murphy*, *Watt*, *Mercier and Camier*, and *Molloy*; (c) the definite noun phrase in titles such as *The Expelled* and *The Unnamable*; and (d) titles that lack highly specified semantic content, such as *Enough*, *Still*, *All Strange Away* (17). And Chiara Montini suggests yet another four-phase division, based on Beckett’s use of different languages: (a) *le monolinguisme polyglotte* (1929–1937); (b) *le bilinguisme à dominance Anglophone* (1937–1945); (c) *le*

bilinguisme à dominance francophone (1946–1953); and (d) *le bilinguisme mixte* (“Sinking” 65).

The various opinions may lead us to a relativistic conclusion that periodization is merely a way of giving in to an urge to turn time into a plot, which Frank Kermode famously described in *The Sense of an Ending* as the “tick-tock” model, an organization which humanizes time by giving it a form (45). But perhaps this tendency to categorize or periodize an oeuvre is more than a subjective feeling. It is possible that it is based on stylistic developments that the human interpreter “senses” but cannot always illustrate with empirical data. The hermeneutic approach thus resorts to explanations such as the ones described previously, resulting in almost as many different periodizations as the number of critics that devise them. We therefore wondered what a nonhuman “interpreter” would come up with as a periodization by means of stylometry.

STYLOMETRY AND STYLOCHRONOMETRY

Stylometry or computational stylistics is a multidisciplinary research domain that uses computational methods to study writing style (Holmes). To some extent, it can be considered a form of Artificial Reading, in which computers are used to simulate and potentially enhance or complement human reading. Because of its heavy reliance on quantification, stylometry is commonly considered a part of the Digital Humanities or Humanities Computing (Schreibman and Siemens) and it is indeed in this domain that stylometry has seen some of its seminal applications, such as the well-known study of authorship in the *Federalist Papers* by Mosteller and Wallace or John Burrows’s pioneering 1987 study of the speech of the different characters in Jane Austen’s novels.

Stylometric research is typically concerned with the relationship between a text’s writing style and its metadata. In the popular field of authorship attribution (Love; Juola, “Authorship”; Stammatatos; Koppel, Schler, and Argamon), for instance, researchers attempt to establish a quantitative link between an author’s identity and her/his unique writing style. In many cases, these methodologies have been able to attribute anonymous texts to a known author. Especially in the fields of ancient and medieval literature (which abound in anonymous texts), stylometry has generated a number of thought-provoking results, for instance, as to the attribution of disputed

texts, such as the Shakespeare–Marlowe controversy (Burrows, “Second Opinion”). Further research has focused on other forms of “author profiling,” such as the automatic detection of the age or gender of a text’s author (Daelemans). In literary studies, stylometry has moreover proved useful in order to arrive at a better understanding of the stylistic (dis)similarities between texts from different genres or periods (Jannidis and Lauer).

This study of periodization in Samuel Beckett’s works sets out from a slightly less studied subfield of stylometry, usually referred to as “stylochronometry” (Stamou), in which a text’s writing style is studied as a function of its date of composition. For instance, research into medieval charters has shown that quantitative measurements enable us to date undated charters on the basis of stylistic features (Tilahun, Feuerverger, and Gervers). Other recent research in stylochronometry has used diachronic models to study the development of the writing style in the works of individual authors. A textbook example is David Hoover’s recent analysis of Henry James’s oeuvre, in which he was able to detect striking diachronic trends. In the field of Anglo-Saxon literature, other studies have reported similar findings for the works of W. B. Yeats (Forsyth) or Jack London (Juola, “Becoming”).

Most stylometric research in literary studies today focuses on stylistic features that can be automatically extracted from (digitized) texts. Whereas seminal studies focused on the analysis of rather superficial characteristics, such as average word or sentence length, a variety of more advanced stylistic features has been explored in the recent literature, including stylistic characteristics related to semantics, morpho-syntax, metaphors, intertextuality, and so on (Stamatatos). In this article, we focus on a feature type that has proven surprisingly successful to model writing style across various languages, authors, and genres, namely “function words” (Binongo; Rybicki and Eder).

Function words refer to the small set of (typically short) words, including articles (“the”), prepositions (“of”), conjunctions (“or”), and pronouns (“she”). These words serve a predominantly grammatical function in a language and they do not carry any straightforward meaning when used in isolation—as opposed to, for instance, the much more straightforward semantics of most nouns (“cat”) or verbs (“eat”), which are often called “content words.” Function words tend to occur very commonly in (not only literary) texts, whereas the much larger and diverse group of content words occurs much less frequently. Apart from the practical advantage that function words can normally be easily identified by computer algorithms (in English or French

texts), there are also a number of theoretical considerations that help explain the attractiveness of function words for stylistic research (Kestemont).

Function words appear frequently throughout texts, which makes them a rich basis for quantitative research (Binongo). Because most texts written in a particular language will have to use the same set of function words—one can hardly write an English text without using *the* or *and*—these function words constitute a reliable base unit for textual comparison. Empirical studies have shown that precisely function words capture a variety of stylistic choices, giving important clues as to an author's gender, identity, or sentiment (Stamatatos). Finally, research in authorship attribution suggests that authors use function words to a large extent *unconsciously*—which seems especially the case in longer prose texts (Pennebaker). This makes function words a refreshing object of research, because they lend themselves well to modeling stylistic aspects that are less deliberately constructed by authors. In Beckett studies, stylistic research has so far mainly focused on rhetoric (Clément), self-translation (Sardin; Montini, *Bataille*; Mooney) and the so-called “tattered” syntax of the late works (Banfield). Of these studies, only the latter explicitly takes function words into account. The following section delineates the methodology by means of which we applied the study of function words in a stylochronometric analysis of Beckett's works.

MATERIALS AND PREPROCESSING

In this study, we analyzed Beckett's prose fiction, in both French and English. Naturally, the limitation to these (lengthier) prose writings does not do justice to Beckett's achievements as a playwright, poet, and essay writer, although future research might easily expand the scope of our investigations to include other genres. We obtained Beckett's oeuvre in digital form, mostly through purchasing the prose works in epub format and extracting the plain text. From this corpus, we removed all non-authorial paratexts, such as prefaces or page numbers. To further process these texts, we identified individual words in these materials using a standard tokenizer (Bird, Klein, and Loper).³

Table 1 gives an overview of the materials we collected, including the publication dates of the editions we used,⁴ the text's size, and other metadata. In terms of chronology, we distinguish between a text's official publication date in French and/or English, as well as the moment

