

Name: Pascale Feldkamp Moreira  
Student number: 5062280  
Thesis Supervisor: Kiene Brillenburg Wurth  
Second Assessor: Olivia Fialho

# Writing with the left hand

## Reading(s) of Bilingual Authors Style(s)

### Acknowledgements

This thesis would not have been possible without the support and guidance of both teachers, supervisor, thesis-group, peers, friends, and family. A special thank you are due to Pedro Moreira and Donart Simnica for baby-sitting, proofreading, and listening.

## Abstract

It is a common belief that the mother-tongue is the only language in which one may properly think, feel, and express oneself. In this view, second language (L2) writing is often considered second-rate writing. One line of argumentation, often adopted by L2 writers themselves, suggests that literary “style” is the crowning achievement of language-mastery, and so reserved for the native speaker. Yet, systematic studies of the features of L2 style have been scarce. That is why I seek to answer the question: what does literary L2 writing look like stylistically?

To answer this question, I examine the L2 writing style of two authors who wrote intermittently in L1 and L2: Samuel Beckett and of Jhumpa Lahiri. Combining distant- and close reading, I compare L1 and L2 style by use of function words, vocabulary, literary devices, tropes, and language-use as well as lexical diversity and sentence-length with the aim to assess the importance of multilingual competencies for literary writing. Moreover, I examine whether acquiring and writing in L2 has an impact on later L1 style. I show that L2 writing is different from L1 writing, that it tends to be more personal and introspective, employs significantly more first-person narrative perspective, is lexically more diverse than L1 writing, and uses language in a novel and idiosyncratic way, bearing witness to an engagement with L2 textual constructions of meaning. These findings are relevant not only for rethinking the status of L2 writing, but also for evaluating the importance of multilingual competences in, for example, education.

## Table of contents

1. Writing with the left hand: Introduction .....	5
1.1. Literary style(s) and L2 writing .....	13
1.2. Research design and methods .....	17
2. How is L2 writing different? .....	20
2.1. The deficiency idea .....	20
2.2. The linguistic consciousness idea .....	22
2.3. L2 writing between deficiency and linguistic consciousness .....	25
2.4. The language specificity idea.....	26
2.5. The L2 specificity idea.....	29
3. Reading L2 style: Methodology .....	31
4. Jhumpa Lahiri: Analysis .....	40
4.1. Distant reading .....	42
5. Samuel Beckett: Analysis .....	56
5.1. Distant reading .....	59
5.1. Close reading .....	72
“The End”/“La Fin” .....	73
<i>Fizzles/Foirades</i> .....	80
6. Two authors’ left-hand style: Discussion.....	90
7. L2 style: Conclusions.....	96

# 1 Writing with the left hand

## Introduction

It's as if I were writing with my left hand, my weak hand, the one I'm not supposed to write with. It seems a transgression, a rebellion, an act of stupidity. (Lahiri, *Other Words* 55)

That is how the multilingual author Jhumpa Lahiri describes writing in Italian, which is not her native language, but one of her literary languages. Most writers, if not all writers, are bi- or multilingual,<sup>1</sup> though most writers only use their first language for literary composition.<sup>2</sup> It is estimated that more than half of the world population is bi- or multilingual, and many speak and write in another language than their first at school, at work, in public or at home (Grosjean 13). It seems, then, rather strange that more writers do not write in another language than their first. Elsewhere, an increasing number of people compose non-literary, especially scientific texts, in acquired languages (Hamel). Still, it is a common belief that *literary* composition requires a “language-mastery” and “style”, which only a native speaker can attain (Kellman, *Translingual* ix). Consequently, second language (L2) writing is often construed as second-rate writing. It is in that context that Lahiri considers her L2 writing a transgression, rebellion, or even an act of stupidity.

Yet, considering the history of European literature, writerly monolingualism might be considered a brief intermission in an otherwise multilingual literary tradition (Yildiz 4; Coulmas 28). Using an acquired language (Latin) as a literary language was a longstanding European norm, and Persian once had a similar role in the Balkans, West-, South- and central Asia, as did Classical Chinese in east-Asia (文言, lit. “text speak”)(Coulmas 28). Yasemin Yildiz, who studies multilingualism in literature, argues that the reason we do not find writerly monolingualism strange today is because multilingualism has been framed as the anomaly in a “monolingual paradigm” that is historically quite recent (4). It is a persistent

---

<sup>1</sup> See Kellman’s efforts to name one monolingual author in *Nimble*, pp. 8-9.

<sup>2</sup> I follow Grosjean in using the term multilingual to describe a person who speaks more than one language regularly, irrespective of their level of proficiency (20). This subsumes a bilingual, one who speaks two languages regularly.

paradigm, in which writing in another language was (and is) thought to impair literary expression, style, and authenticity, and where the “mother tongue” was seen as the only language in which one was thought to properly “think, feel and express oneself” (Kellman, *Translingual* ix; Yildiz 7-10). We see this idea of the L1 as the “proper” literary language reflected in the insecurity some bi- or multilingual authors express when describing their L2 writing. As for Lahiri, in Italian she claims she is “writing with [the] left hand”, and lacks the “true mastery” that is required of the literary author over their language (Lahiri, *In Other Words* 55, 83). Similarly, Samuel Beckett stated that writing in his L2 French was a way to “impoverish” himself as a writer and to write “without style” (Kager 72; Forster 87). Yet what is this “true mastery” of a language? Or rather, if we ask this in an examinable, backward way: what does “writing with the left hand” or “poverty” of L2 style *look like*? That is the central question of the present study, where I examine L2 writing style by way of two case-studies: Beckett and Lahiri, comparing their L1 and L2 literary style.

L2 writing is a topic at the intersection of literary studies, second-language acquisition (SLA) research, cognitive linguistics and linguistics proper, several perspectives that each frame the phenomenon of L2 writing in their own disciplinary language and examine it by different methods.<sup>3</sup> Therefore, methodology plays a prominent role in the present study, not least because when studying language, methodology is essentially a way of using a (disciplinary) language to examine language. The inquiry into L2 writing is therefore a question *about* language, as much as it is a question *of* language. It should become clear in the following both how differently the phenomenon of L2 writing is construed across disciplines, and how a careful combination of methods and perspectives can help us understand it better stylistically. In the present chapter, I outline conceptions of L2 use in general, and L2 writing in particular before turning to the concept of style and how to examine it.

Pace the “anomaly” of L2 writing, many authors have written in their L2. One *selective* bibliography of L2 writers which only considers works by authors who either mix English with another language or have English as their L2, spans an impressive 137 names for *creative fiction only* (Kellman and Lvovich). Some of the most famous examples of L2 writers are Joseph Conrad and Vladimir Nabokov (not to mention the much earlier Apuleius,

---

<sup>3</sup> I use the term “L2 writing” which is standard in linguistics, while terms like “exophonic” or “translingual” writing are common in the literary field. The latter terms are not used consistently to describe the same phenomenon. By using the term “L2 writing”, I denote the act of writing in another language than L1, regardless of the order of acquisition of that language and regardless of whether that speaker is a literary author or not.

Hafez, and Rumi). Yet their example has generally been framed as an outstanding and exceptional achievement. In SLA research, such exceptional L2 speakers are even compared to Olympic athletes or opera singers (as in Cook 191). Additionally, a monolingual paradigm in the field of linguistics has meant that “anomalous” bilinguals were generally avoided in language research (Pavlenko, *Emotions* 4). The variance in proficiency among bilinguals was seen as the central problem: There is no consensus about the “ideal” or “normal” bilingual, which is why bilingual language-use, including L2 writing, has been generally under-explored (Ibid.).

This exceptionalism of L2 writers today is closely connected to prevalent, normative ideas about first and second languages: On one hand, a widespread popular belief that “younger-is-better” for L2 learning, and, on the other hand, a broad acceptance of the critical period hypothesis in SLA research. According to this hypothesis, native-speaker-like mastery of an acquired language is near impossible after a certain age (often set around puberty). Therefore, especially in SLA research, L2 use was (and is often) seen as a deficient version of L1 use, to which it may but aspire (Hanauer, ‘Appreciating’ 11). Studies throughout the 70s and 80s has confirmed (and assumed) as much when matching L2 speakers against native speaker, looking at, for example, L2 speakers grammar (Johnson and Newport), pronunciation (Asher and García), accent and phonology (Oyama, ‘A Sensitive Period’; Seliger et al.), and language comprehension (Oyama, ‘Sensitive Period and Comprehension’). Yet, Piller argues that measuring L2 competence has been skewed by a phono-syntactic bias, whereby research has looked at only few aspects of language-use (Piller 182-183). As such a broad consensus about L2 pro- or deficiency, has only more recently been broadly questioned (Pavlenko, *Emotions*; Piller). Today, it is becoming increasingly more recognised that “passing for” or even surpassing a native speaker is not a superhuman feat. Language competence, in L1 and L2 speakers alike, is instead seen as dynamic and unequal across areas of language-use, i.e., a speaker may be more proficient in writing than in speaking (Pavlenko, *Emotions* 10). Language, in this view, is not considered as an object to be either acquired or “mastered” *tout court*, but as a set of skills which evolve through practice for L1 and L2 speakers alike.

L2 writing as a subfield of L2 use has been marked by a similar bias, that is, L2 writing has been generally understood as deficient L1 writing. Moreover, due to bilingual exceptionalism, L2 writing was not studied systematically in its own right, especially as a creative endeavour. Generally, it is said that L2 writing displays a simpler vocabulary, less lexical diversity and grammatical complexity, and contains more calques (Birdsong; Oyama,

‘Sensitive Period and Comprehension’).<sup>4</sup> Yet Crossley and McNamara note that, for example, a lower lexical diversity of L2 writing has been but an assumption (120). More recently, there has been a greater focus on features of L2 writing in the context of SLA research with regard to evaluating and assessing skill-specific proficiency (Piller 183). One line of inquiry in SLA research has also encouraged the practice of creative writing in the SLA-classroom, with the development of an L2 “voice” as a main goal (Disney). There are also scholars in the SLA context such as David Hanauer, who seek to study L2 writing in its own right; not only compare L2 writing to a native norm. Hanauer’s focus is above all on the innovation of language in L2 writing, but also on the process of L2 learning as a situation brimming with potential interaction between cultures and languages that further new ways of thinking and expressing (Hanauer, “Appreciating” 4). Therefore, Hanauer contends, L2 writing “is beautiful, aesthetically pleasing, innovative and designed to surprise” (“Appreciating” 2). Yet, while developing “style” or voice has become more important in language teaching, it is still rare to find studies that look for it and examine at L2 writing in terms of aesthetic appreciation (Hanauer, “Appreciating” 11).

Also in the literary field, Steven Kellman, who published a pioneer study of the subject of L2 writing (2014) seems to regard the L2 as a sort of handicap for literary composition. Writing about Lahiri, he argues that her Italian is “purged of all the nuances and felicities that Lahiri could count on in the use of English” (Kellman, *Nimble Tongues*, 132). Moreover, studies of L2 writers in the literary field have tended to focus on the political and personal motivations for L2 writing (Coulmas; Kellman, *Translingual*; Kremnitz). That is, on reasons why authors choose the language they do for writing, and not the way authors subsequently *use* that language. Now, particularly in the literary field, there has been little work on the *de facto* stylistic characteristics of L2 writing, which would look at the L2 as a resource for literary composition. That is what the present study aims to do. While prior research within literary studies has focused on writers’ motivations for changing language, I intend to focus on what actually comes about when they do, that is, style *per se*.

Yet, in asking whether language of composition may be a factor for style, I do not intend to only look at the ways L2 style may fall short of L1 style. Rather, I seek to examine whether writing in L2 can be considered an advantage rather than an inhibition to literary style, with the aim to assess the importance of multilingual competences for literary writing.

---

<sup>4</sup> Calques are word-for-word translations transposed from one language to the other, such as “earworm” from the German *Ohrwurm*.

In the following section I outline approaches to L2 use and L2 writing that take the advantage perspective on bilingualism and L2 writing. Ideas about L2 writing, both from a deficiency point of view and an advantage point of view, will be the different perspectives from where to view L2 writing.

Concurrent with the development away from bilingual exceptionalism in linguistics, there has been interest in possible advantages of bilingualism, as well as in the relation between bilingualism and creativity in creativity studies. A string of studies match bilinguals against monolinguals in a variety of tasks (Blom et al.; Kharkhurin; Leikin et al.). Among “the amazing benefits of being bilingual” are listed (i.a.): forestalling dementia, higher creativity, better executive functioning and selective attention (Blom et al.; Kharkhurin; Vince). Yet often, the danger in the line of inquiry on bilingual advantage is assuming bilingual exceptionalism in another flavour. In these studies, the problem is not so much that the “normal” bilingual does not exist, but that the “normal” *monolingual* is just as rare. Yet the homogeneity of monolinguals, that merely constitute the “control group”, is hardly ever problematised. Even monolinguals who only speak one standardised language may know geographical and/or generational variations. Moreover, boundaries between languages are blurry, but so too are those between a language and a dialect, particularly when looking beyond a European context. Especially in post-colonial space, boundaries between language are more commonly seen and treated as permeable and mutable (Bennett and Barros). Really, language boundaries are ontologically blurry altogether. It should be remembered that standardised languages are artificial constructions historically engendered in the project of European nation-building (Yildiz). What Yildiz calls attention to when speaking of a “monolingual paradigm”, is the historical normalisation and forgetting of this circumstance.

There are attempts in various fields to move past the monolingual paradigm, standard-language ideology, as well as bi- and multilingual exceptionalism. In his seminal *The Monolingualism of the Other*, Jacques Derrida argues that a speaker is never monolingual, because languages are neither pure nor distinct (8). Rather than seeing language as a transparent medium of communication that a speaker uses to freely express an (extra-linguistic) idea, Derrida argues that we are preceded by and entrenched in language and cannot therefore relate to it from a position of “mastery” outside of it (Ibid.). To “master” a language, also in the sense Lahiri described it, would presuppose both the unity and homogeneity of a language as well as an extra-linguistic position from where to “master” it. If there is no extra-linguistic position, that would imply that we might see the “bilingual advantage” as a quality of any one language itself, and not as a skill that “acquiring” or

“mastering” a language might bestow on the speaker. Following the idea of Derrida, Dijk et al. argue that bilingualism disposes toward creativity precisely because bilinguals are entrenched in more than one linguistic system, and that it is the qualities of languages that dispose bilingual to being creative. Seeing language from an embodied view, they argue that linguistic systems are not only linked to meaning, but to culturally specific, sensori-motor experiences, and to ways of experiencing the world (9). For Dijk et al., bilinguals thus draw on and inhabit multiple linguistic, experiential and conceptual systems.