Table 1: Table of the analyzed corpus

Start_ date	Title_en_long	Title_en	Pub_ date_en	Container_ text	Title_fr_long	Title_fr	Pub_ date_fr	Orig_ lang	Words (en)	Word (fr)
1931	More Pricks than Kicks	kicks	1934		NA	NA	NA	en	54,510	NA
1932	Dream of Fair to Middling Women	dream	1992		NA	NA	NA	en	72,411	NA
1935	Murphy	murphy	1938		Murphy	murphy	1947	en	58,560	64,597
1941	Watt	watt	1953		Watt	watt	1968	en	84,132	89,746
1947	Molloy	molloy	1955		Molloy	molloy	1951	fr	94,122	82,554
1948	Malone Dies	malone	1956		Malone meurt	malone	1951	fr	49,750	52,313
1949	The Unnamable	unnamable	1958		L'Innomable	innomable	1953	fr	60,178	63,055
1946	The End	end	1954	Nouvelles	La fin	fin	1955	fr	8,867	9,252
1946	The Expelled	expelled	1962	Nouvelles	L'expulse	expulse	1946	fr	5,691	5,996
1946	First Love	love	1973	Nouvelles	Premier amour	amour	1970	fr	8,391	9,456
1946	The calmativ	calmativ	1967	Nouvelles	Le calmant	calmant	1955	fr	6,534	7,033
1958	How It Is	how	1964		Comment c'est	comment	1961	fr	37,028	38,022
1946	Mercier and Camier	mercier	1974		Mercier et Camier	mercier	1970	fr	28,424	36,008
1976	Company	company	1980		Compagnie	compagnie	1980	en	9,574	10,643

(Continued)

Start_ date	Title_en_long	Title_en	Pub_ date_en	Container_ text	Title_fr_long	Title_fr	Pub_ date_ fr	Orig_ lang	Words (en)	Word (fr)
1980	Ill Seen Ill Said	ill	1981		Mal vu mal dit	mal	1981	fr	7,860	8,533
1981	Worstward Ho	ho	1983		NA	NA	NA	en	4,501	NA
1983	Stirrings Still	still	1989		Soubresauts	soubresauts	1989	en	1,915	1,995
1951	Texts for Nothing	nothing	1967	Texts for Nothing	Textes pour rien	rien	1955	fr	17,496	17,866
1955	From an Abandoned Work	abandoned	1956	Texts for Nothing	D'un ouvrage abandonné	abandonne	1967	en	3,762	4,111
1965	Imagination Dead Imagine	imagine	1965	Texts for Nothing	Imagination morte imaginez	imaginez	1965	fr	1,106	1,196
1965	Enough	enough	1967	Texts for Nothing	Assez	assez	1966	fr	2,283	2,420
1970	The Lost Ones	ones	1972	Texts for Nothing	Le depeupleur	depeupleur	1970	fr	8,268	8,717
1960	Fizzles	fizzles	1976	Fizzles	Foirades	foirades	1973	fr	5,566	5,808

Note: For each text (in each language), we list the original title, an abbreviated title, the language in which Beckett first composed a text, and the text's length (word count), as well as the publication date of the first edition. The "start of composition" column provides an indication as to when Beckett started writing a particular text (in either language). Some texts were originally published as part of a larger collection, which we identify in the "container text" column. NAs indicate missing information, for example, the French title of a text that was only written in English.

when Beckett started working on a text in either language (Van Hulle and Verhulst). Finally, Table 1 lists a convenient abbreviation for each text (in each language), which we will use as a shorthand to identify texts in the following graphs. While international copyright law does not allow us to redistribute the original texts, all software code necessary to replicate our findings, is freely available.⁵

As is common across a variety of studies in stylometry (Burrows, “Measure of Stylistic”; Hoover, “Multivariate”), we defined a relevant list of function words by extracting an initial list of the 300 most frequent words (MFW) from both the English and the French corpus. This number seems an acceptable choice against the background of previous studies (Stamatatos) and allows us to mine the most common function words from Beckett’s oeuvre. From this list, we have manually removed words that did not clearly fit the category of function words. In light of the focus of our analysis, we therefore removed items that correlated too strongly with the topic of particular texts, rather than with their writing style (Hoover, “Multivariate”). Removed items include the names of characters, numerals, and too topic-specific nouns (e.g., “mouth” or “chien”) or verbs (e.g., “leave” or “passer”). In Table 2, we list the 300 MFW we obtained for each corpus: items preceded by a hash tag were manually removed. After this removal, we were left with 162 function words for the English corpus and 169 for the French. Note that we refrained from removing common auxiliary verbs, such as inflected forms of *have*, because they are interestingly tied to a text’s narrative perspective, as will become clear in the following. Likewise, we have not removed any personal pronouns from these word lists (as is common in attribution studies) because we are especially interested in the stylistic shifts this category of words might capture.

PRELIMINARY ANALYSES

In this section, we describe an exploratory analysis of the material, using an established method from stylometry for data visualization: principal components analysis (PCA) (Binongo and Smith). This technique is an unsupervised procedure, in the sense that it does not require any additional information other than the texts themselves to produce a result. In this section, we do not yet explicitly integrate any chronological information in the analyses and try to establish to what extent Beckett’s oeuvre might have a

Table 2: List of the most frequent words in the English and French corpus

(a) English Corpus
<p>a, about, after, again, against, #air, all, alone, always, am, an, and, another, any, anything, are, as, at, away, back, be, because, #bed, been, before, behind, being, #belacqua, #best, #better, between, #black, #body, but, by, #came, #camier, can, #case, #celia, #certain, #come, #coming, could, #course, #dark, #day, #days, #dead, did, do, does, #don, done, #door, down, each, either, else, #end, enough, even, #evening, ever, every, #eye, #eyes, #face, #far, #feel, #feet, #felt, few, #find, #first, for, #found, #four, from, #get, #give, #go, #god, #going, #gone, #good, #got, #great, #ground, had, #hand, #hands, has, have, having, he, #head, #hear, #heard, #heart, her, here, him, himself, his, #home, #house, how, i, if, in, into, is, it, its, just, #kind, #knew, #knott, #know, #last, #least, #leave, #left, less, let, #life, #light, like, #little, ll, #long, #look, #looked, #love, m, #made, #make, #man, many, #matter, may, me, #mean, #mercier, might, #mind, mine, #miss, #moment, more, #morning, most, #mother, #mouth, #mr, much, #murphy, must, my, myself, #name, #neary, #need, never, #next, #night, no, none, nor, not, nothing, now, of, off, often, oh, #old, on, once, one, only, #open, or, other, others, our, out, over, own, #part, perhaps, #pim, #place, #point, #possible, #put, #question, quite, rather, #rest, #right, #room, #round, s, #said, same, #saw, #say, #saying, #second, #see, seemed, #seen, #set, shall, she, should, #side, #silence, since, #sky, so, some, something, sometimes, soon, #speak, still, #stop, such, t, #take, #tell, than, that, the, their, them, then, there, these, they, #thing, #things, #think, this, those, though, #thought, #three, through, till, #time, to, together, #told, too, #took, towards, #true, #turn, #turned, #two, under, until, up, us, very, #voice, #want, was, #watt, #way, we, well, #went, were, what, when, where, which, while, #white, who, why, will, #window, with, without, #woman, #word, #words, #worse, would, #wylie, yes, yet, you, your</p>
(b) French Corpus
<p>a, #abord, ai, ailleurs, ainsi, #air, #aller, alors, après, assez, au, aurais, aurait, aussi, autre, autres, aux, avaient, avais, avait, avant, avec, avoir, ayant, #bas, #besoin, #bien, #bon, #bonne, #bouche, #boue, #bout, #bras, #bruit, c, #camier, car, #cas, ce, cela, celle, celui, ces, cet, cette, ceux, #chambre, chaque, #chemin, chez, ~#chien, #chose, #choses, #ciel, comme, comment, contre, #corps, #counihan, #crois, #célia, #côté, #cœur, d, dans, de, depuis, des, dessus, #deux, devant, #dire, #dis, #dit, #dix, donc, dont, #dos, #doute, #droite, du, déjà, dû, #effet, elle, elles, en, encore, #endroit, enfin, #entendre, entre, #esprit, est, et, eu, eux, #exemple, faire, faisait, fait, faut, #façon, #femme, #fenêtre, #fils, #fin, #fois, fut, #gauche, #grand, #grande, #haut, #heure, #heures, #homme, ici, il, ils, #instant,</p>