There are L2 writers who feel like they live in different worlds and embody multiple selves and therefore claim, like Dijk et al., that bilingualism is a resource rather than an impairment to literary composition. Rosario Ferré who writes both in English and Spanish, contends that “a bilingual writer is really two different writers, has two very different voices, writes in two different styles, and, most important, looks at the world through two different sets of glasses” (Ferré 138). Lahiri even speaks of language as an entryway, noting that she can “access” a different reality in Italian for exploring “ideas, feelings, impressions” (Lahiri "Interview" 23:00-23:55). If it is true that language is like a gateway unlocking a different writer, voice, style, and view on the world, that may be why some writers choose to write in L2.

This view has been underpinned by more recent research on cross-linguistic cognitive variations of bilinguals, especially in psychology and cognitive linguistics, not least prompted by bilinguals’ own accounts (Besemeres). Identity, cognition, memory, and affect have been found to vary in bilinguals, in such a way that we may speak of language-specific identities (Marian and Kaushanskaya), memories (Aragno and Schlachet), cognition (Fausey and Boroditsky), and emotions or attitudes of bi- and multilinguals (Byford; Ervin; Pavlenko, *Emotions*). In continuation of this line of inquiry, approaching L2 writing from the standpoint of linguistic relativity and language embodiment, I seek to test whether we may also speak of language-specific literary style.

Yet, there are other perspectives on bilingualism advantage. In literary studies, one tradition sees bilingualism as a resource for literary writing more generally, whether the author writes in their L2 or L1. In *The Dialogic Imagination*, Bakhtin noted that Roman literary imagination benefitted from Latin/Greek bilingualism, because it introduced them to cultural and linguistic otherness (Bakhtin 61). For Bakhtin, Roman authors were able break to free from any one set view of the world, because “the Latin literary word viewed itself *through the eyes* of the Greek word” (Bakhtin 61, italics in original). By in this way defamiliarizing their own culture and language, Bakhtin argued, bilinguals attain a “linguistic

consciousness” more perceptive of language (61). Following Bakhtin, Kellman argues that linguistic consciousness shows in L2 writing, in the way L2 texts are “metalingual in their self-consciousness of their own linguistic medium ... they make language itself strange” (Kellman *Nimble* 12). Like Hanauer, Kellman finds that L2 writers are unconventional and surprising, defamiliarizing language through language games, puns, neologisms, and by emphasising the polysemy of words (Ibid.).

In this light, it seems that the idea of a raised linguistic consciousness in the writing of bilingual authors directly contradicts the idea suggested in the linguistics and SLA context, that L2 writing is a deficient form of L1 writing. Because of a higher awareness of language, L2 writers may pay closer attention to their use of words and grammar. In continuation of Bakhtin and Kellman’s ideas, I examine whether bilingual authors may display such “linguistic consciousness”.

I adopt these three overarching ideas about L2 writing from the different fields to test in my analysis. I call these *the deficiency idea*, *the specificity idea*, and *the linguistic consciousness idea*, which will be further discussed in the theory chapter, where I formulate hypotheses on their basis to be tested in the analysis. In testing these, the present study seeks to explore how and whether L2 writing is different from L1 writing, and how writing in an L2 may shift an authors’ style. It does so by asking: Are L1 and L2 writing stylistically distinguishable, and, if so, what are the distinguishing features? Three sub-questions suggest themselves based on the three overarching ideas outlined. Firstly: Is L2 writing lexically and syntactically simpler, so that we may understand L2 writing as deficient L1 writing? Secondly: Is L1 and L2 texts different in a way that we may speak of language-specific styles? And finally: Does writing in L2 have an impact on L1 literary style, so that writing in L2 alters an author’s way of writing more generally?

While there is yet a need to study the style of L2 writing from a literary perspective, that does not mean that literary scholars have not thought about L2 writing. Various scholars have studied the role of language for L2 writers (Arndt et al.; Forster; Helgesson and Kullberg; Kellman, *Translingual*; Wilson); especially for Beckett there is a large scholarship (i.a., Beer; Bolin; Germoni and Sardin; Hulle and Kestemont; Kager; Rabinovitz; Slote). There are also various interviews and essays by L2 authors, including Beckett and Lahiri (Kellman, *Switching*; Lahiri, *Other Words*; Lahiri, ‘I Am in Italian’; Forster). I draw on this secondary literature, aiming to examine how well more overarching ideas about L2 writing describe actual sentence-level stylistic features.

Still, there has been no structured method for studying language-specific literary style from the literary perspective. Kellman offers close readings of L2 texts but does not study characteristics of L2 style as a general phenomenon (see Kellman 132). To understand style in L2 writing more broadly, Hanauer contends that we need to combine linguistics and SLA methods with a literary perspective (Hanauer, "Appreciating" 11). I argue that a careful combination of methods and perspectives, looking at the multilingual phenomenon of L2 writing by way of multiple (disciplinary) languages, offers a deeper understanding of L2 writing style and a way to study it systematically. Therefore, I combine a literary close reading in the tradition of stylistics with stylometric distant reading and with quantitative measures of style from a traditional SLA context.

As this study is situated at the intersection of various fields, it engages in multiple debates and furthers a dialogue between them. Firstly, in the context of literary studies, this study seeks to complement literary approaches to L2 writing by focusing on style with a computational approach. I suggest that the literary field benefits greatly from a closer dialogue with linguistics and computational methodologies. Thus, this study aspires to bridge a gap between the fields of literature and linguistics on the topic of L2 writing.

Secondly, this study engages in the debate on multilingual proficiency and language acquisition after the critical age in the context linguistics and SLA research. It seeks to contribute to the debate by examining whether L2 writing can be seen as deficient L1 writing, but also by focusing on *differences* between L1 and L2 writing. It thus complements a linguistics and SLA approach to L2 writing by looking at literariness and style.

Lastly, in the context of stylometry, this study seeks to add to the methodological debate on close- and distant reading by considering the gap between approaches to and notions of style. Here, I seek to explore whether combining stylistics and stylometry is productive for examining L2 writing.

Apart from the relevance of the present study to the specific debates, assessing the importance of multilingual competencies for writing is more broadly. As has been noted, studying and rethinking the status of L2 writing is important for going beyond a monolingual paradigm and the bilingual exceptionalism linked to it. Moreover, assessing the importance of multilingual competences for writing is also relevant for evaluating the status and space we give L2 learning in an educational system that seeks to enable literacy, linguistic awareness and creativity, and in society at large.

As the terms stylistics and stylometry suggest, these methods examine style. An investigation of style calls for a clarification of this nebulous term. Therefore, in the

following section, I begin by discussing the term style in different traditions and how to study style in practice, before moving on to a brief outline of research design and methodology.

## 1 .1. Literary style(s) and L2 writing

The scarcity of stylistic studies of translingual writers in literary studies is variously related to the problematic nature of the term itself. The term “style” has been used to denote the characteristics a particular work, an individual, a group, a school, a time or place; but style has also been taken to be a holistic impression of texts (Baldick, ‘Style’; Dolven). As Dolven notes, the versatility of the term is “a challenge for theory”, but also what makes it an apt instrument in cultural analysis and popular discourse alike (1369).

Style was first treated as an aspect of rhetoric by Aristotle and Cicero. Here, style or decorum were thought to consist of formal features (syntax, diction, and figuration), which distinguished a good rhetorician (Dolven 1369). The reason why Beckett and Lahiri can say they write without style in L2, is because style was in derivation of this tradition perceived as a marker of language-mastery (Dolven 1369). That is, you either have it or you don’t. However, the concept of style in the tradition of stylistics is more generous. Stylistics has been above all focused not on *what* a text means, but *how* it means (Leech and Short 11). As such, stylistic analyses have focused on the formal aspects, literary devices, and language, that account for the “literariness” of texts, even if “there is little agreement as to exactly what these are” (Attridge, *Singularity* 10-11). Moreover, in computational stylistics as a supposed continuation of stylistics proper by other (computational) means, the idea about style is even wider. Recently, Herrmann et al. have defined style as “a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively” (Herrmann et al. 44). In this sense, all texts have a “style”: there are always a certain ensemble of observable formal features in any text. The advantage of Herrmann et al.’s definition is that style may belong to either author, group of authors or indeed a genre or period, and that style may be studied both qualitatively and quantitatively.

A second reason why the study of stylistic studies of L2 writing have been scarce in literary studies is because approaches that rely on close reading, have come up short in looking at style across multiple authors’ *oeuvres*, simply because of scope. Traditional stylistics have focused on particular works or schools, or periods and events in an individual

authors' life (Halliday 1971; Chatman). Going beyond local studies of style would mean close reading and comparing across a very large corpus. In other words, the problem of studying and comparing authorial styles is very much a methodological problem. Pioneers like Maria Tarlinskaja and Willie van Peer undertook large-scale, quantitative stylistic analysis even before the computer (1987; 1986), yet computational methods facilitate quantitative and large-scale literary studies in a new way.

Today we may speak of a (re)turn to stylistics by computational methods. Among more recent studies in quantitative stylistics are large-scale studies of stylistic variation across an authors' *oeuvre* (chronostylometry), or across genre and mode of composition (Hulle and Kestemont; Hoover "Modes"; Haverals et al.). My study should be seen in the context of this computational approach to style, exploring language of composition as a factor for stylistic variation. Importantly, I set the "re" in "return to stylistics" in parenthesis, because the computational approach entails a new methodology and therefore a different conception of the object "style" itself. Stylistics first linked the study of literature and of language (Zyngier 227). In continuation, Russian formalists such as Roman Jakobson in the early 20<sup>th</sup> century fortified that link by turning away from the author and toward language and text (Dolven). Stylistics in this tradition examines a particular language-use, the "literariness" of language, or what Attridge calls a "peculiar language", with a focus on inventive language, formal features, literary devices and tropes (Attridge, *Peculiar 2*). The application of statistics and computational methods in more recent computational approaches to style are even more language-oriented, studying style through a statistics lens, and often looking at the most frequent words (MFW)(Zyngier 227, Craig).

Curiously, this focus on language does not, however, coincide with a radical turn away from the author (Craig). Actually, the branch of quantitative stylistics which Fialho and Zyngier call stylostistics or stylometry is geared toward examining a "stylistic profile" or authorial style (Fialho and Zyngier 330; Craig). Stylometry has come to study an authors' "distinctive identity in language", which it can effectively identify, according to Craig, thanks to a happy marriage between language, a finite set of linguistic features, and statistics methodology, apt for detecting small numerical variations (Craig). Thus, Craig argues that stylometry "is not a crude, tone-deaf approximation to the *literary* but a method in harmony with the fundamental principles of *language*" (Craig, my italics). True to the close connection between language and the literary in stylistics, Craig seems to equate the terms here. Certainly, it may be that stylometry is sensitive to the variations in language-use. However, it should be noted that stylometry – often based on simple wordcounts – overlooks a great deal

of “the literary” considered in traditional stylistics. It has little regard for polysemy or figurative language, and classic literary devices such as irony, metaphor, tropes and formal features, which it is (yet) unable to process adequately (Boot and Naber 54; Herrmann et al. 46). While the advantage of stylometry is that it may “reduce the nebulous phenomenon of style into a number of quantifiable features” (Choiński et al. 360) – it is still a reduction.

In looking for authorial style, stylometric studies study general features observable across individual works. A common selection of features has been most frequent words, predominantly function words (because they are very frequent), such as prepositions, conjunctions, articles, etc. The idiosyncratic use of function words has been commonly referred to as an authors’ “stylome” or the author’s textual fingerprint (Daelemans 451). This stylome is genre and content-independent to a certain degree and supposedly used unconsciously, wherefore it is considered less easily modified by the author or adopted by imitators (Halteren et al. 66; Kestemont et al. 206). Looking at the stylome has proven effective in practical applications, such as in forensics and in digital security, for example in identifying authors of computer viruses (Fialho and Zyngier 330). However, it has been shown that the idea of a stylome or textual fingerprint is not fully tenable; it is not essential to an author but may change, for example with age (Hirst and Wei Feng), and changes in physical and mental health (Pennebaker, *The Secret Life of Pronouns* 25). Hulle and Kestemont’s study of Beckett also seems to suggest that style is not constant across an authors’ languages (197) – a question that the present study will contribute to exploring further.

Yet, as Eder points out, there is no consensus on the operationalisation of the concept style or stylome, nor on which or how many features to examine (Eder, ‘Style-Markers’ 105; Stamatatos 539). Many recent studies distinguish between authors taking a very narrow vision of style or stylome, looking at only the 100 MFW (Eder, ‘Style-Markers’ 105).

Still, however we tweak the parameters, computational stylometry will not substitute a human close reading in the tradition of stylistics, because it examines a different object altogether. Variations in frequency among the MFW is not generally considered an author’s literary style – or is at least a minimal definition. That does not mean that computational stylometry is wholly pointless endeavour. In fact, there is a sense in which its shortcomings are also its greatest advantage. While the idea of author’s having a textual fingerprint may be questioned, the usage of the most frequent function words are clearly still the most invisible and forgettable elements of an authors’ “style”. As Pennebaker notes, function words are a mere 0,04 percent of the average English-speaker’s total vocabulary, yet if we look at only

the *twenty* most frequent function words, these amount to a stunning 30 percent of any text (Pennebaker, *Secret* 25–27). Function words are typically short and hard to perceive. They are the “forgettable” words in a sentence, the words a human reader is bound to overlook (Pennebaker, ‘Forgettable Words’). We generally pay attention to content over function words, which are processed differently in the brain (Pennebaker, *Secret* 29). In fact, one may, in particular cases of brain-damage, lose the ability to use function words, but still be able to use content words, and *vice versa* (Ibid.). Literary scholars are not exempt to unseeing function words. As Burrows notes at the outset of his seminal book *Computation into Criticism*, when discussing literature, we usually “proceed as if a third, two-fifths, a half of our material were not really there” (Burrows 1). Thus, computational stylometry can be a powerful complement to stylistic analysis because it looks at that in style which is most often overlooked in close reading. That is why I take a minimal view of style in the following analyses by computational stylometry, focusing on the topmost frequent function words and features.

Each method explores a different aspect of a “style” here, which is also why combining methods may help us gain a deeper understanding of L2 style. A computational approach combines the large-scale with the minuscule; it helps to process large amounts of data and find general patterns *on the sentence-level* of L2 and L1 writing. When referring to style in the sense of a computational stylometry in this study, “minimal style” is used, while when referring to the notion of style in traditional stylistics, “expanded style” is used.