(Continued)

Table 2: List of the most frequent words in the English and French corpus (*Continued*)

j, jamais, je, #jour, #jours, jusqu, #knott, l, la, le, les, leur, leurs, #lit, #loin, #long, longtemps, lui, #lumière, là, m, ma, #mademoiselle, #main, #mains, maintenant, mais, #maison, #mal, me, #mercier, mes, #mieux, moi, moins, #moment, mon, #monde, #monsieur, #mort, #mot, #mots, #murphy, même, n, ne, #neary, ni, #noir, non, nous, #nouveau, #nuit, on, ont, ou, oui, où, par, #parle, #parler, #part, #partie, pas, #passer, peine, pendant, #personne, #petit, #petite, peu, peut, peux, #pieds, #pim, #place, plus, plutôt, point, #porte, #possible, pour, pourquoi, pouvait, pouvoir, #premier, #première, presque, #près, pu, puis, qu, quand, #quatre, que, quel, quelle, quelque, quelques, #question, qui, quoi, #raison, #reste, rien, s, sa, #sac, #sais, #sait, sans, #savoir, se, #sens, sera, serait, ses, #seul, #seule, seulement, si, #silence, sinon, #soir, soit, son, sont, #sorte, sous, souvent, suis, #suite, sur, t, tant, tantôt, #tard, tel, #temps, #terre, toujours, #tour, tous, tout, toute, toutes, #trois, trop, très, tu, #tête, un, une, #va, #vais, vers, veux, #vie, #vieux, #visage, #vite, voilà, #voir, #vois, #voix, vous, #vrai, #vu, #watt, #wylie, y, #yeux, à, ça, étaient, étais, était, été, être, #œil
--

Note: The function word vocabulary which forms the basis of all subsequent analyses in this article. Words preceded by a hash tag were manually removed from the data to reduce the effect of, for instance, topic-related features, which are less related to writing style.

“natural” chronological structure when it comes to the author’s writing style. Previous research in stylochronometry suggests that some oeuvres indeed display an important diachronic structure, which can be detected with little supervision by scholars (Juola, “Becoming”; Hoover, “Conversation”).

The prose texts in the Beckett corpus differ significantly in length, ranging from novel-size works to the *Nouvelles* and even shorter texts (consult Table 1). To compensate for this skewedness, we sliced the texts into consecutive, non-overlapping samples of a fixed size (expressed in token length), which is a common “windowing procedure” in segmentation studies (e.g., Brooke, Hammond, and Hirst). We dropped trailing words at the end of texts, if they did not constitute an entire sample anymore. While longer texts, inevitably, still have a quantitative advantage (they contribute more samples), this procedure allowed us to study Beckett’s prose writings as a collection of equally sized text slices.

We produced a purely numerical representation of these slices of text, which can serve as the input for quantitative algorithms. Each text slice is represented as a list of real values (or “vector”), which will hold the relative frequencies of the items in the function word vocabulary we consider.

We obtained a word's relative frequency by dividing its absolute frequency in a slice by the slice's total word length. In technical terms, we “vectorized” these textual samples under a so-called bag-of-words assumption (Manning, Raghavan, and Schütze): because we only collect the frequency of individual words—as if the slice were a randomly jumbled “bag of words”—this representation is ignorant of a term's original position in a document (or document sample), or even the original word order. This is a common and efficient strategy in Computer Science.

In stylometry, it is common practice to use PCA to produce a visualization of the original texts as points in a scatterplot, which is said to capture the main stylistic variation in a data set (see Figures 1a and b). In this figure, we plot the texts using a label, which consists of the initial characters of the work's title, followed by an index (e.g., “merci6” represents the sixth slice we extracted from *Mercier and Camier*). As an aid to interpretation, we have automatically assigned the texts to four clusters and we have colored the samples accordingly. By doing so, we can quickly identify the four primary clusters in the texts.⁶

PCA has the considerable advantage that the sample labels in the plot can be overlaid with the so-called “loadings” of the original words (in a smaller, grey-colored font). These loadings offer an indication of how the original vocabulary contributed to the structure that emerges from this plot. Samples plotted in a similar region of the scatterplot can be said to be more stylistically similar. The overlaid loadings of the individual word features (in grey) reveal how individual words contribute to the position of samples in the scatterplot. For instance, word loadings plotted to the far left in Figures 1a and b are typical of samples that can be found in the same area.

From the resulting scatterplots (see Figures 1a and b), we can draw a number of interesting observations. For the English data in Figure 1a, we see how the horizontal spread is dominated by a threefold clustering, with some of Beckett's earlier works clustering in the far left. Interestingly, the loadings reveal that these works are characterized by a relatively elevated frequency of words related to a third-person narrative perspective (*he, she, his, has*, etc.). To the far right of Figure 1a, we find a dense group of labels corresponding to some of Beckett's postwar works, such as *The Unnamable* and *Texts for Nothing*. According to the loadings, these texts can be characterized by the use of first-person pronouns (*I, me*)

Figure 1 · PCA scatterplot of the most frequent words selected previously, for the (a) English and (b) French corpus, respectively. We plot the position of 3,761-word slices; label names indicate the work from which a sample was extracted. Word loadings have been overlaid in grey. A similar temporal structure is evident for both the English and French data, especially in the horizontal distribution of samples.

in combination with impersonal pronouns such as *it* and *there*, which suggests that these texts focus on the relation between an “I” and his nonpersonal surroundings. We see how texts from the in-between period (such as *Molloy* and *Watt*) also hold the middle in the horizontal distribution of samples. Interestingly, the visualization shows that a single sample jumps out with respect to the vertical dimension of the plot: *Worstward Ho* (and to a lesser extent, *How It Is*) is characterized by what seems, at first sight, a rich mix of fairly abstract quantifiers and determiners (*other, none, sometimes*, etc.), often with an indeterminate semantics ([*n*]ever, same, again, all, etc.).

As to the French scatterplot in Figure 1b, it is surprising how similar the structure is that arises from the French prose. To the far left, we find some of the earlier works (*Watt, Murphy*), which are also characterized by a dominant third-person perspective (cf. pronouns *ses, sa, son*, etc.). These works are placed in opposition to the sample cloud containing the *L'Innommable* samples to the far right, in which the use of the first-person perspective is striking (*j', je, me, suis, moi*, etc.), again in combination with the impersonal *c'est*. For the French data, too, we find that the vertical variation focuses on displaying aspects of Beckett's later writings, most notably *Comment c'est*, and to a lesser extent *Foirades* and *Mal vu mal dit*. Here, too, the PCA reveals a shift in Beckett's vocabulary to a set of indeterminate words (*tous, autre, peu*, etc.), including abstract oppositions such as *toujours* versus *jamais* or *oui* versus *non*.

Thus, interestingly, both scatterplots horizontally create an opposition between earlier and later writings in Beckett's oeuvre, focusing on an opposition between a first-person and a third-person perspective. With respect to the vertical dimension, both analyses show a tendency to focus on a vocabulary shift, which seems to occur in Beckett's later, shorter writings, toward a more abstract and indeterminate vocabulary, dominated by adverbs and particles with an indeterminate and abstract semantics. This tendency corresponds with Ann Banfield's argument that “the late style is extracted from the lexical formatives with the least semantic content” (20). *From an Abandoned Work*, again, proves to represent an exception to this rule, being closest to a group of *Molloy* in the English plot (Figure 1a) and occupying a relatively central, even neutral position in the French plot (Figure 1b), in between samples of *Mercier and Camier* and *Comment c'est*.

CHRONOLOGICAL ANALYSIS

In the previous section, we have shown how an exploratory statistical technique can be used to analyze the stylistic structure of Beckett's prose fiction. However, this analysis was still ignorant of the diachronic structure of the data, meaning that samples from Beckett's early works could in principle just as easily cluster with his later writings. This prevented us so far from identifying clear turning points in Beckett's career. Therefore, we now move to a quantitative method that is able to take into account the actual chronology of Beckett's works.

Variability-based Nearest Neighbor Clustering (VNC) is a clustering technique that has recently been introduced by Gries and Hilpert in the context of the diachronic analysis of corpus linguistic data. For this research, Gries and Hilpert considered corpora with an important diachronic component, such as child language acquisition data. Their aim was to identify distinct temporal stages by pinpointing the main turning points. Gries and Hilpert proposed a simple adaptation of existing clustering techniques. Traditionally, cluster analyses of texts are graphically represented using a "dendrogram": a kind of tree representation in which the connections between texts express which texts are more similar to one another (see Figures 2a and b for an example). Agglomerative clustering, for instance, is a "bottom-up" procedure, which will start at the outer twigs of the tree, that is, the original text represent the leaves of the tree (Hoover, "Multivariate").⁷ Iteratively, these newly created twigs will in turn be merged into higher-level nodes, until all nodes eventually join in a single root node or "stem."

In conventional dendrograms, each node can be freely combined with any other node in the tree, thus potentially scrambling the original chronological order of the data (i.e., an early text could easily cluster with later text if they happen to be similar). In their VNC approach, Gries and Hilpert add the simple constraint that only consecutive nodes, which are immediately adjacent in time, can form new clusters. This restriction enables analyses in which the chronological structure of the data is reflected in the top branches of trees, representing the main diachronic stages in the data.⁸

We have run a VNC on our data: the resulting trees offer a clear insight into the chronological structure of Beckett's oeuvre. Figure 2a, for the English prose, displays a clear initial cluster of Beckett's earliest two novels, *More Pricks than Kicks* and *Dream of Fair to Middling Women*. This initial cluster

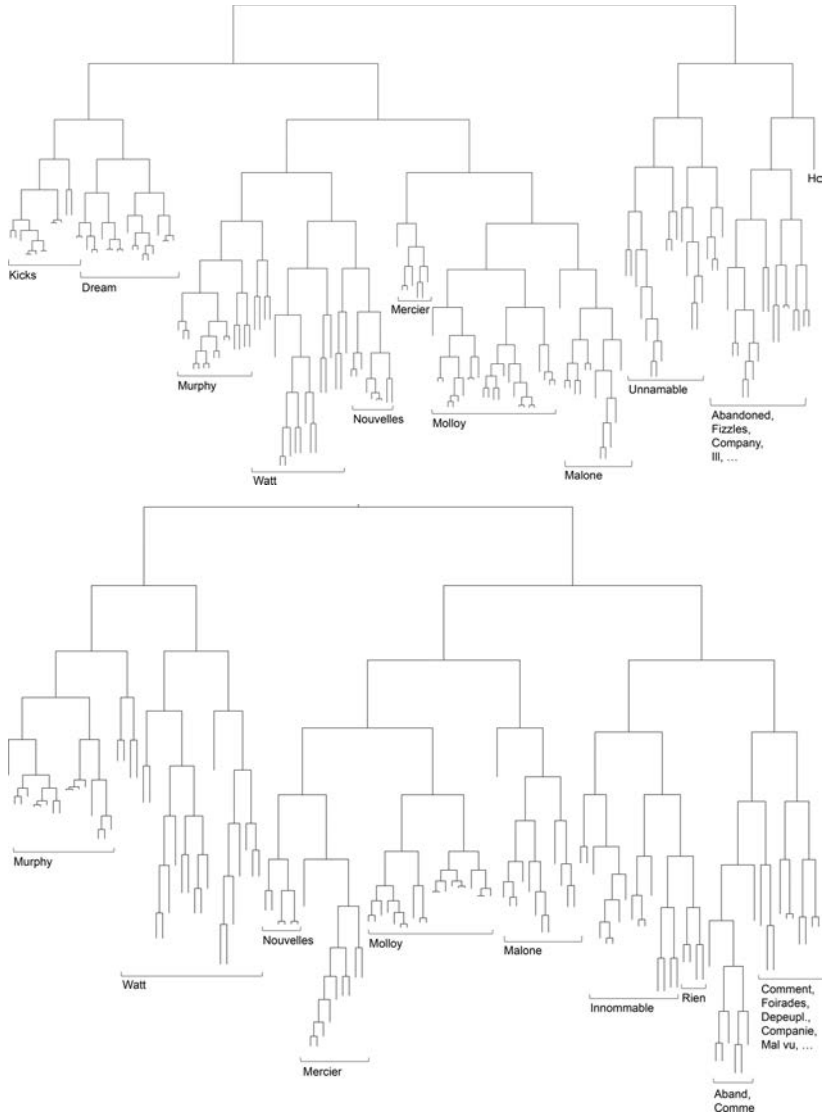


Figure 2 · A Variability-Based Neighbor Analysis of Beckett's prose writings, (a) English and (b) French. These visualizations offer a tree-like representation of Beckett's oeuvre: the connections in such "dendrograms" show which texts are more similar to one another. Starting from the leaves of the tree (i.e., the original samples from the texts), newly created twigs are merged into higher-level nodes. To produce a diachronically insightful result, the VNC analysis only allows nodes to merge, if they are immediately adjacent in time. Diachronically speaking, the trees can be read from left to right.

is of course absent from the French prose's VNC dendrogram in Figure 2b, since these texts lack a translated counterpart.⁹ Otherwise, the structures of Figures 2a and b run remarkably parallel. *Murphy* and *Watt* constitute the second chronological cluster of works, together with the *Nouvelles*. Only at a higher level is the former group paired with the cluster consisting of *Mercier and Camier*, *Molloy* and *Malone Dies*. Interestingly, the original French versions of the *Nouvelles* are joined with the next cluster, whereas the English versions are joined with the previous cluster, which indicates a different status of this collection in both languages. The last major branch for both languages holds the tight group of branches uniting *The Unnamable*, *Texts for Nothing*, *How It Is*, and the series of shorter late works. In the English tree, it is worth noting that *Worstward Ho* occupies a fairly pronounced position, emphasizing again the special status of this work. Beckett claimed that he could not translate *Worstward Ho* and this "untranslatability" is directly connected to the style of this short text (as we will discuss in the last section of this article).