Yet it is by no means *only* possible (nor preferable) to study style quantitatively. Literary studies boast a long tradition of formal and stylistic analysis that has fruitfully relied on close reading with a focus on formal features and literary devices. It may even be argued that a literary studies that forfeits formal analysis misses its object, if the “literary” in literature is understood as the use of a “peculiar language” (Attridge, *Peculiar Language*). A close reading that pays attention to the literariness of L2 writing is particularly apt in this case, considering the inventiveness that i.a. Hanauer claims for L2 writing. In traditional stylistics, Attridge stresses the importance of looking at the inventiveness, uniqueness, and strangeness of language in what he sees as the singularity of the literary, that effects an “experience of increased possibilities for thought and feeling” (Attridge, *Singularity* 76). Insofar as the literariness of literature holds this kind of strangeness, and insofar L2 writing is characterised by innovation in language (according to Hanauer and Kellman), as well as in thinking (according to Bakhtin), L2 writing may well be literary *par excellence*.

Therefore, a close reading in the tradition of Attridge, with attention to the literariness of the language, to formal features, tropes and literary devices such as narrative perspective, is essential to gaining a deeper understanding of literary style in L1/L2. Still, the problem of a close reading is, again, a problem of scope. It would be outside the range of this thesis to close read the entire *oeuvre* of Beckett and Lahiri. The aim is therefore to combine a computational, quantitative method with *strategic* close readings. For that, I choose two works by Beckett, where it is known that one part has been written in English and the other in French, to compare L1 and L2 style within these works by close reading.

## 1.2. Research design and methods

The analysis of L2 writing will proceed as follows. Firstly, I will be looking at the literary style of authors in their L1 and L2 *oeuvre* through the package Stylo in the program R, which is especially developed for stylometric analysis by features such as MFW (Eder et al.). Stylometry with Stylo can measure and visualise similarities between authors' use of function words and features – i.e. their minimal style. Often, methods such as cluster analysis are used for measuring the variance between authors. Yet here, I measure variance *within* Beckett and Lahiri's *oeuvres* (i.e., between their L1 and L2 texts). One pitfall here is that stylometry is therefore applied to the author's *oeuvre* consisting of *originals and translations*, because Stylo can only work with same-language texts. That is, I will be examining, for example, Beckett's entire prose *oeuvre* in English (with French originals translated into English) and then again in French (with English originals translated into French). It is important because the fact that some of the texts are translations may bias the results. To ensure that there are notable differences in style and that difference is not just the effect of translations on the results, I inspect which function words and features were distinctive for L1 and L2 texts by using Nearest Shrunken Centroid (NSC) classification. I outline the procedure of both cluster analysis and classification in detail in chapter 3.

However, computational stylometry of original texts and translations is only one part of this study. To compliment an analysis of purely word frequencies and to examine *the deficiency idea*, I examine two other quantitative measures of style because they are often employed in SLA research in assessing proficiency and matching L2 against L1 texts

(Hanauer, ‘Measuring Voice’; Jarvis 89): vocabulary, sentence length and lexical richness. These analyses are done through the Zeta in the package Stylo, scripts, and the text-analysis package koRpus. The scripts are written to measure sentence-length and lexical richness<sup>5</sup>. Furthermore, computational stylometry will be supplemented by strategic close readings of Beckett’s *Fizzles* and the short story “The End”. The focus here is on difference in style between passages originally written in English and passages originally written in French.

To answer the question whether L2 writing styles have something in common as a more general phenomenon, I study multiple L2 writers. Restricted by the scope of the thesis, two writers who compose in both L1 and L2 are included. As noted, L2 writing is not unusual, and there are many writers from antiquity until today who have written in both their L1 and L2. Among these, for the present study, the selection of authors has been restricted by several methodological considerations. Digital access to texts was important, but also that they should write in prose, since stylistic comparison across genres has proven difficult (Rybicki and Eder; Kestemont et al.). Prose is also chosen for reliability, because prose commonly has an acceptable length for stylometric analyses (in contrast to poetry). I have chosen to look only at languages which I command, to be able to understand the results and to close read Beckett’s original French. Lastly, I chose writers proximate in time, so as not to compare literary style across centuries.

Jhumpa Lahiri and Samuel Beckett have been selected because they conform to the selection criteria, but also because each of their cases are similar, but also different. In both Beckett and Lahiri first wrote in their L1 for a period, then in their L2 for a period, and then in their L1 again. Thus, we might also see the influence of L2 writing on subsequent L1 style. Moreover, Lahiri is an interesting case because she learned and began writing in her L2 Italian at an adult age, in contrast to Beckett, who learned his L2 French in school. Comparing Beckett and Lahiri therefore tells us about the role of proficiency in L2 writing style. Beckett is an interesting case because the genesis of his (bilingual) texts is well documented. Thus, various of his texts are recorded to have been composed in two languages (i.a., “The End”, *Fizzles*), so that I may examine whether the switch from L1 to L2 is stylistically visible within these texts in close reading. I have chosen Beckett’s works “The End” and *Fizzles* for the *strategic* close reading. It is strategic because these works may be

---

<sup>5</sup> The corpora cannot be distributed here for because of international copyright law. However, to make the findings of the present study easy to replicate, all scripts used in the analysis is available at: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

seen as key works for examining the L1/L2 variance of Beckett's style. "The End" is a key work, because it is the first work in which Beckett used his L2 French for literary composition. *Fizzles* is a key work because it is a late work, in which the influence of having written in L2 on L1 style may be examined.

In the following chapter, I will clarify the three main ideas about L2 writing and formulate hypotheses on their basis, which will be tested in the analysis of the two case studies (chapter 2). The practical methods are described in the following chapter on methodology (chapter 3), before they are applied in each case-study, Lahiri and Beckett in turn (chapter 4 and 5). Chapter 6 discusses and synthesises the findings in each case, and chapter 7 concludes.

## 2 How is L2 writing different? Theory and hypotheses

As noted previously, there are two opposed views on L2 writing style. On the one hand, the idea that L2 writing is deficient L1 writing. That is, L2 writing is *like* L1 writing, but is lexically and grammatically simpler. On the other hand, that L2 writing defamiliarizes and denaturalises one's native language and thereby favour a certain play with and attention to language. Additionally, there is the idea of language-specificity mainly from psycholinguistics, that the way we write, think, and feel should be different in different languages. Building on previous studies of L2 writing I outline these three theoretical ideas below.

### 2.1. The deficiency idea

As has been noted, the idea of L2 writing as deficient writing has been dominant in linguistics and SLA research before the 1980s. As has been noted, L2 writing was often connected to a lack of language mastery, or as Lahiri put it, a “stylelessness”. Beckett also used this deficiency idea as a rationale for a writing in French. Beckett is one of the most well-known multilingual writers, with a large *oeuvre* both in English and French. He was taught French in school, a subject he later studied. He moved to Paris for a lectureship in 1927, where he settled more permanently in 1937 (Kager 69). Here, after the war and after writing *Watt*, Beckett began his first prose piece in French – or rather, he began “La Fin” as “Suite” in English but switched to French midway (Slote 114). In answering another critic about this switch, Beckett described how he was able to keep his apartment after the liberation, and so: “j’y revins, et me remis à écrire – en français avec le désir de m’appauvrir encore davantage” (I went back there and sat down to write – in French, with the desire to further impoverish myself) (quoted in Kager 72, my translation). As such, even if it is seen as a kind of deficient writing, choosing to write in L2 can be construed as literary strategy

insofar as it is an intentionally imposed constraint, a way for a writer to “impoverish” themselves.

In the linguistics and SLA context, a wealth of studies assessing L2 writing compared to an L1 norm show what such “poverty” may look like. Crossley and McNamara note that the general conclusion has been that compared to L1 writing, L2 writing has “less lexical variety, specificity and sophistication”, but that in fact, aspects of L2 writing such as lexicon has been neglected in SLA research (120). Moreover, studies of L2 writing have generally studied students in an SLA classroom or students and researchers in an academic environment (clearly because the data is more readily available in these contexts). Such studies, comparing novice and more advanced L2 speakers, naturally have a larger vocabulary and grammatical repertoire (Hanauer, *Poetry* 38). Furthermore, L2 writing is found to employ less polysemous words, and simpler as well as shorter clauses and sentences even when comparing advanced and native speakers (Barone and Cargile 14, 11; Yan and Xu 331). To test the deficiency idea in the analysis, I construct the following hypothesis:

*HP1: L2 writing is lexically and grammatically poorer than L1 writing, it employs less polysemous words and sentences are shorter.*

Still, it is unclear whether the findings of these studies may be extended beyond a (Western) academic and SLA context. Findings differ, moreover, depending on which kind of L2 speaker is studied, naturally in terms of L2 proficiency (a weeklong learner of English may be supposed to have a small vocabulary), but also in terms of what the speakers’ L1 language is. For example, Reid found that L1 Spanish speakers writing in English tended to use longer sentences and more synonyms than L1 speakers of English (91). In general, these kinds of studies have focused on comparing L2 writing to a norm set by native speakers. That is, they have compared texts by for example, L2 and native English speakers, finding (and assuming) L2 texts to be deficient when compared with L1 texts. Because studies have focused on assessing L2 proficiency in relation to an L1 norm, they are more inclined to notice what L2 writers do not do, than what they actually do.

A good example of such L1/L2 comparison is Hinkel (2002), who studied L2 English writing in academia and found, for example, that these “need to learn more... idioms and collocations to develop a substantial arsenal to improve their writing in English” (247). The “need” is very much normative here, setting the native speaker’s language as the norm toward which L2 writers may aspire. In the SLA classroom and in academia, comparing and

seeking to bring L2 speakers to an L1 norm may not be a problem, since a standard language is valued in these contexts, for example to maintain a common linguistic ground. Where it does become problematic is in transferring the same ideal of the L2 speaker (as a proxy L1 speaker) to a literary context. It is contradictory to say that bringing the L2 writer to a native norm, or that, for example, learning idioms would improve L2 texts, if it is by discomfoting norms of language that texts become literary (Attridge 5).

## 2.2. The linguistic consciousness idea

The discomfoting quality of literary texts is a main tenet of formalist literary theory from Jan Mukařovský to Roman Jakobson and Willie van Peer, for whom a key element of the literariness of texts is a language-use that instils a sense of linguistic strangeness, known as defamiliarization (from the Russian *ostranenie*) or foregrounding (*aktualisace*). Czech theorist Mukařovský, who first introduced the latter term, argued that literary language is distinct from standard language insofar as it intentionally distorts it (18). Literary language may affect such distortions through literary devices, tropes, but also through non-standard language-use, neologisms, play with the polysemy of words, and non-standard sentence structures (Mukařovský 29-30). Foregrounding is for Mukařovský when such literary language is used to “place in the foreground the act of expression, the act of speech itself” (19). From this perspective, yes, it may be true that L2 writers do not use their L2 like a native, but equally true that L1 writers do not use their L1 like a native – rather, they use literary language.

Literary defamiliarization is often associated with L2 writing, as Doris Sommer does in her book *Bilingual Aesthetics*. For Sommer, there is a defamiliarizing roughness in multilingual writers’ texts, which “can irritate the senses pleasantly enough to notice both the artist at work and a refreshed world that may have grayed from inattention” (Sommer 30). This view is echoed by various literary scholars, such as Kellman, who argues that L2 texts wilfully employ language-games such as anagrams, palinodes, puns, neologisms, and calques – intended or not (Kellman *Nimble* 11-12). For Kellman, L2 writers are more “attuned to ambiguity”, and display “metalingual awareness, manifested in ostentatious verbal play and in reflexive constructions that lay bare the devices of their art” (Kellman *Nimble* 14). As has

been noted, Bakhtin is most known for the idea that bilingualism raises such “meta-lingual awareness” or “linguistic consciousness”. Based on these ideas of linguistic consciousness and defamiliarization in L2 writing, a hypothesis may be set up in the following way:

*HP2: In L2 writing, language is defamiliarized by literary devices, tropes, and non-standard language use and language-games, such as puns, neologisms, calques, etc.*

As noted, while there have been no systematic studies of such linguistic consciousness in L2 writing, there are various studies on the subject of linguistic awareness in L2 reading. I outline ideas about L2 reading here because it informs how we might understand linguistic consciousness.

In general, it can be said that reading in an acquired language is slower (Hall 84). While L1 readers possess a certain automaticity in processing language and focus on content, the L2 reader pours over the textual construction of meaning (Ibid.). Studies of even proficient speakers reading L2 have emphasised a heightened attention to surface level textual features over content knowledge (Hanauer, *Poetry* 34). The experience of L2 reading is poignantly described by Lahiri in her reflections on learning Italian, *In Other Words* (2015). Lahiri is an American writer, born in London and of Bengali descent, who won the Pulitzer Prize for her first collection of short stories *The Interpreter of Maladies* (1999). Lahiri spoke English and Bengali from a young age. She began learning Italian at an adult age (in her 30s), a language in which she has subsequently published essays, short stories and three books of fiction and nonfiction. Learning Italian and moving to Italy in 2012, Lahiri began to slowly, “painstakingly”, to read and write in Italian (Ibid.). In *In Altre Parole (In Other Words)*, Lahiri describes reading in Italian not only as a laborious, but as a creative process that is brimming with potential. She notes that when reading Italian, she is “more active, more involved”, and “[e]very page seems to have a light covering of mist... Every new construction seems a marvel. Every unknown word a jewel” (42-43). Lahiri brings a heightened attention and a sense of wonder to the L2 text and describes the encounter even with individual words in terms of discovery, almost as if she were climbing a mountain and looking over unknown lands. The positive yet disorienting encounter with the other language at the sentence and word level is illustrated throughout *In Other Words* and in her new novel *Dove mi Trovo (Whereabouts)* with short and detailed exegeses of Italian words such as the following:

*Portagioie*, the Italian word for jewelry box, is a compound of two polyvalent words. Gioia (pl. *gioie*) means both “joy” and “jewel”. Porta... belongs to a constellation of words pertaining to acts of bearing, bringing, carrying, and transporting, which in turn give rise to terms for “door,” “gate,” and “port”. *Portagioie*, therefore, could also be interpreted, in Italian, not only as a box of jewels, but a container of joy, a doorway or gateway to joy, something that brings you joy. (*Whereabouts* 158)

Similar to how Lahiri described words as jewels before, words are here described in a most material sense, connected to specific objects and actions. Lahiri emphasises sensori-motor experiences connected to the word, “bearing, bringing, carrying, and transporting”. As noted, when viewing language-learning from the standpoint of an embodied theory of language, language-learning not only implies learning a lexicon, but learning sets of culturally specific sensori-motor experiences connected to words. In the above, as throughout Lahiri’s Italian works *In Other Words* and *Whereabouts*, the sensori-motor experience of words is emphasised, showing an intense, and not purely intellectual engagement with language.