The VNC analysis brings us closer to answering our original research question, but it does not yet identify *exact* turning points. To achieve this, we had the clustering procedure exactly identify the n text slices that it considers to be the most significant turning points. Apart from the chronological ordering of the original data points, this procedure only required us to predetermine the number of diachronic stages or "segments" we wanted to obtain; for example, for the setting $n = 3$, we would obtain three diachronic stages, indicating an early, middle, and late Beckett. We derived this segmentation approach from parallel studies in audio research, where efficient libraries exist to divide audio fragments such as songs into a number of consecutive segments. In Table 3, we list the results obtained for increasing numbers of turning points.¹⁰

When asked to identify a single turning point in Beckett's oeuvre, the analysis for French singles out a sample from the *Nouvelles*, when considering the entire MFW vocabulary. Interestingly, the English analysis opts for the first sample of *The Unnamable*. We also observe a difference in stylistic status for *Premier amour* [*First Love*] in both languages. Our analyses suggest that the English version of *First Love*, together with the other translated *Nouvelles*, behaves more like Beckett's early prose in English than the story's French counterpart, which is remarkable, since *Premier amour* was written in 1946, whereas the English translation was made more than two

Table 3: Segmentation of Beckett's French and English writings through time

<i>n</i> (Number of segments)	Turning points: English prose	Turning points: French prose
2	unnamable_1	amour_1
3	watt_5, unnamable_1	amour_1, abandonne_1
4	watt_5, unnamable_1, abandoned_1	amour_1, innommable_1, abandonne_1,
5	watt_5, unnamable_1, abandoned_1, ho_1	amour_1, innommable_1, abandonne_1, comment_9

Note: This analysis considers all MFW listed previously, and offers the results for various numbers of turning points (2–5).

decades later, in the early 1970s. At first sight, this result is largely in line with previous scholarship, which had the start of the “middle Beckett” coincide with the *Nouvelles*, even though these periodizations are based on different grounds. If we zoom in, however, the stylometric analyses do show remarkable stylistic oddities that nuance the rough-and-ready, three-phase periodization. For the three-phase periodization, the results of the stylometric analysis show different turning points depending on the language of the corpus (*Watt* and *The Unnamable* for the English texts; *Premier amour* and *D'un ouvrage abandonné* for the French texts).

When asked to choose additional turning points, both analyses identify the start of *The Unnamable* and *From an Abandoned Work*, but not in the same order. *L'Innommable* [*The Unnamable*] is the least surprising turning point, as most of the existing periodizations mention it as a pivotal work. But usually it is the *end* of this work that is pinpointed as the watershed. This traditional periodization is partly inspired by the novel's famous closing line “I can't go on, I'll go on,” marking both an end (“I can't go on” in the present tense) and a beginning (“I'll go on” in the future tense). These semantic connotations do not influence the stylochronometric periodization, which interestingly marks the *beginning* of *L'Innommable* [*The Unnamable*] as a turning point, thereby breaking up the “unit” of the so-called “trilogy.” Opening with its three basic narratological questions (Where? Who? When?), the novel's beginning marks a break in terms of narrative conventions. Whereas, from a narratological point of view, *Molloy* and *Malone Dies* still presented more or less coherent story lines, *The*

Unnamable shifts the attention radically from story to discourse. The opening lines of *L'Innommable* [*The Unnamable*] are a good example of the novel's stylistic features, such as the emphasis on negation, repetition, hesitation, and contradiction. As in the case of *Worstward Ho*, *L'Innommable* proved to be almost untranslatable. But whereas Beckett eventually could not translate *Worstward Ho*, he did manage to translate *L'Innommable* into English:

Où maintenant ? Quand maintenant ? Qui maintenant ? Sans me le demander. Dire je. Sans le penser. Appeler ça des questions, des hypothèses. Aller de l'avant, appeler ça aller, appeler ça de l'avant. Se peut-il qu'un jour, premier pas va, j'y sois simplement resté, où, au lieu de sortir, selon une vieille habitude, passer jour et nuit aussi loin que possible de chez moi, ce n'était pas loin. Cela a pu commencer ainsi. (*L'Innommable* 7)

Where now? Who now? When now? Unquestioning. I, say I. Unbelieving. Questions, hypotheses, call them that. Keep going, going on, call that going, call that on. Can it be that one day, off it goes on, that one day I simply stayed in, in where, instead of going out, in the old way, out to spend day and night as far away as possible, it wasn't far. Perhaps that is how it began. (*The Unnamable* 1)

In combination with the preceding and following turning points, an interesting pattern emerges: Beckett's career as a writer of prose fiction spans a period of more than fifty years; in this half century, the three major stylistic turning points occur within a comparatively limited interval of only a decade (1946–56). Finally, when identifying a fourth turning point, the analysis singles out a sample from *Comment c'est* from the French prose samples and *Worstward Ho* from the English series. Especially for the English corpus, this is a remarkable result, given the limited length of Beckett's final prose work, stressing the exceptional nature of *Worstward Ho*. One interesting result is that the segmentation for the English corpus singles out a sample from the beginning of *Watt* as the second turning point, while such a trend is absent from the French data.

DISCUSSION

From our analyses, a number of interesting trends have emerged, which might offer a valuable framework to reinspect some of the periodization issues discussed in Beckett scholarship. The VNC analysis generally

supported the periodization of Beckett's oeuvre into an early, middle, and late "cluster." In English these periods would cover, first, Beckett's early works, *More Pricks than Kicks* and *Dream of Fair to Middling Women*; second, the mid-career works, ranging from *Murphy* to *Malone Dies*; and third, a series of later works starting with *The Unnamable*. From the French prose, a similar periodization arises—although it clearly reflects the absence of any translated counterparts for *More Pricks than Kicks* and *Dream of Fair to Middling Women*. Here, we could distinguish a first period roughly coinciding with the French versions of *Murphy* and *Watt*, a middle period starting with the *Nouvelles* and, finally, the late works starting with *L'Innommable*. Interestingly, the *Nouvelles* only seemed to coincide with a minor turning point in the second period in the English prose. Especially interesting is the fact that the *Nouvelles* appear to reflect Beckett's earlier style in English, more so than in French.

The initial periodization roughly coincides with some of the periodization schemes proposed in the earlier literature, albeit on completely different grounds. An important difference, however, is that our analyses invariably point to the beginning, rather than the end, of *L'Innommable* [*The Unnamable*] as a major stylistic turning point, thus breaking up the unity of the so-called postwar "trilogy." Indeed, in terms of function word frequencies, *L'Innommable* [*The Unnamable*] fits in better with Beckett's later works than the first two novels of the trilogy, which themselves appear to be more similar to, for instance, the *Nouvelles*. As the earlier quotations indicate, the style of *L'Innommable* [*The Unnamable*] (for instance "I, say I") presages much of what Beckett will later apply even more radically in *Worstward Ho*: "On. Say on. Be said on. Somehow on. Till nohow on. Said nohow on" (81).

Apart from the VNC analyses, we have also attempted to obtain a "hard" diachronic segmentation of Beckett's oeuvre. For the French corpus, this segmentation yielded stable results, with the start of *Premier amour* and *L'Innommable* being singled out as the most significant fault lines. Additionally, *From an Abandoned Work* [*D'un ouvrage abandonné*], notwithstanding its short length—also often emerged as a major watershed in both languages. The question, however, is whether this result is to be interpreted as a turning point leading the way for Beckett's later, experimental works, or as a sort of "re-turning point," marking a temporary relapse (after *L'Innommable*) into the idiom of the *Nouvelles*. The narrative

style of the opening paragraph contrasts sharply with the opening lines of *The Unnamable* quoted earlier:

Up bright and early that day, I was young then, feeling awful, and out, mother hanging out of the window in her nightdress weeping and waving. Nice fresh morning, bright too early as so often. Feeling really awful, very violent. The sky would soon darken and rain fall and go on falling, all day, till evening. Then blue and sun again a second, then night. (*From an Abandoned Work* 57)

Debout au petit matin ce jour-là, j'étais jeune alors, dans un état, et dehors, ma mère pendue à la fenêtre en chemise de nuit pleurant et gesticulant. Beau matin frais, clair trop tôt comme si souvent, mais alors dans un état, très violent. Le ciel allait bientôt foncer et la pluie tomber et tomber toujours, toute la journée, jusqu'au soir. Puis de nouveau bleu et soleil une seconde, puis nuit. (*D'un ouvrage abandonné* 9)

This text does not seem to have played a significant role in the periodization debate so far. In fact, this work is arguably a blind spot in Beckett studies. John Pilling calls it one of Beckett's oddest "odds" and refers to the critical literature on this work as "what little there is" ("Early Beckett" 173). The only existing periodization that takes this work into account is Chiara Montini's, whose fourth phase (marked by mixed bilingualism) begins with *From an Abandoned Work*. If the emergence of *From an Abandoned Work* [*D'un ouvrage abandonné*] from the results of the stylochronometric analysis can be interpreted as a temporary stylistic relapse, it would constitute an instance of *reculer pour mieux sauter*, before the next major stylistic watershed (*Comment c'est* in the French prose, *Worstward Ho* in the English fiction). In that case, this relapse might explain why the work has been somewhat neglected in Beckett studies.

Another remarkable result is that *Worstward Ho*'s unparalleled style is also reflected in many of our results. As a work that was written in the last decade of Beckett's career, it constitutes no less than a fourth turning point, which indicates that, even when he was in his late seventies, Beckett still kept reinventing himself stylistically. As indicated previously, *Worstward Ho* is probably the only text Beckett tried but failed to translate himself and this "untranslatability" is to a large extent due to the

stylistic specificity of this text. Beckett did try to make a translation, but as James Knowlson notes:

His efforts to translate *Worstward Ho* into French soon ground to a halt. How, he asked me, do you translate even the first words of the book “On. Say on” – without losing its force? It was not until after his death that his friend, Edith Fournier, translated the book, although she had discussed it with him and he had chosen her title, *Cap au pire*, from among several that she suggested. (684–85)

In a recently discovered document among Beckett’s correspondence to his French publisher Jérôme Lindon, he appears to have translated only a few fragments of the text (Van Hulle). In these fragments, the “untranslatable” opening word is rendered in French as “plus loin,” which is both less ambiguous and less powerful than the monosyllabic “on.” Most of the sentences in *Worstward Ho* lack a verb or are otherwise elliptic:

Worse less. By no stretch more. Worse for want of better less. Less best. No. Naught best. Best worse. No. Not best worse. Naught not best worse. Less best worse. No. Least. Least best worse. Least never to be naught. Never to naught be brought. Never by naught be nulled. Unnullable least. Say that best worse. With leastening words say least best worse. For want of worser worst. Unlessenable least best worse. (94)

Due to this lack of grammatical context to reduce the potential meanings, the semantic potential of every single word proliferates. In addition, Beckett introduces several neologisms, such as “leastward,” “undimmed,” “dimmost,” “leastmost,” “unlessenable,” “unworsenable.”

Interestingly, the segmentation approach for the English corpus also singled out *The Unnamable* and *From an Abandoned Work* as fault lines, but had a harder time identifying a clear break between the early and middle period: especially *Watt* seems a good candidate for a stylistic turning point in this respect. With the methodology we applied, it turns out to be more difficult in English than in French to model the transition from a young to a middle Beckett. This result possibly reflects the fact that the original, English version of *Watt* was written relatively early (during the war), whereas its translation was made much later (in cooperation with Agnès and Ludovic Janvier) and was not published in French until 1968. In any case, this particular result

offers grounds for a reexamination of the difference in evolution between Beckett's French and English prose production, and in particular the role of *Watt* as a transitional novel.

CONCLUSION AND PERSPECTIVES

In this article, we reported the results of a quantitative analysis of the stylistic evolution of Samuel Beckett's prose writings. The stylometric analyses disrupt the neatness of the traditional three-part periodization as they show different turning points depending on the language of the corpus (*Watt* and *The Unnamable* for the English texts; *Premier amour* and *D'un ouvrage abandonné* for the French texts). In other words, in terms of stylistic developments, Beckett's English oeuvre is not identical to his French oeuvre. Our analyses suggest that the start of Beckett's "late" writings coincides with the beginning of *L'Innommable* [*The Unnamable*], thus breaking up the unity of the postwar "trilogy" in both languages. This corresponds with the genesis of these three novels, which was interrupted: between *Malone meurt* and *L'Innommable*, Beckett wrote his play *En attendant Godot*. As for the "late" prose, the unique position of *Worstward Ho* in the stylometric analyses indicates that, when Beckett was almost eighty years old, he still managed to write a work that was stylistically so innovative that it constitutes yet another turning point.

The main novelty of the stylometric analyses lies in the fact that it is restricted to the frequencies of function words, an inconspicuous category of words, seldom taken into account in traditional literary studies. Indeed, as pioneering researcher John Burrows once said: "It is a truth not generally acknowledged that, in most discussions of works of English fiction, we proceed as if a third, two-fifths, a half of our material were not really there" (*Computation* 1). The advantage of taking this neglected category into account is that it allows us to nuance and adjust existing periodizations. In this sense, the qualitative analysis of the existing periodizations and the quantitative method proposed in this article do not overlap, but complement each other. Especially the clear watershed of the *beginning*, rather than the ending, of *L'Innommable* [*The Unnamable*] is notable. The discovery of the stylistic importance of *From an Abandoned Work* calls for a reappraisal of this understudied work. Most periodizations assume a linear stylistic evolution. The stylometric analyses show that there are also forms of stylistic relapse, which disturb this linearity.

But also the more fuzzy result regarding the first turning point is important as it indicates that the decisive moment that separates the early work from the mature work is not unequivocally linked to the transition from writing in English (*Watt*) to writing prose fiction in French (the *Nouvelles*). Beckett's decision to start writing in French did play a role in the stylistic development of his prose, but the clearest turning point occurred with some delay, with *L'Innommable*, after Beckett had already written a few novels in French. After this turning point, it became more difficult to continue writing this new kind of prose, resulting in a writer's block after *Textes pour rien*. When Beckett tried to write in English again with *From an Abandoned Work*, he apparently relapsed into his "old style." Not until he had managed, with difficulty, to translate *L'Innommable* into English did he find a way to reconnect with the style of *L'Innommable* and move on with *Comment c'est*. In this sense, the stylometric analyses also suggest the important role of the act of translation and the effect of bilingual writing in Beckett's stylistic evolution. This bilingual dimension is also relevant to the unique position of *Worstward Ho* in this development. For even though it is not a bilingual work, it has a long bilingual prehistory, which most periodizations tend to disregard.