It is this attention to words and grammatical structure, and close attention to polysemy, nuance, and differences in the sensori-motor experience connected to words, that L2 readers bring to the L2 text. Hanauer notes that L2 reading implies an attentiveness especially to word choice, grammar, and to formal features, which is reminiscent of the way L1 readers read more difficult or creative texts (Hanauer, *Poetry* 34). In fact, Hanauer suggests that poetry turns L1 readers into L2 readers (Ibid.). Reading (in) an “unfamiliar language”, whether this is L1 or L2, makes one stall. It may be said that the L2 reader brings an attention to surface features of all texts, that the L1 reader only brings to literary language.

Yet does that mean that L2 writers not only read with a heightened attention to surface features in general, but also write in this way? In literary studies, this view has as noted been especially forwarded by Bakhtin, who argued that bilingualism heightens “linguistic consciousness” as a general capacity in writers. In this way, Bakhtin holds that bilingual writers, who stand “on the boundary between languages and styles”, do much the same L2 readers do (60). That is, they pay close attention to and “stylize” language, focusing on surface features, how textual meaning is constructed and how language constructs reality (60). It is similar to what Kellman argued, yet it should be highlighted that in Bakhtin’s sense “linguistic consciousness” is a feature more general than the language games Kellman claimed to be characteristic for L2 writing. We might say that it is the development of a

general capacity or awareness of language, which for Bakhtin affects the way bilinguals write *both in their L1 or L2*. It implies that learning an L2 may shift the way we use our L1 too.

Lahiri reflects on experiencing such a shift as she “returns” to English after writing in Italian: “I knew that the idea of ‘coming home’ [to writing in English] was no longer an option. I had gone too deep into Italian... My center of gravity had shifted; or at least, it had begun to shift back and forth” (Lahiri, ‘Where I Find Myself’). Here, Lahiri attests to linguistic consciousness as a double perspective, where the bilingual author is shifting “back and forth”. We might test whether “linguistic consciousness” or heightened attention to surface features is visible in L2 writing, and whether a shift is stylistically palpable in Lahiri and Beckett’s L1 writing after “returning” from L2 writing. The hypothesis may be put as follows:

*HP3: A heightened, “more active, more involved” awareness of language is stylistically apparent in L2 writing. This is also palpable in L1 texts when a writer “returns” to L1 after having written in L2.*

## 2.3. L2 writing between deficiency and linguistic consciousness

As has already been noted, Hanauer’s approach to L2 writing is situated somewhere between literary studies and SLA research, studying L2 writing by SLA research methodology, but also by distant reading methods and with a literary perspective. Clearly, Hanauer is interested in, if you will, the “surplus value” of L2 writing as a creative endeavour, not in deficiencies of L2 writing. As noted before, a key point in this approach is the idea that L2 writing is innovative and creative with language. Thus, his approach may be situated within the L2 deficiency approach methodologically, but within the context of the linguistic consciousness idea theoretically.

In Hanauer's study (2010), he systematically analyses 80 L2 poetry books by quantitative methods.<sup>6</sup> It should be noted that his findings relate to poetry, a genre that encourages certain features (e.g., short sentences). Moreover, that these poetry books were written in the academic and English Second Language environment by a culturally and linguistically heterogeneous group of student-poets (for a list, see Hanauer, *Poetry* 39). General characteristics of these L2 poems were found to be as follows: They are short poems employing short lines (5 words per line in average). First-person pronouns are markedly more frequent than other pronouns (seven times more frequent than third-person singular pronouns). They employ low frequencies of conjunctions, negations and quantifiers, which suggests to Hanauer that his L2 corpus is "characterized by a direct, descriptive style without extensive meta-discursive qualification or argumentation" (*Poetry* 43). Furthermore, L2 student-poets use a high frequency vocabulary, that is, 84 percent of the words are within the 2,000 most frequent words in the English language (Hanauer, *Poetry* 44). Also, there is a high presence of emotional vocabulary (*Poetry* 51). Overall, Hanauer describes the style of this corpus of L2 poetry as "personal, direct and descriptive" (*Ibid.*). What is important to note here is that L2 writing as a general phenomenon is suggested to have "a very specific style", despite the heterogeneity of the student-poets in term of cultural and linguistic background (Hanauer 52). Are these findings discernible beyond Hanauer's specific corpus and beyond poetry as a genre? In the following analysis, they are tested on L2 prose corpora. The hypothesis as grounded on Hanauer's findings is the following:

*HP4: L2 writing tends to be personal, direct, and descriptive. It tends to take a first-person perspective with a high presence of emotional vocabulary, little meta-discursive qualification and argumentation, and employs high frequency vocabulary.*

## 2.4. The language specificity idea

I have called it the specificity idea that bilinguals should behave differently in different languages, that is, have language-specific attitudes, worldviews, and literary styles.

---

<sup>6</sup> The poems were prompted, but results that are closely related to the prompt are excluded in the following summary.

This idea is different from the idea of L2 deficiency and linguistic consciousness because it is grounded in linguistic relativity, i.e., that different languages imply different ways of thinking, propounded by Sapir and Whorf (Pavlenko, *Bilingual* 1). The Sapir-Whorf hypothesis, as it is commonly known, has been subject to much debate. Yet the hypothesis is not as radical as we may believe (nor as radical as it has been made out to be in its historical reconfiguration<sup>7</sup>). While a strong version of a Sapir-Whorf hypothesis is frequently questioned, is not uncommon to find linguists taking an evolutionary view of language, that is, that languages adapt to the needs of a particular culture through time (Pavlenko, *Bilingual* 25). Thus, by positioning us in a culture, language positions us to habitually pay closer attention to some phenomena over others (Ibid.). That is why Pavlenko calls language an “attention-directing mechanism” (Pavlenko, *Emotions* 53).

If language is seen as an attention directing mechanism, acquiring an L2 may be seen as a reweighing of attention, so that acquiring an L2 influences the conceptions and categories of reality of inherent to the L1 (Pavlenko, *Emotions* 62). As noted, such influence of the L2 on the L1, or *vice versa*, is commonly known as cross-linguistic influence and may apply equally to the way we write as to the way we think.

Let us return to Lahiri’s example of the word jewellery-box, which she encounters in Italian as *portagioie*, and which holds the allusions of joy and gateway. Surely, such aspects of the jewellery box (as something that brings or opens up to joy) may be brought forth in English as well, such as by adding descriptors – yet they are not given in the same way as they are in Italian. These aspects of *portagioie* are *said* whether a speaker intends to or not. Still, it does not mean that they are necessarily noticed. As previously discussed, L1 speakers process words with a certain automaticity, and “unsee” surface meaning of words out of habit, while L2 speakers may be more attentive to the meaning and nuances of words. Moreover, because of these aspects of the Italian word *portagioie*, by learning Italian Lahiri may experience cross-linguistic influence and come to think of the object in a new way (as a bringer or portal to joy), in contrast to how she thought of it in English.

Yet we should not consider this re-weighing of attention by L2 acquisition a purely intellectual shift. Dijk et al. (2019) argue that learning an L2 means an expansion of not only a person’s lexical repertoire, but that bilinguals have access to different embodied conceptual and cultural repertoires, where lexically equivalent terms imply different perceptual and

---

<sup>7</sup> See the review of the disciplinary travels and rethinking of Sapir and Whorf’s ideas as the Sapir-Whorf hypothesis, in Pavlenko *Bilingual* pp. 1-18.

sensori-motor connotations in L1 and L2. Adams notes, for example, that while the concept “bread” may imply white toast for an English-speaker in the United States, for a French-speaker “*pain* might evoke a different image, one of a baguette that you buy from the boulangerie” (4). Languages thus evoke cultural notions that involve particular perceptions and sensori-motor experiences (e.g. going to the *boulangerie*, holding it under your arm, or grabbing a slice of toast from the package). Thus, a concept in L2 may have an equivalent in L1 on the lexical level, but differ on the conceptual, perceptual, or emotional level. Pavlenko has noted that bilinguals will invent words, use foreign terms, or “violat[e] both semantic and morphosyntactic constraints” of a language, to express concepts that do not have lexical equivalents in the language they seek to express it in (*Bilingual Mind* 261-3).

Two main points may be inferred from the literature on linguistic relativism and cross-linguistic influence. Firstly, that bicultural bilinguals are inclined to be creative with language in the sense of changing it, such as by using calques or neologisms, or by breaking rules of grammar and morphosyntax to express concepts, words, and nuances of a language that do not have an equivalent in the language they seek to express it in. This idea is in line HP2 but should be distinguished from it since different mechanisms are at work here. A hypothesis may be formulated on the basis of this idea:

*HP5 Authors use language creatively compared to standard language-use to express the nuances of a word or phrase in one language, which does not have an equivalent in the other.*

Secondly, it can be inferred that bilinguals may prefer certain words, concepts, or topics in one language over the other (eventually because the word or concept does not exist in one of the languages). For example, because of the conceptual and sensori-motor nuances of the term *portagioie* in Italian in comparison to English, Lahiri may be inclined to use it more in her Italian than in her English works. A sixth hypothesis follows:

*HP6: Authors prefer certain words, concepts, or topics in one language in comparison to the other.*

There is, however, another sense in which authors may be said to have language-specific style, if we do not think of style as specific to a particular language (Lahiri’s Italian style, Beckett’s English style, etc.), but as specific to either L1 or L2 as theoretical categories (Lahiri’s L2 style, Beckett’s L1 style, etc.). That is, instead of thinking that style differs

according to language, style differs according to whether the author is writing in an acquired or native language.

## 2.5. The L2 specificity idea

Hanauer found it to be a characteristic of L2 writing that it is more emotional or personal than L1 writing. Yet the idea that L2 writing should be more emotional than L1 writing is not generally supposed about L2 use. Actually, especially studies of bilinguals in psychology have suggested the opposite, namely that the mother tongue has a privileged access and more primary relation to the speakers' emotions. Pavlenko notes that psychological studies of the mid 20th-century have firmly established "the view of the first language as the language of emotions and the second as the language of distance and detachment" (30). Psychoanalysts became attentive to the significance of the language of analysis in the wake of the Second World War, and found that switching to L1 could prompt patients to recount and connect emotionally, to remember more vividly and overcome repressions (Aragno and Schlachet 24). The second language was thus conceived as more distanced, or even as a means of emotional repression (Aragno and Schlachet; Byford; Foster; Javier). There is evidence to support this view. It has been noted even from the time of Freud that patients in psychoanalysis seem to prefer using their L2 for swear words and for certain subjects (Pavlenko, 'Affective processing in bilingual speakers' 410, see also Pavlenko *Emotions* 29). More recently, it has been shown that bilinguals are differently affected by emotional stimuli in their L1 and L2. For example, people react more emotionally to swear words in L1, which Harris et al. measured by skin-conductance, what Pavlenko calls the "L2 detachment effect" (Harris et al., Pavlenko "Affective Processing" 421).

As such it would seem like a widely acknowledged fact that the L2 does not bear the emotional weight of the mother tongue. Does that, contrary to Hanauer's findings, mean that L2 writing is less personal? Not necessarily. A similar detachment effect or "distance" to the L2 has been noted by L2 writers, especially of the Second World War generation such as Jerzy Kosinski. Yet here, it seems that the "detachment effect" not only implies an emotional disconnection to L2, but may even enable authors to explore thoughts and feelings L2 more freely. Reflecting on his writing in English, Kosinski notes: "[i]t was a great surprise to me...

that when I began speaking English, I felt freer to express myself... all the thoughts that I would have found difficult to reveal in my native tongue” (Kosiński 125). That there should be a certain emotional “distance” in using the L2 is a topos that recurs in the reflections of L2 writers across time and continents. The Spanish/English writer Ariel Dorfman, for example reflects that in writing his memoir *Heading South, Looking North*, that English “allowed me to see the events in a different (or at least tolerable) light... show myself, perhaps reveal myself, use the distance, treat myself as an almost fictional object” (Dorfman 212). Yet, when writing about the same events in Spanish Dorfman notes that: “I would find myself sick and trembling, faint with anxiety” (Ibid.). This topos of distance is not exclusive to the reflections of literary writers. Also philosopher Hannah Arendt notes: “Ich schreibe in Englisch, aber ich habe die Distanz nie verloren” (Arendt 24). To clarify, it may be that L2 writing is more impersonal or “distant” than L1 writing, yet it may also be that author’s use that distance to express themselves more freely. Therefore, it is interesting to test whether an L2 detachment effect is apparent in L2 writing of Beckett and Lahiri, so that their L2 writing is either more or less personal and emotional. A hypothesis may be constructed in the following way (and may of course be rebutted):

*HP7: L2 writing is more personal and emotional, due to the L2 detachment effect.*

The theoretical framework outlined in this chapter proposes differences between L1 and L2 writing as the basis for the following analysis, offering possible explanations for these differences. I am aware of nuances subsumed under the headings, and these ideas should above all be considered as heuristic constructs for the analysis, as should the hypotheses. Methods of analysis have been adjusted to examine the above hypotheses, as shall be discussed in the following chapter. Each analysis will take stock of these hypotheses, which are listed in appendix 1.

### 3 Reading L2 style Methodology

Before turning to the analysis, it is important to “look under the hood” of especially the distant reading methods employed in the present study, since a thorough understanding of these is the basis for a good interpretation of results. In the present chapter I discuss the selection of features and texts for examining minimal style by cluster analysis and classification in Stylo as well as their reliability and describe the procedure of these two methods. Moreover, I outline the procedure for examining quantitative features of a more expanded notion of style, sentence-length and lexical diversity and conclude with remarks on close reading.

As noted in the introduction, studies of computational stylometry often operate with the idea of an author-specific “stylome” which may be understood as an author’s idiosyncratic usage of the topmost frequent function words or “minimal style”. However, what actually goes into the categories function- and content-words is not clear-cut. In theory, function words are words with a grammatical function, they indicate how content-words, which carry semantic meaning, go together (Corver and Riemsdijk 1).

Yet nuances are lost in this distinction between content and function words. There are function words with lexical content and content words with grammatical properties, such as prepositions, which may carry semantic meaning. For example, as in: “I count on you” in contrast to the purely prepositional use of “on”, as in: “coffee is on the table” (Corver and Riemsdijk 4). There are also classes of content words that have a grammatical function. Corver and Riemsdijk list the following classes of (discrete) words: grammatical nouns (i.a., one, self, thing), grammatical verbs (auxiliary verbs and verbs such as, i.a., want, make), grammatical adjectives (i.a., other, same, such)(6). I see no reason not to include these classes in my selection of function words, seeing that Hulle and Kestemont also include auxiliary verbs in their stylometric study of Beckett (182).