The stylochronometric analyses thus help us reassess certain critical commonplaces and prompt us to keep questioning not just our periodizations, but also our urge to periodize *tout court*. Kermode's "tick-tock" model, mentioned previously, not only applies to plots in fiction, but also to critics' attempts at periodizing an author's oeuvre. By trying to humanize time and make sense of an oeuvre, we also turn it into a plot. Especially in Beckett's case, the dangerous neatness of such a narrative with a beginning, a middle, and an end tends to obscure the continuous incompleteness of his work. That does not mean we should stop periodizing; if anything, it is an invitation to keep trying and failing. More periodizations imply more reassessments. The more nuanced the plots become, the better they may cope with the plotless continuum of his work's "mere gress."

DIRK VAN HULLE, PhD, is a professor of English Literature and director of the Centre for Manuscript Genetics at the University of Antwerp. He is the editor of *The New Cambridge Companion to Samuel Beckett* (2015), board member of the Samuel Beckett Society, and a trustee of the International James Joyce Foundation. With Mark Nixon, he is codirector of the *Beckett Digital Manuscript Project* and editor-in-chief of the *Journal of Beckett Studies*. He is the author of *Textual Awareness* (2004), *Manuscript*

Genetics: Joyce's Know-How, Beckett's Nobow (2008), *Modern Manuscripts: The Extended Mind and Creative Undoing* (2014), coauthor with Mark Nixon of *Samuel Beckett's Library* (2013) and with Shane Weller of *The Making of Samuel Beckett's L'Innommable /The Unnamable* (2014).

MIKE KESTEMONT, PhD, is an assistant professor in Digital Text Analysis at the University of Antwerp. He specializes in computational text analysis for the humanities, in particular stylometry or computational stylistics. He has published on the topic of authorship attribution in various fields, such as classics or medieval European literature. Mike actively engages in the debate surrounding the digital humanities and attempts to merge methods from artificial intelligence with traditional scholarship in the humanities. His website (www.mike-kestemont.org) contains pointers to his recent scholarly activities, including an open access scientific documentary about stylometry and Hildegard of Bingen (vimeo.com/70881172).

NOTES

Dirk Van Hulle's research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n° 313609. Mike Kestemont was partially funded for this research as a postdoctoral researcher for the Research Foundation of Flanders. The authors would like to thank Pim Verhulst for his valuable feedback and Folgert Karsdorp for kindly sharing software implementations of various methods used in this article (e.g., <https://github.com/fbkarsdorp>).

1. The term "trilogy" was used for the 1959 Olympia Press edition of the three novels *Molloy*, *Malone Dies*, and *The Unnamable* (Van Hulle and Weller 81). Beckett, however, expressed his dislike of this term at several occasions, for instance in letters to John Calder, Aidan Higgins, and Barbara Bray (Tucker 23).

2. In an earlier study (1984), Abbott situated the "key divide" in Beckett's career "between *Malone Dies* and *The Unnamable*" (*Diary Fiction* 184).

3. For the French texts, this tokenization strategy introduces a number of minor issues; for example, *aujourd'hui* will be considered as two tokens (*aujourd* and *hui*). Because we limit analyses to function words, however, this has not proven to pose a major problem. To maximize the comparability across texts, we have only considered lower-case, alphabetical character strings, ignoring for instance punctuation marks and digits.

4. For the French texts, we used the publications by *Les Éditions de Minuit*; for the English ones, mostly the publications by Faber and Faber (2009–10).

5. github.com/mikekestemont/beckett.

6. Technical comment: we scaled each column using z-scores, which is a recommended practice (Eder). For the clustering, we used Ward linkage and Euclidean distances.

7. We use Ward's linkage to construct our cluster hierarchy, a common metric (Eder) that continually monitors the overall reduction in variance in a tree when selecting which nodes to merge next.

8. For these analyses, we rely on an implementation by Folger Karsdorp, available at <https://github.com/fbkarsdorp/HACluster>.

9. They were not translated by the author himself.

10. Note that for these analyses, we lower the sample size to 1,106 words (the length of the English *Imagination Dead Imagine*), in order to be able to include all texts. We used the *librosa* library for this purpose.

WORKS CITED

- Abbott, H. Porter. *Beckett Writing Beckett: The Author in the Autograph*. Ithaca: Cornell UP, 1996. Print.
- . *Diary Fiction: Writing as Action*. Ithaca and London: Cornell UP, 1984. Print.
- Bair, Deirdre. *Samuel Beckett: A Biography*. London: Jonathan Cape, 1978. Print.
- Banfield, Ann. "Beckett's Tattered Syntax." *Representations* 84.1 (2003): 6–29. Print.
- Beckett, Samuel. *Company/Ill Seen Ill Said/Worstward Ho/Stirrings Still*. Ed. Dirk Van Hulle. London: Faber and Faber, 2009. Print.
- . *L'Innommable*. Paris: Les Editions de Minuit, 1953. Print.
- . *The Letters of Samuel Beckett*, Vol. I, 1929–1940. Eds. Martha Dow Fehsenfeld and Lois More Overbeck. Cambridge: Cambridge UP, 2009. Print.
- . *Texts for Nothing and Other Shorter Prose, 1950–1976*. Ed. Mark Nixon. London: Faber and Faber, 2010. Print.
- . *The Unnamable*. Ed. Steven Connor. London: Faber and Faber, 2010. Print.
- Binongo, José Nilo G. "Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution." *Chance* 16.2 (2003): 9–17. Print.
- Binongo, José Nilo G., and W. M. A. Smith. "The Application of Principal Components Analysis to Stylography." *Literary and Linguistic Computing* 14.4 (1999): 445–66. Print.
- Bird, Steven, Ewan Klein, and Edward Loper. *Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit*. Sebastopol: O'Reilly Media, 2009. Print.
- Bolin, John. *Beckett and the Modern Novel*. Cambridge: Cambridge UP, 2013. Print.
- Boxall, Peter. "Still Stirrings: Beckett's Prose from *Texts for Nothing* to *Stirrings Still*." *The New Cambridge Companion to Samuel Beckett*. Ed. Dirk van Hulle. Cambridge: Cambridge UP, 2015. Print.
- Brooke, Julian, Adam Hammond, and Graeme Hirst. "Unsupervised Stylistic Segmentation of Poetry with Change Curves and Extrinsic Features." *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*. Eds. David Elson, Anna Kazantseva, Rada Mihalcea, and Stan Szpakowicz. Montreal: ACL, 2012. Web. 4 March 2016.