Moreover, Jocker’s observations on the importance of punctuation marks, such as the comma or exclamation mark, in distinguishing authorial from each other should be taken into consideration in selecting features for stylometric analysis (Jockers 77). Jockers notes that the most important minimal features in predicting the author of a text are marks of punctuation

(99). These are often excluded seeing that editors may have a hand in how they are used. That is, because they point to the editor over the author. Yet, it is well-known that Beckett and Lahiri had the last say in the editing process, and were attentive to their use of punctuation-marks (Germoni and Sardin, Lahiri, *Other Words* 177). As such, I see no reason to exclude punctuation marks from the most frequent features (MFF) of minimal style here, as they, like function words, perform a grammatical function. In fact, they may even substitute a function word, like a semicolon may replace a conjunction.

Thus, the cluster analysis and classification performed in the present thesis consider the function words and features, the most common adverbs (such as *very*, *really*), articles, auxiliary verbs, conjunctions, interjections, modals, negations, prepositions, pronouns, quantifiers, and question words within the 600 most frequent features (MFF), including the grammatical content words listed above and punctuation marks. The usage of these words and features is what I call an author's "minimal style". An example of the selection of function words is shown in table 1<sup>8</sup>. In the analysis, I refer to the selected features as "function words and features".

<p>,, ., de, et, la, le, à, je, que, il, -, les, pas, ne, en, un, ce, dans, qui, se, dit, une, plus, pour, mais, des, me, du, au, tout, ou, ", elle, c'est, son, ?, sur, par, comme, si, qu'il, lui, moi, sans, avec, est, être, même, sa, ça, était, #monsieur, on, avait, rien, peut, #bien, là, mon, non, où, cette, #temps, #dire, vous, nous, peu, y, jamais, -, ses, ils, encore, ma, a, car, d'un, fait, faire, cela, deux, toujours, moins, d'une, aussi, puis, quand, ni, aux, #watt, alors, mes, avant, c'était, autre, chose, tu, !, j'ai, n'est, suis, ces, #tête, fois, #mercier, #camier, #jour, vers, l'autre, dont, #moment, ainsi, quoi, tous, qu'elle, #loin, qu'on, quelque, #nuit, mal, #yeux, #dis, petit, n'y, sous, leur, sont, oui, été, jusqu'à, donc, #vie, comment, voilà, #seul, après, n'était, mieux, #murphy, #voir, enfin, s'il, celui, #voix, maintenant, tant, serait, #sais, #main, soit</p>
---

Table 1: The top 150 MFF in Beckett's *oeuvre* in French descending order. Words preceded by # have been excluded.

It should be noted that some words can belong to more than one class of words, such as *moins* (a preposition and noun). For these, I decided to err on the side of inclusion rather than exclusion. I have chosen to look at the function word and features within the 600 MFF, because this number is higher than in the analysis of Hulle and Kestemont (300 MFW) yet may still be seen to capture the *minimal* style of an author. Yet, after selection this number is much lower<sup>9</sup>.

<sup>8</sup> For the total list (with selection) used in each analysis, see:

[https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

<sup>9</sup> I note the number of actual features examined in the specific analyses of Beckett and Lahiri.

Many computational studies of literary texts yield impressive results, yet, as Eder notes, the question of their reliability often remains implicit (Eder, ‘Does Size Matter?’ 167). In the last decade stylometric methods have developed greatly by larger developments in general Natural Language Processing (NLP)(Daelemans 425). Therefore, ways of ensuring reliability have been implemented, such as testing both method and measures on benchmark corpora (Ibid., Eder, ‘Does Size Matter?’). Eder has done this for stylometric literary analysis that relies on MFW, looking at reliability according to sample length on benchmark corpora. He finds any length of samples (/texts) below 3,000 “disastrous” for reliability in English-language analyses (Eder, ‘Does Size Matter?’ 170). Still, it should be noted that results are increasingly better up to 5,000 words, at which point authorship attribution by classification in English is correct more than 80 percent of times on a benchmark corpus (Eder, ‘Does Size Matter?’ 180). Taking reliability into consideration, I only look at texts which are around 4,000 words or more. The reason for not setting the limit higher is because of good results with samples of 4,000 words, and because many texts would then be excluded (particularly of Beckett’s *oeuvre*). This minimum limit applies to the texts after pre-processing. In pre-processing, any non-authorial introduction or afterword, editorial information, footnotes, references, glossaries, acknowledgements, and page numbers have been deleted from them<sup>10</sup>.

Studies of computational stylometry often use a type of machine-learning method to test the strength of groupings initially indicated by another method, for example, cluster analysis (Eder and Rybicki, Hulle and Kestemont). I do the same here, first “exploring” the corpora by cluster analysis, and then testing groupings by classification. Cluster analysis is an exploratory statistical method, which, when using delta-distance, will visualise differences between texts based on normalised word frequencies (so called z-scores) and group similar texts together in a dendrogram. Z-scores indicate the difference between the frequency of a word or feature in a given text, compared its average frequency in the whole corpus. To simplify, the result is that, generally, a text which has a very low frequencies of some words will not be grouped with a text that has a very high frequency of the same words. Cluster analysis will thus show groupings of what are assumed to be similar texts in terms of how they use certain words and features (here, function words and features within the 600 MFF).

In classification, in contrast to cluster analysis, is “supervised”, that is, the number of groupings or “classes” are specified in advance. To classify a text og group of texts, a

---

<sup>10</sup> For a detailed list of the works included, see: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

prediction model is trained on a set of texts (the training-set), which computes a prototype for each of the specified number of classes based on the frequencies of selected features (in the present study, function words and features within the 600 MFF). The classifier is then asked to predict the belonging of any one additional text (the test-set) to one of the classes in the training-set, based on the similarity of the text to class-prototypes. Often, this method is used in authorship attribution for attributing an unknown text to classes of known authors, determining whose style the text is the most alike (Eder and Rybicki). Yet, we may also have both the training- and test-set consist of the same (known) authors' texts, to see how well the classifier is able to compute a "prototype" for these authors' styles. For example, we assume that the early works of Beckett written in English and the middle-period works written in French are each a stylistic unity. We then train the classifier on a training-set consisting of two classes, one of early English texts and one of middle-period French texts, for each of which the classifier computes a "prototype". Then we see how well the classifier can predict, say, whether *More Pricks than Kicks* belongs to the early English works (it does), in comparison to how many times it predicts the text to belong to other classes, such as Beckett's middle-period French works. If the classifier could assign *More Pricks than Kicks* with 100 percent accuracy to the class of early English works, the text is indeed much more similar to early English works than to another class. I use classification in this sense, not to predict the author of an unknown work, but to assess "the signal strength" of the groups that were suggested in the cluster analysis (Jockers 79). That is, how alike the texts of each author are, or how coherent they are in *minimal* style.

However, how well the classifier can attribute a certain text to a certain author depends very much on which texts one chooses to train to model with. If the model is trained on texts that are characteristic for the author (e.g., for Goethe, perhaps *Wilhelm Meister*), it will be better at predicting the right class, compared to training the model on texts that are uncharacteristic (e.g., Goethe's *Farbenlehre*) (Eder and Rybicki 230). In classification there is a method to minimise the effect of nonrepresentative training-sets, known as cross-validation. With cross-validation, classification is performed several times during which, for each iteration, the model swaps texts between training- and test-set. Cross-validation mitigates the risk of either getting too optimistic or pessimistic results, by examining different compositions of training and test-sets in each iteration, and so minimises possible biases. When setting cross-validation to 100, there are 100 iterations. Usually, doing 10 cross-validation rounds is satisfactory, but when treating the highly complex data that texts are, it

has been shown that it may not be (Eder and Rybicki). Considering the reservation of Eder and Rybicki, the following classifications are performed with a 100-fold cross-validation.

There are a number of different measures to estimate how well a classifier performs. *Accuracy* indicates *how many times the classifier matched the correct text to the correct class, divided by the total number of texts*. That is, if we have 10 texts, 80 percent accuracy would indicate that the classifier attributed 8 texts to the correct class. Yet reporting only *accuracy* often overstates the results (Eder, ‘Performance Measures’).

*Precision* indicates *how many times a classifier attributed a text to a class, in comparison to how many times it was right*. For example, if it guesses that 5 out of 10 texts belong to Beckett’s early English works (even if, actually, only 3 do), and out of those 5 times, it was correct only 2 times, the *precision* would be  $2/5$ . Precision can in this way be thought of as a sensitivity measure. If precision is low, the classifier is too “greedy” at guessing.

*Recall* indicates *how many of the texts were attributed to the right class, in comparison to how many texts actually belong to that class*. For example, if it guesses that 5 texts belonged to Beckett’s early English texts (while only 3 do) and out of these guesses, it was right 2 times. In this case, the *recall* would be  $2/3$ . A low recall indicates that the classifier is too “picky”.

The *F-score*, in contrast, *combines precision and recall, to measure the overall performance of the classifier*. Importantly, the *F-score* tells us about the stylistic consistency of the class itself, i.e., how sharply classes are distinguished from each other. If the *F-score* is low, that would not merely suggest that the classifier is wrong, but that we were wrong in our initial assumptions about the stylistic unity of, for example, Beckett’s early English texts and Beckett’s middle-period French texts. That is, a low *F-score* would indicate that the classes are not that different, because the classifier is constantly confusing early English texts with middle-period French texts.

*F-score*, however, tells us nothing useful about the times when the classifier guesses wrong. It may be important to inspect misclassifications, since misclassifications may suggest that another class is stylistically similar to the class the text actually belong to. Therefore, it is useful to create a table (also known as a confusion-matrix), which list the predicted results, i.e., the classes the classifier predicted the texts to belong to, against the expected results, i.e.,

the classes we know the texts to belong to<sup>11</sup>. In the following analysis, I report confusion matrix and *F-score*, yet details on cross-validation rounds, *accuracy*, *precision*, and *recall* are available at the online repository<sup>12</sup>.

I have been calling what the classifier computes for each class a “prototype”. Yet in truth, classification methods have very different ways of measuring similarity. There are several methods of classification, such as delta, random forests, nearest shrunken centroids (NSC), support vector machines (SVM), etc., each measuring the distance between classes of texts differently. NSC computes prototypes or “shrunken centroids” for each class. These are technically at the geometrical centre, the “centroid”, of the features in each class of the training-set, when they are represented as data-points in multidimensional space. In the case of the present study, the dimensions are the frequencies of our selected features within the 600 MFF (i.e., up to 600 dimensions). The test-samples are then classified according to which of these class-centroids they are most alike, or “nearest” in that multidimensional space. That centroids are “shrunken” means that they have been filtered in such a way as to exclude outliers in each class (for a detailed account on shrinkage in NSC, see Tibshirani et al.). NSC has the advantage of being more “readable” in comparison to, say, SVM, because NSC displays which words or features were most important for computing a class-prototype in comparison to other class-prototypes. Examining these important words and features gives us an idea of how use of function words and features differ between classes.

Further quantitative analysis of vocabulary in L1 and L2 texts is performed with Zeta in Stylo on the same corpus with the same pre-processing as in cluster analysis and classification. Zeta measures the distinctiveness (or keyness) of a given feature in a group of texts compared to another group – in the present study L1 compared to L2 texts. It is one way to compare vocabulary between groups of texts. I use (Craig’s) Zeta<sup>13</sup> here, since it one of the methods supplied in the Stylo package, yet different methods do measure distinctiveness somewhat differently. It has been shown that some methods that are based on word frequencies (such as the Chi-square) cannot deal appropriately with text-length and the possible overrepresentation of some words in but one of the texts in a corpus (Kilgarriff). For example, if we have 3 literary texts in group A, but 2 literary texts and a longer medical text in group B, methods such as Chi-square may indicate that medical terms are distinctive for

---

<sup>11</sup> For the script involved here, see: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

<sup>12</sup> See: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

<sup>13</sup> First proposed by John Burrows, it has since been developed by Hugh Craig (Burrows, "All the Way"; Craig and Kinney).

group B as a whole, though this is only true for one text. Zeta is more reliable in this regard since it measures high or low *consistent use* of words.

With Zeta, texts are split into segments of a specified size (commonly around 2,000 words) in which occurrences of words are counted, by which the Zeta-score of each word in each group is returned (Hoover, "Zeta" 2). The Zeta-score is calculated by counting binarily (0 or 1) whether a given word is found in a segment at least once. The percentage of segments where the word is present is then subtracted by the percentage of segments where the word is absent. A consistency percentage of 0,7 would mean that the word occurred in 70% of segments, so that Zeta in fact measures the *consistency* of a word in a group of texts. The distinctiveness (Zeta-score) of a given word in a group of texts is then calculated by comparing the consistency of the word in one group of texts to its consistency in the other group, subtracting the latter percentage from the former. If, say, the word "he" has a consistency percentage of 0,5 in group A, and 0,9 in group B, the Zeta-score would be -0,4, and "I" would be more distinctive for group B. If a word has a high Zeta-score closer to 1 or -1 it is more distinctive, while scores approaching 0 would indicate that those words are used with similar consistency in both groups of texts.

The other quantitative measures of style in the present study, mean sentence length and lexical diversity are, again, also applied to the same corpus with the same pre-processing, comparing L1 texts to L2 texts of Beckett and Lahiri's *oeuvres*. While computing sentence-length is relatively straightforward,<sup>14</sup> a word of caution on lexical diversity measures. As noted, it has been argued that L2 writing is less lexically diverse than L1 writing. Moreover, Hanauer found that L2 English writing uses words found among the topmost common words in English (Hanauer, *Poetry* 44). He determines Lexical Frequency Profile (LFP) for his L2 poetry corpus, a method that compares the vocabulary of a certain text to most frequent words in English and to an academic word list (see Laufer and Nation). However, such profiling is specifically developed for English and there is (yet) no corresponding measure in other languages, why it is not implemented in the present study. Instead, I opt to look only at lexical diversity, which in theory indicate how varied vocabulary is. That is, lexical diversity is lower if words are repeated, and if there are few *hapax legomena*, i.e., words occurring only once. I assume that if L2 writing only uses the most frequent words of the L2 language (so Hanauer), lexical diversity will also be lower.

---

<sup>14</sup> See code at: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

While various methods have been proposed to measure lexical diversity, the problem is that many are sensitive to the length of the text. Often, studies employ a variation of the Type Token-Ratio (TTR), in which the total number of unique words (types) in a text is divided by the total number of words (tokens). Ultimately, TTR measures lexical diversity by measuring repetition. For example, “to be or not to be, that is the question” consists of 10 tokens, but 8 types, so that the TTR is 8 divided by 10 (i.e., 0.8). If TTR is 0.8, 8 out of 10 tokens are types, if it is 0.3, only 3 out of 10 are types, etc. The closer the TTR is to 1, the higher the lexical diversity of the text. However, clearly, if a text is very long, the number of types will also be lower, since the chances that the same word will be used again are bigger. If a text is very long, the number of total tokens will also be higher, so that lexical diversity will ultimately be lower the longer the text. There are, however, measures of textual richness that take text length into account such as the Mean Segmental TTR (MSTTR) and Moving Average TTR (MATTR). These calculate a *mean* TTR for either a segment or a moving window. The MSTTR method slices texts into same-sized segments, calculates the mean TTR in each segment, and then calculates MSTTR, the mean of all segment-means of a text taken together. As per the same logic as when calculating TTR, MSTTR score will be lower the bigger the size of the segments. In the following analysis, I use what has been suggested as the standard size of 100 words per segment (i.e., MSTTR-100)(Torruella and Capsada 449). I use the koRpus package in R for this analysis<sup>15</sup>.