- Burrows, J. F. *Computation into Criticism: A Study of Jane Austen's Novels*. Oxford: Oxford UP, 1987. Print.
- . "Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17.3 (2002): 267–87. Print.
- . "A Second Opinion on 'Shakespeare and Authorship Studies in the Twenty-First Century.'" *Shakespeare Quarterly* 63.3 (2012): 355–92. Print.
- Byrne, Paula. "Young Jane." *Times Literary Supplement* 20 February 2015: 9–10. Print.
- Clément, Bruno. *L'Œuvre sans qualités: Rhétorique de Samuel Beckett* [*The Work without Qualities: Samuel Beckett's Rhetoric*]. Paris: Seuil, 1994. Print.
- Daelemans, Walter. "Explanation in Computational Stylometry." *Computational Linguistics and Intelligent Text Processing, 14th International Conference*. Ed. Alexander Gelbukh, 2013. Heidelberg: Springer, 2013. Print.
- Eder, Maciej. "Computational Stylistics and Biblical Translation: How Reliable Can a Dendrogram Be?" *The Translator and the Computer*. Eds. Tadeus Piotrowski and Łukasz Grabowski. Wrocław: WSF Press, 2013. Print.
- Federman, Raymond, and John Fletcher. *Samuel Beckett: His Works and His Critics. An Essay in Bibliography*. Berkeley and London: U of California P, 1970. Print.
- Forsyth, R. S. "Stylochronometry with Substrings, or: A Poet Young and Old." *Literary and Linguistic Computing* 14.4 (1999): 467–78. Print.
- Gontarski, Stanley E. "Staging Himself, or Beckett's Late Style in the Theatre." *Samuel Beckett Today/Aujourd'hui* 6 (1997): 87–97. Print.
- Gries, Stefan Th., and Martin Hilpert. "Variability-Based Neighbor Clustering: A Bottom-up Approach to Periodization in Historical Linguistics." *The Oxford Handbook of the History of English*. Eds. Terttu Nevalainen and Elizabeth Closs Traugott. Oxford: Oxford UP, 2012. Print.
- Holmes, David I. "The Evolution of Stylometry in Humanities Scholarship." *Literary and Linguistic Computing* 13.3 (1998): 111–17. Print.
- Hoover, David L. "A Conversation Among Himself. Change and the Styles of Henry James." *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*. Eds. David L. Hoover, Jonathan Culpeper, and Kieran O'Halloran. New York: Routledge, 2014. Print.
- . "Multivariate Analysis and the Study of Style Variation." *Literary and Linguistic Computing* 18.4 (2003): 341–60. Print.
- Jannidis, Fotis, and Gerhard Lauer. "Burrows's Delta and Its Use in German Literary History." *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*. Eds. Matt Erlin and Lynne Tatlock. Rochester: Camden House, 2014. Print.
- Juola, Patrick. "Authorship Attribution." *Foundations and Trends in Information Retrieval* 1.3 (2008): 233–34. Web. 1 March 2016.
- . "Becoming Jack London." *Journal of Quantitative Linguistics* 14.2 (2007): 145–47. Print.
- Kermode, Frank. *The Sense of an Ending: Studies in the Theory of Fiction*. Oxford: Oxford UP, 1967. Print.
- Kestemont, Mike. "Function Words in Authorship Attribution. From Black Magic to Theory?" *Proceedings of the Third Computational Linguistics for Literature Workshop*.

- Eds. Anna Feldman, Anna Kazantseva, and Stan Szpakowicz. Gothenburg: ACL, 2014. Web. 4 March 2016.
- Knowlson, James. *Damned to Fame: The Life of Samuel Beckett*. London: Bloomsbury, 1996. Print.
- Koppel, Moshe, Jonathan Schler, and Shlomo Argamon. "Computational Methods in Authorship Attribution." *Journal of the American Society for Information Science and Technology* 60.1 (2008): 9–26. Print.
- Love, Harold. *Attributing Authorship: An Introduction*. Cambridge: Cambridge UP, 2002. Print.
- Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge: Cambridge UP, 2008. Print.
- McHale, Brian. *Constructing Postmodernism*. London: Routledge, 1992. Print.
- Montini, Chiara. *La Bataille du soliloque: Genèse de la poésie bilingue de Samuel Beckett* [*The Battle of Soliloquy: The Genesis of Samuel Beckett's Bilingual Poetics*]. Amsterdam: Rodopi, 2007. Print.
- . "Sinking in the Mud: *From an Abandoned Work* et le difficile retour à l'anglais." *Samuel Beckett Today/Aujourd'hui* 20 (2008): 63–69. Print.
- Mooney, Sinéad. *A Tongue Not Mine: Beckett and Translation*. Oxford: Oxford UP, 2011. Print.
- Mosteller, Frederick, and David L. Wallace. *Inference and Disputed Authorship: The Federalist*. Reading: Addison-Wesley, 1964. Print.
- O'Toole, Fintan. "Samuel Beckett: The Private Voice." *The New York Review of Books* 19 March 2015: 34–36. Print.
- Pennebaker, James. *The Secret Life of Pronouns: What Our Words Say about Us*. London: Bloomsbury, 2011. Print.
- Pilling, John. "Early Beckett: The One Looking Through His Fingers." *The New Cambridge Companion to Samuel Beckett*. Ed. Dirk van Hulle. Cambridge: Cambridge UP, 2015. Print.
- . "From an Abandoned Work: All the Variants of the One." *Samuel Beckett Today/Aujourd'hui* 18 (2007): 173–83. Print.
- Rybicki, Jan, and Maciej Eder. "Deeper Delta across Genres and Languages: Do We Really Need the Most Frequent Words?" *Literary and Linguistic Computing* 26.3 (2011): 315–21. Print.
- Sardin, Pascale. *Samuel Beckett auto-traducteur ou l'art de l'empêchement: Lecture bilingue et génétique des textes courts auto-traduits (1946–1980)* [*Samuel Beckett, Self-translator; or: The Art of Impediment: A Bilingual and Genetic Reading of the Short Self-translated Texts (1946–1980)*]. Artois: Artois Presses Université, 2002. Print.
- Schreibman, Susan, and Ray Siemens. *A Companion to Digital Literary Studies*. Oxford: Blackwell, 2004. Print.
- Stamatatos, Efstathios. "A Survey of Modern Authorship Attribution Methods." *Journal of the American Society for Information Science and Technology* 60.3 (2009): 538–56. Print.

- Stamou, Constantina. "Stylochronometry: Stylistic Development, Sequence of Composition, and Relative Dating." *Literary and Linguistic Computing* 23.2 (2008): 181–99. Print.
- Tilahun, Gelila, Andrey Feuerverger, and Michael Gervers. "Dating Medieval English Charters." *The Annals of Applied Statistics* 6.4 (2012): 1615–40. Print.
- Tucker, David. "Introduction: 'No One Wanders Unpunished': Revisiting *Molloy*, *Malone meurt*/*Malone Dies* and *L'Innommable*/*The Unnamable*." *Samuel Beckett Today/Aujourd'hui* 27 (2014): 11–24. Print.
- Van Hulle, Dirk. "Translation and Genetic Criticism: Genetic and Editorial Approaches to the 'Untranslatable' in Joyce and Beckett." *Linguistica Antverpiensia, New Series: Themes in Translation Studies* 14 (2015): 47–60. Print.
- Van Hulle, Dirk, and Pim Verhulst. "Introduction: A Beckett Continuum and a Chronology of Beckett's Writings." *The New Cambridge Companion to Samuel Beckett*. Ed. Dirk van Hulle. Cambridge: Cambridge UP, 2015. Print.
- Van Hulle, Dirk, and Shane Weller. *The Making of Samuel Beckett's L'Innommable/The Unnamable*. London: Bloomsbury Academic, 2014. Print.