As noted in the introduction, I intend to combine quantitative measures of style in L1 and L2 texts with a *strategic* close reading, whereby style in a wider sense may be examined, with special attention to literary devices, form, tropes and language-use (Attridge, *Peculiar* 2). As the approach of my close reading has already been discussed, I will not discuss it at length here. Still, I want to clarify basic terms from traditional stylistics and narratology that are used in the analysis. These are, firstly, the literary devices *narrative perspective* and *focalisation*. *Narrative perspective* (or point of view) is the mode of the narrating voice, such as the first-person narrative perspective, where the narrator speaks in the “I”-form, often with an “intimate effect”, or the third-person narrative perspective, where an (often omniscient and impersonal) narrator speaks about a character (Booth 150). Another term to describe the way a story is narrated is *focalisation*, which can be internal or external (Genette 10–11). Internal focalisation is where the story is narrated through a character, that is, telling only what the

---

<sup>15</sup> For validation of the results, MSTTR was performed in both koRpus and quanteda, another similar package. The (similar) results, are compared here: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

character knows, with insight into the character's thoughts; external focalisation is narrated about a character, telling only what can be seen from an outside perspective and without insight into the character's thoughts (Genette 189). Secondly, another term in the analysis is *language-use*, whereby I mean the choice of words and phrases in a text, as often compared to norms of communication in a given language (Mukařovský 17). For example, using idiomatic expressions is a standard language-use, while changing idiomatic expressions or coining new words (neologisms) is a non-standard language-use. Thirdly, when referring to tropes, I mean particular figures of speech such as metaphor, syllepsis, antanaclasis, etc., which use language in a specific way, often to point beyond the literal meaning of words (Baldick, 'Trope'). Lastly, when using the term form, I refer to formal elements of the story, such as the structure and arrangement of the literary work in a broad sense.

The following analysis examines a minimal notion of style computationally, as well as other features of an expanded notion of style by looking at sentence length, vocabulary, and by close reading the two selected works by Beckett (chapter 5). The overall aim is to examine whether there is a difference between the L1 and L2 style of Beckett and Lahiri, and if those differences can be described by the hypotheses 1-7. First, I examine the minimal style of these texts by cluster analysis and classification, looking only at the 600 MFF<sup>16</sup>. Thereafter, I look at features according to an expanded notion of style, the average sentence length and vocabulary. In the part of the analysis of Beckett's *oeuvre*, I conclude with the two close readings.

---

<sup>16</sup> For a list of total features examined in the analysis of Lahiri's *oeuvre*, see: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

## 4 Jhumpa Lahiri Analysis

In contrast to the analysis of Beckett's *oeuvre*, both fiction and nonfiction works are included in Lahiri's case, since she has many publications of nonfiction in each language (in contrast to Beckett). One of Lahiri's works of nonfiction in particular demands attention when asking about language-specific style, which is *Translating Myself and Others* (2022). This work consists of essays originally composed in both English and Italian. In the following analysis, I divide these according to language of composition, so that there is a *Translating Myself and Others* consisting of the essays originally written in English and a *Translating Myself and Others* consisting of the essays originally written in Italian<sup>17</sup>. Moreover, one of the essays in this collection is written in both Italian and English, "Where I Find Myself". In the following English-language analysis, the essays of this collection originally written in English are identified by the abbreviated title of the collection preceded "En", while the essays originally written in Italian in the same collection are similarly identified but preceded by "It" ("It\_Translating"). The essay "Where I find myself" is identified by that title as a mixed-language composition ("En/It\_Where"). In the Italian-language analysis, I consider only four of the essays included in *Translating Myself and Others*, since only four of them are available. Three of these were originally published in Italy<sup>18</sup>, and are examined jointly under an abbreviated title "It\_Lingua-Calvino-TreUltime". Moreover, the essay "Where I find Myself" of mixed-language composition was translated into Italian by Domenico Starnone and included (in Italian) in the English publication of *Translating Myself and Others*. In the following analysis, it is identified by the abbreviation of its Italian title, again as a mixed-language composition ("En/It\_Traduttrice").

While it is not optimal to work with different versions of *Translating Myself and Others* in this way, it is important because the essays were written proximately, in Rome and Princeton 2019-2021 and published in 2022. As such, by comparing the essays of this

---

<sup>17</sup> The language of composition for each essay is explicitly noted by the author herself in *Translating Myself and Others*, pp. 159-160.

<sup>18</sup> "Lingua" was first published in *Corriere della Sera*, Oct. 16, 2021. "Calvino Abroad" was first published in *La Lettura*, Sept. 19, 2021. Lahiri originally delivered a version of "Tre Ultime Metafores" as a speech when receiving an honorary degree in Specialized Translation at the University of Bologna. It was subsequently published in *Domani*, Nov. 5, 2021.

collection internally, we can investigate whether they display stylistic variance according to language of composition, despite their proximity in time. Moreover, the 2022 collection contains the first texts Lahiri has written in English since *The Lowland*, nine years prior. It is therefore an important work to examine when asking whether the way Lahiri writes in English has changed when “returning” from writing in Italian for an extended period of eight years.

Table 1 shows the total materials considered in the following analysis of Lahiri’s L1 and L2 writing. Few of Lahiri’s works have been excluded (only pieces published in *The New Yorker* that were shorter than 4,000 words). Yet it should be noted that some works have not been translated into Italian (marked NA below), even if all her Italian works have been translated into English. The Italian language analysis is therefore based on a smaller corpus, which should be kept in mind. Each work is assigned an abbreviated title, which identify the works in the following visualisations.

*Table 2: Lahiri’s oeuvre* (chronological by year of publication). Note that some works only exist in the original English and are therefore marked “Not Applicable” (NA) in the columns of Italian data. Note also that the most recent work *Translating Myself and Others*, is only partly published in Italian. Thus, the Italian version of this text is abbreviated: *Lingua-Calvino-TreUtime*. The total count of features in English and Italian includes, as discussed, words *and* punctuation marks.

Title (EN)	Abbreviated title (En)	Abbreviated title (It)	Year of publication	Published in collection / journal	Original language	Total features (En)	Total features (It)	Genre
Interpreter of Maladies	Interpreter	Interprete	1999		en	73819	67313	Fiction
The Namesake	Namesake	Omonimo	2003		en	121539	113713	Fiction
Unaccustomed Earth	Unaccustomed	NuovaTerra	2008		en	129189	129301	Fiction
The Lowland	Lowland	Moglie	2013		en	129991	131742	Fiction
Brotherly Love	Brotherly	NA	2012	<i>The New Yorker</i>	en	16299	NA	Fiction
Rhode Island	Rhode	NA	2008	<i>State by State: A Panoramic Portrait of America</i>	en	5491	NA	Nonfiction
The Space between the Pictures	TheSpace	NA	2011	<i>The Suspension of Time: Reflections on Simon Dinnerstein and the Fullbright Triptych</i>	en	4565	NA	Nonfiction
In Other Words	OtherWords	AlterParole	2016		it	27916	26267	Nonfiction
The Clothing of Books	Clothing	Vestito	2016		it	9042	8966	Nonfiction
Whereabouts	Whereabouts	DoveMiTrovo	2018		it	33742	29404	Fiction
The Boundary	Boundary	NA	2018	<i>The New Yorker</i>	it	3631	NA	Fiction
Casting Shadows	Casting	NA	2021	<i>The New Yorker</i>	it	8164	NA	Nonfiction
Where I Find Myself	Where	Traduttrice	2022	<i>Translating Myself and Others</i>	mixed	5394	5126	Nonfiction
Translating Myself and Others	TranslatingMyself	Lingua-Calvino-TreUtime	2022	<i>Translating Myself and Others</i>	mixed	9059	7440	Nonfiction

# 4

## .1. Distant reading

### Cluster analysis

When looking at Lahiri’s bilingual *oeuvre* by exploratory cluster analysis, we clearly see a division along the language line.<sup>19</sup> The cluster analysis in Italian most strongly renders this division (fig 1.), where texts originally written in Italian (green) and texts originally written in English (red) are considered clustered together each on a branch of their own in the dendrogram. The mixed-language composition *Traduttriccce di me Stessa* (blue) is situated in-between the two clusters yet is more similar to texts originally in Italian than to texts originally in English. There is also a slight difference in the use of function word and features between genres, as the topmost cluster of originally Italian nonfiction texts *In Altre Parole* and the three essays (“Lingua”, Calvino and “Tre Ultime”) are more similar to each other than to the work of fiction *Dove Mi Trovo* which is one step removed. Yet, as noted, the analysis in Italian is based on a smaller corpus, while the analysis in English allows for a more complex picture (fig. 2).

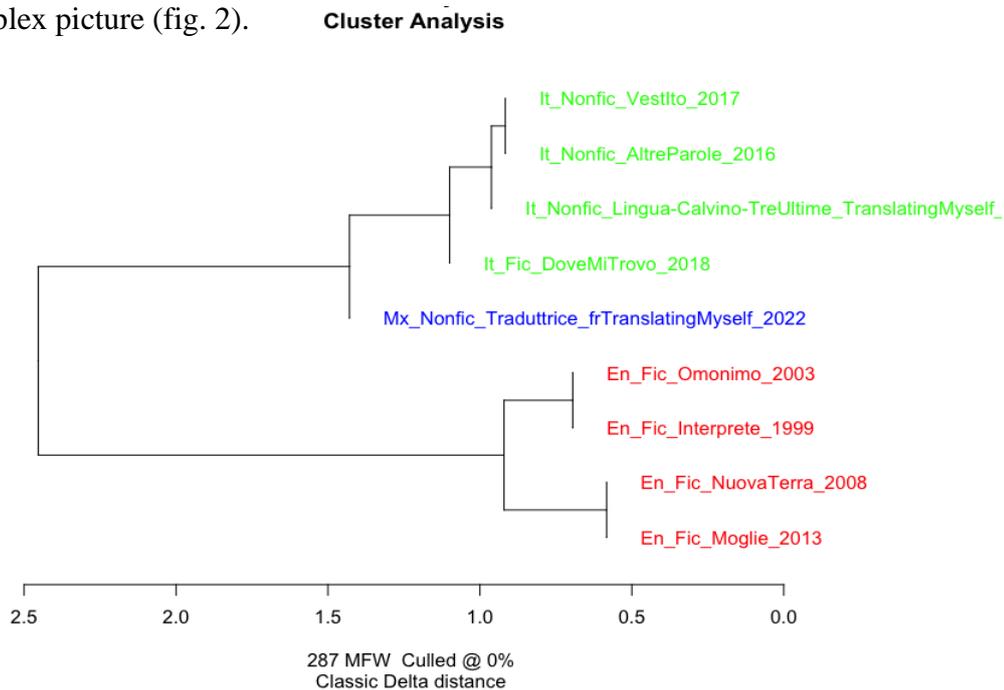


Figure 1: cluster analysis of the English-language corpus visualized in a dendrogram. Reading the dendrogram from right to left, the most similar works are to the rightmost side and decreasing toward the left, so that, e.g., *La Moglie* is more similar to *Nuova Terra* than to *L’Omonimo*. Language of composition is indicated by either “It” or “En”, and colours also refer to language of composition (here, green for Italian, red for English, and blue for the mixed language composition “Where I Find Myself” which is part of the collection *Translating Myself and Others*).

<sup>19</sup> All visualisations from the analysis part of this study are also to be found at: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

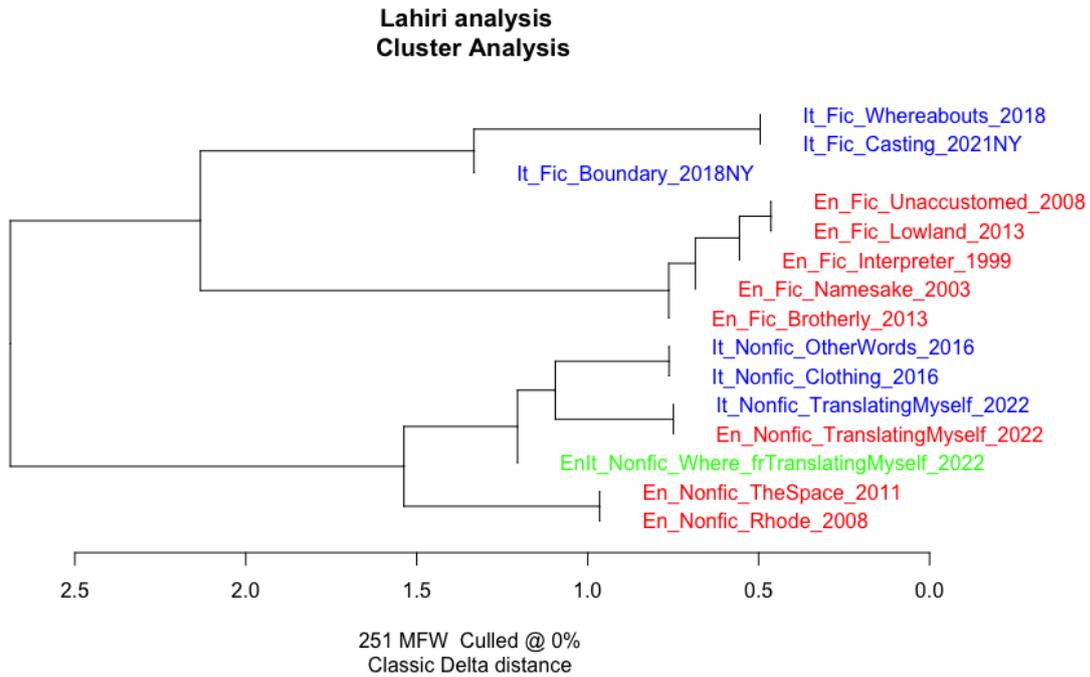


Figure 2: Cluster Analysis of the English-language corpus visualized in a dendrogram. The language of composition is indicated by either “It” or “En”, and colors also refer to language of composition (blue for Italian, red for English, and green for the mixed language composition “Where I Find Myself” which is part of *Translating Myself and Others*).

The analysis in English shows the same tendency as the analysis in English, that texts are considered similar which are written in the same language (fig. 2). In figure 2, works originally written in Italian (blue), and works originally written in English (red) are shown as belonging to distinct groups. Especially the topmost branch is clearly divided into originally English and originally Italian works. That *Whereabouts* and “Casting Shadows” should be considered very similar in the topmost cluster of Italian fiction is no surprise, since parts of *Whereabouts* were reused and elaborated for “Casting Shadows”, which was not published in its original Italian but first published in English in *The New Yorker*. Therefore, “The Boundary” also stands out a little compared to these two works. Yet overall, the cluster analysis identifies these three Italian works as belonging to the same “branch”, distinct from the branch below of English fiction.

Moreover, fiction and nonfiction are clearly considered stylistically distinct. That is notable, since here, only function words and features among the 600 MFF are analysed. That nonfiction and fiction is grouped distinctly in this analysis, implies that Lahiri’s use of function words differs considerably according to genre. In figure 1, we thus see two main branches, fiction and nonfiction, and four sub-branches, texts in each genre grouping according to language of composition. If we look at the top fiction branch, we see a clear distinction between fiction written in English and Italian; so too in the lower cluster, between

nonfiction written in Italian and English – but only before 2022. That is, there is a clear clustering according to language of composition if we disregard the three middle-position texts that belong to the collection *Translating Myself and Others* published in 2022. The essays in *Translating Myself and Others*, whether they were originally composed in Italian (blue), English (red), or in a mix of the two (green), are clearly alike, as they are grouped close together, although the mixed-language essay “Where I Find Myself” is at a remove. Yet it seems that the 2022 essays written in English are closer to Lahiri’s minimal style in Italian nonfiction, such as in *The Clothing of Books* (2016), than they are to her style in English nonfiction, such as *Rhode Island* (2008). Indeed, something seems to have happened to her style in her newest nonfiction collection *Translating Myself and Others*. In fact, if we disregard this latest collection, there is a perfect division along the language line between English and Italian fiction, as well as between English and Italian nonfiction. That is, Lahiri clearly has language-specific styles in each genre *before 2022*.

### Classification of language-style groupings

As noted, classification helps us to examine the strenght of groupings seen in the cluster analysis, and the NSC in particular gives us a hint as to the distinguishing features. When running the NSC classification on Lahiri’s corpus in English, considering English and Italian fiction, as well as English and Italian nonfiction as four distinct classes, these classes are shown to be quite different. The NSC classifier generally attributes texts to belong to the right classes, with an average *F-score* of 0.87. In the same classification performed on the corpus in Italian, I have specified only two classes, texts originally in English, and texts originally in Italian (since there is only one nonfiction texts in the smaller corpus in Italian). Here, the *F-score* is even higher, at 0.98, implying that the classes are very distinct in their use of function words and features in the Italian.

Thus, there are more misclassifications when trying to classify texts in English (lower *F-score*). The errors of classification are also telling and can be seen in the confusion tables below (table 3 and 4).

*Table 3:* Confusion matrix of the classification in English, which indicates the misattributions (“confusions”) of the classifier. For a perfect classification, values should only group along the diagonal (marked in grey), so that, for example, English fiction is only attributed to its proper class, English fiction. Classifications are rarely perfect. The numbers are high because cross-validation was performed 100 times. High numbers are not as important as the distribution of numbers along the horizontal line.

	English Fiction	English nonfiction	Italian Fiction	Italian Nonfiction
English Fiction	195	0	1	0
English Nonfiction	4	67	0	19
Italian Fiction	1	0	99	0
Italian Nonfiction	0	33	0	81

*Table 4:* Confusion matrix of the classification in Italian. Note that since the Italian corpus is smaller, there are not as many classes. The classification is set to classify texts as either originally English or originally Italian.

	English texts	Italian texts
English texts	197	0
Italian texts	3	100

In the Italian analysis (table 4), the classifier only misclassified texts 3 times, and every time it classified an Italian text as one originally written in English. In the English analysis (table 3), the picture is more complex since the classifier is working with more classes (fiction and nonfiction, originally English and originally Italian texts). Out of all the times the classifier guessed where English fiction-texts belong, it only guessed wrong once, where it attributed an English fiction text to Italian fiction. The same can be seen for Italian fiction, which the classifier only once misattributed an Italian fiction text to English fiction. As such, English and Italian fiction-texts are very distinct in their use of function words and features and are easily classified correctly.

Nonfiction written in Italian and English, however, is more similar. Italian nonfiction-texts were classified as English nonfiction 33 times, and correctly classified as Italian nonfiction 81 times. Similarly, English nonfiction-text were misclassified 19 times as Italian nonfiction. Even if these results are still significantly higher than chance (for four classes, 25 percent), the classifier cannot easily distinguish English nonfiction from Italian nonfiction. The similarity of English and Italian nonfiction may be attributed to the fact that the style of texts in *Translation Myself and Others* composed in English is, as shown in fig. 1, more alike the minimal style of Italian nonfiction than English nonfiction. Instead of thinking of nonfiction as 2 classes, one for English nonfiction and one for Italian fiction, it would be more fitting to think of it as 3 classes, one for early English nonfiction (2008-2011), written when Lahiri was living in the United States, one for Italian nonfiction (2016), written while Lahiri was living in Italy, and one for later nonfiction (2022) written when Lahiri *returned* to the United States. In fact, when looking at the (mis)classifications in each round, it is found

that only the 2022 work *Translating Myself and Others* was responsible for the misclassifications in both its language-versions<sup>20</sup>. The essays in this collection composed in English are misclassified as Italian nonfiction, and the essays composed in Italian are misclassified as English nonfiction. It would suggest that in this 2022 work, Lahiri minimal style in English is not as different to her minimal style in Italian as it was before 2022.

To see what words and features are responsible for stylistic difference between Lahiri's L1 and L2 texts, I examine the centroids computed for these two classes with the distinctive function words for each class. Note that the full vocabulary is not examined here, but only differences in use of the selected function words and features among the 600 MFF, like in the above cluster analysis and classification. As noted, the NSC classifier may show which features were most distinctive for attributing the texts to the classes. Features are distinctive for a class insofar as they are used frequently in comparison to their mean frequency in the whole corpus and the other class. If a word is used more frequently than the mean frequency in the corpus in class a, and less frequently than the corpus means in class b, that word is distinctive for class a. The words are listed with distinctiveness descending (down and to the right) in tables 5 and 6.

un (It)	del (En)	a (En)	mi (It)	i (En)	lei (En)
, (It)	sono (It)	avrebbe (En)	questa (It)	me (It)	la (En)
le (En)	questo (It)	gli (En)	sempre (It)	il (En)	con (En)
era (En)	aveva (En)	si (En)	disse (En)	sul (En)	mia (It)
erano (En)	una (It)	ogni (It)	ho (It)	: (It)	ma (It)

*Table 5 and 6:* Topmost 30 distinctive words for Lahiri's English and Italian texts. In the table above, the words were taken from the NSC classification performed in English, and below in Italian. Words that are frequent in English texts but infrequent in Italian texts are indicated by (En). Words frequent in Italian but infrequent in English texts are followed by (It).

had (En)	up (It)	I (It)	me (It)	even (It)	would (En)
could (En)	that (It)	one (It)	her (En)	is (It)	his (En)
are (It)	like (It)	get (It)	say (It)	were (En)	can (It)
my (It)	out (It)	said (En)	until (It)	his (En)	says (It)
was (En)	about (It)	this (It)	not (En)	every (It)	- (En)

<sup>20</sup> See the script for inspecting misclassifications at: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

Both when looking at the texts in Italian and in English (table 4 and 5), there are notable patterns in texts originally written in English to texts originally written in Italian. It is clear that first person pronouns (*my, I, me, mi, me, mia*) are more distinctive for Italian texts, while third-person pronouns are more distinctive for English texts (*her, his, lei*). It indicates that Lahiri tends to write more from third person perspective or about other people in English, and from a first-person perspective, about herself or rather, the narrative self, in Italian. The conjugation of verbs in the Italian also confirms this tendency, where verbs conjugated in the third person singular (*era, aveva, avrebbe, disse*) are distinctive for Lahiri's English texts, while the first person singular (*sono, ho*) are distinctive for her Italian texts. Moreover, it seems that the past tense (*had, could, was, said, were, era, aveva, disse*) is more distinctive for English fiction, and the present tense (*are, get, say, is, sono, ho*) is more distinctive for Italian fiction.

In sum, the differences in use of function words and features when looking just at the 600 MFF, what I have called the minimal style, differs between Lahiri's Italian and English writing, both when looking at cluster analysis and classification. Moving on to examine style in a broader sense, these findings should be kept in mind.

## Vocabulary

In the analysis with Zeta I consider a larger vocabulary than previously examined with the NSC classifier, comparing distinctive words in Lahiri's texts composed in English to those in texts composed in French (both fiction and nonfiction). I perform Zeta on the corpus of texts in English and on the corpus in Italian, keeping in mind that the corpus of texts in Italian is substantially smaller than the one in English. As noted (cf. methodology), a Zeta-score of 0 indicates that a word was used in a similar way in both of the groups. A higher score indicates that a word appeared more in one group compared to the other. Zeta lists a lot of words distinctive for each group in the case of Lahiri (with scores higher than a threshold set at 0,2), which indicates that vocabulary differs substantially between her L1 and L2 writing.

Some of the differences in vocabulary between English and Italian texts here are similar to what was seen when examining the distinctive features identified by the NSC classifier before (table 5 and 6). That is, first person pronouns predominate in Lahiri's Italian writing (*I, myself, mine, me, io, miei, mia, mie, mio, me*), and third-person pronouns in her

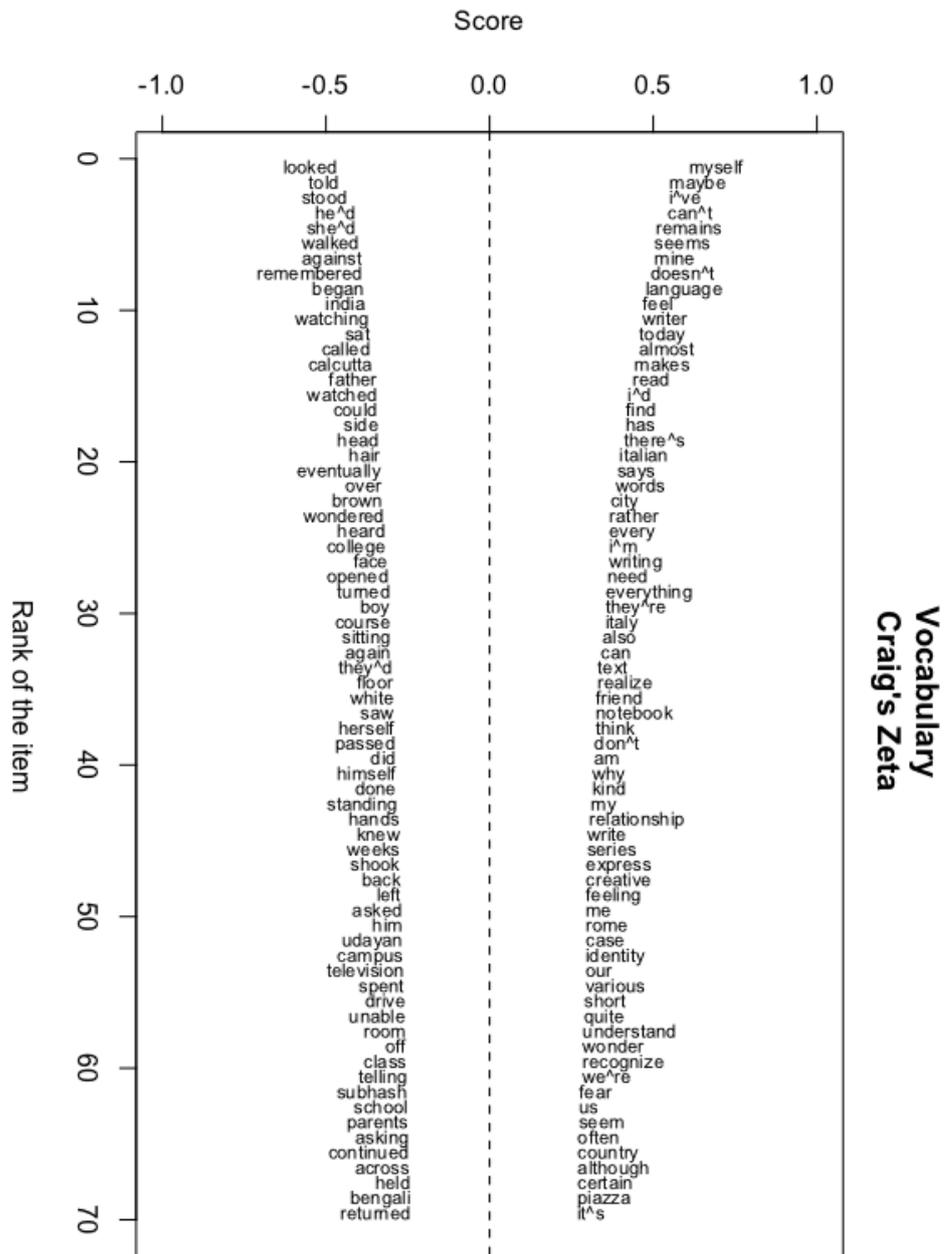
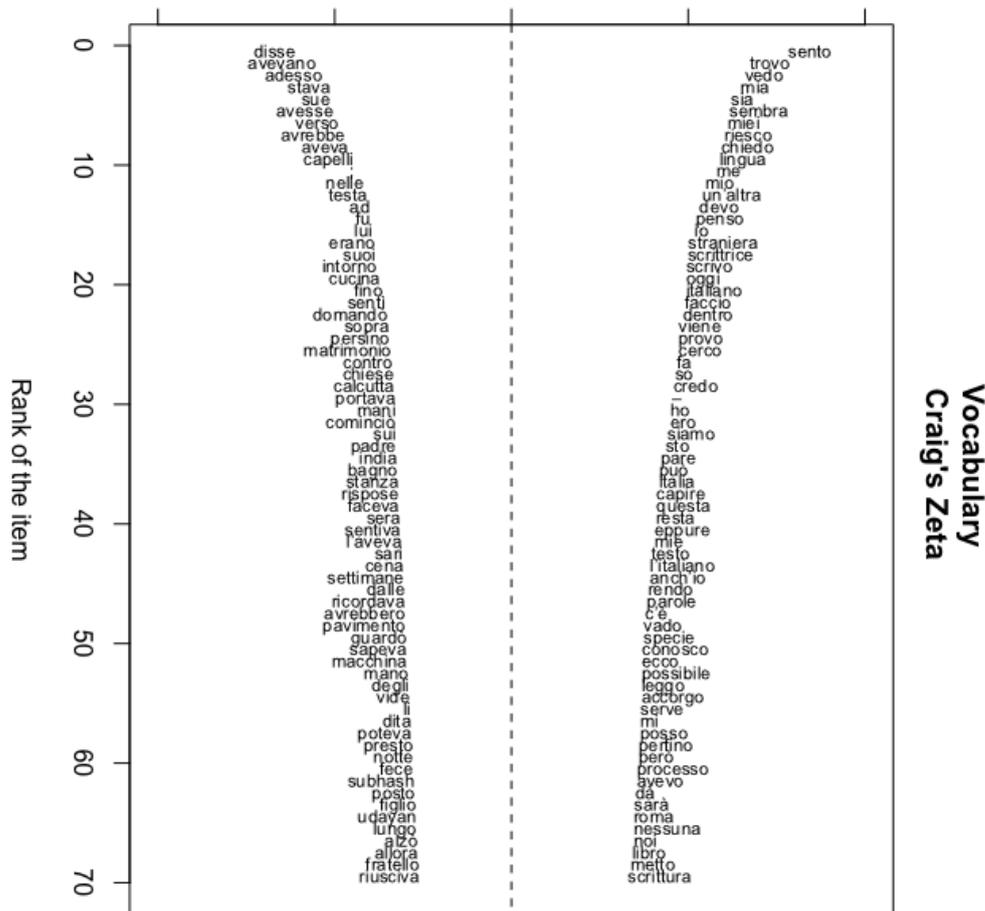


Figure 3: Zeta analysis of vocabulary contrasting distinctive words in Lahiri's L2 (Italian) texts against her L1 (English) texts, analysis performed in English.

English (he, she, lui, sui, sue)(figure 3 and 4). Also as before, the difference in narrative perspective is particularly reflected by the verbs in Italian, predominantly conjugated in the third person singular in Lahiri's English texts (*avesse, avrebbe, faceva*, etc.) and in the first person singular in her Italian texts (*vedo, sento, trovo*, etc.)(fig. 4). Likewise, present tense is more distinctive of Lahiri's Italian writing, and past tense more distinctive of her English writing, as is also indicated by the conjugation of the verbs in Italian.

Another interesting difference between Lahiri's Italian and English writing is apparent when looking at the types of verbs. Motion verbs and verbs of dialogue predominate her

Figure 4: Zeta analysis of vocabulary contrasting distinctive words in Lahiri’s L2 (Italian) texts against her L1 (English) texts, analysis performed in Italian.



English texts (such as, i.a., stood, walked, watching, called, turned, passed, standing, telling, asking, *stave*, *portava*, *guardo*, *rispose*, *chiese*). In contrast, verbs of cognitive and emotional processes, but also verbs related to writing and reading, predominate Lahiri’s Italian texts (feel, realize, think, feeling, understand, express, wonder, recognize, write, read, writing, *sento*, *accorgo*, *penso*, *capire*, *processo*, *credo*, *conosco*, *scrivo*, *leggo*)<sup>21</sup>.

Nouns among the distinctive features reflect the setting of novels. Among distinctive words of English texts are nouns like India, Calcutta, Bengali, college, and campus, reflecting the often American and/or Indian setting of Lahiri’s English works such as *The Namesake*. Lahiri’s Italian works are often set in Italy, with distinctive words such as Italy, Rome,

<sup>21</sup> It should be noted that because Lahiri’s L1 and L2 writing prefer different tenses, the usage of verbs is not properly represented in the table. For example, “walked” might stand out in the English works, because the present tense of the verb (i.e., “walk”) is used by Lahiri in Italian. However, if motion verbs were used similarly in each group, we should see more motion verbs in the present tense in Italian texts. Moreover, the focus is here on differences in the predominant *types of verbs* in each language, not on assessing the frequency of any one verb.

piazza, *Roma*. Moreover, when looking at nouns among the distinctive words, these frequently relate to people in her English texts (father, head, hair, boy, hands, parents, *padre, capelli, testa, mani, fratello*), while the nouns in Lahiri's Italian texts indicate a greater focus on writing (language, writer, words, text, notebook, *lingua, scrittrice, parole, libro, scrittura*), emotions and identity (relationship, feeling, identity). We might thus say that Lahiri's English works are focused on describing actions and people in the past tense, while her Italian works more introspective, focused on writing, reading, and emotions. This difference in focus between L1 and L2 texts is important since it is very much aligned to Hanauer's suggestions (HP4 of this thesis), that L2 writing should be more personal, contrary to the way L2 writing may be viewed as a more distant or impersonal language for expression in the monolingual paradigm (cf. theory).

In sum, there is a clear difference in preferred vocabulary between Lahiri's L1 and L2 writing, which reflects differences in subject matter, but also in style in the expanded sense, such as narrative perspective. Looking now at particular quantitative features of style in the expanded notion, I turn to the analysis of lexical diversity and sentence-length, to see whether Lahiri's L1 and L2 writing displays differences here.

### Lexical Diversity and sentence length

As noted, to examine the idea of L2 deficiency, I look at quantitative measures often applied in a linguistics and SLA context. The following table summarises the lexical diversity (MSTTR) of Italian and English texts examined in their original language. It should be noted that these are MSTTR for both fiction and nonfiction. Importantly, the lexical diversity of Lahiri's texts is not dependent on genre. In fact, her English nonfiction is responsible for both the highest and lowest MSTTR among her works originally written in English.<sup>22</sup>

---

<sup>22</sup> For the MSTTR of each text individually, see the detailed table at: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

	Analysis in Italian	Analysis in English
	Italian texts	English texts
mean	0,78	0,72
median	0,77	0,72
max	0,80	0,75
min	0,76	0,70
standard deviation	0,02	0,01

Table 6: MSTTR-100 values for Lahiri's English and Italian texts analysed in their original language. Max. and min. indicate the highest and lowest MSTTR of texts in a group. Standard deviation indicates the dispersion of values. If standard deviation is low, that indicates that the MSTTR of texts in a group was close to the mean, while a higher standard deviation indicates that the MSTTR of texts in the group was more dispersed.

Looking at the MSTTR for English and Italian texts in their original language, it is notable that the MSTTR are much higher in the Italian works compared to the English, both when looking at mean and median MSTTR. Moreover, the distance between the minimum and maximum MSTTR is smaller for texts composed in Italian, so that they are consistently richer in vocabulary compared to the texts composed in English, with a bigger variation between the min. and max. MSTTR.

The difference in MSTTR of 0.6 between the mean of Lahiri's original Italian texts (0.78) and her original English texts (0.72) may not look like much. Yet it should be remembered that the bigger part of any text is function words which are used frequently and repeatedly. As noted, an MSTTR-100 of 1 would indicate that all words in 100-word segments were unique words. MSTTR values should therefore be compared to other findings by the same method. For example, to the MSTTR-100 of a work like Boccaccio's *Decameron*, the first book of which Torruella and Capsada have shown to be 0.71, or to essays of French as a third language learners, which Lissón and Ballier found to have a mean MSTTR-100 of 0.64 (Torruella and Capsada 450; Lissón and Ballier 25). In literary texts, it would be rare to find an MSTTR below 0.6 (comparable that of the learner of a third language), if it is not a stylistic device, or above 0.9 (the highest MSTTR-100 of a text in Lahiri's corpus is 0.8). That is to say that even small differences within the marginal 0.6 to 0.9 are significant, and that a difference of 0.6 is, in fact, the difference between a third-language learner and Boccaccio. Lexical diversity is thus significantly higher in Lahiri's L2 Italian, than in her L1 English.

To further examine the idea of L2 deficiency (HP1) in Lahiri’s case, I compare sentence lengths of L1 and L2 texts. As noted, sentence length is measure by running a script in R, where the mean sentence length of each text is measured in each language. The figures 3-4 show Lahiri’s corpus arranged by the mean words per sentence (y-axis) and year of publication (x-axis).

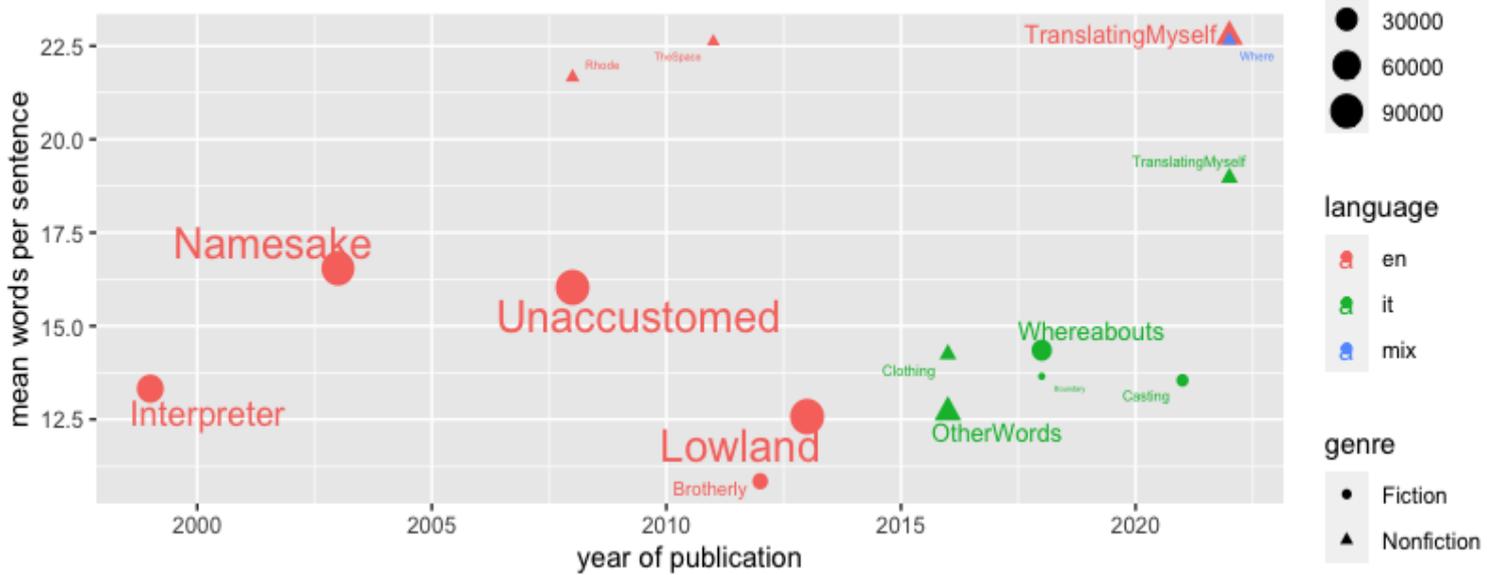
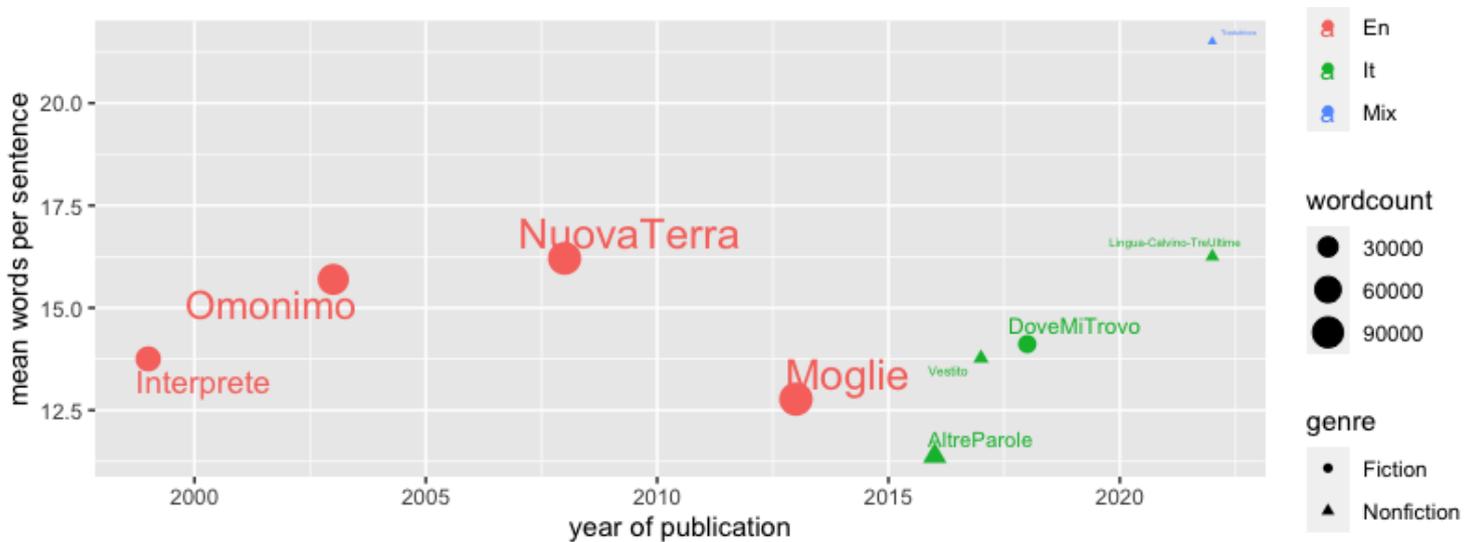


Figure 5-6: Scatterplots of median sentence length per text (y-axis) and year of publication (x-axis). The analysis of works in English is set above and in French below. Colours indicate the original language of composition, size indicates total length of text (in words), and shape indicates genre (fiction or nonfiction).



While sentence-length initially seems to vary between English and Italian texts, if we look at the mean sentence-length, the variance between Lahiri's L1 and L2 texts is not very big. In the English-language analysis, it is a mean of 15,3 words per sentence for works written in English and 14,6 for works originally in Italian. Respectively, in the Italian-language analysis, that number is 13,9 for works originally in Italian and 14,6 for those originally in English. Thus, works in Italian are in average a little less than one word shorter per sentence.<sup>23</sup>

Still, sentence length has seemed to decrease in Lahiri's stylistic evolution (toward 2015) and then increased (toward 2022). This curve may, however, be considered an evolution of Lahiri's style over time (rather than languages) since there is no significant drop when Lahiri starts to write in Italian. Sentence length decreases from *Unaccustomed Earth* (En) over *The Namesake* (En) toward *In Other Words* (It) in both language-analyses (fig. 3-4). Yet, it is notable that the two versions of *Translating Myself and Others* (2022) vary significantly in sentence-length. The smaller of these (green) is, as noted, originally written in Italian, while the larger text (red) is originally written in English. Between the two is a difference in mean words per sentence of 3,8 words. That is, the texts included *Translating Myself and Others* that were originally written in English employ significantly longer sentences. It amounts to about a 19 percent increase in average sentence length for the texts of *Translating Myself and Others* originally written in English compared to the texts originally written in Italian in the same collection. Importantly, too, the short essay "Where I find myself" composed in both English and Italian is here coincident with the sentence-length of the English essays of the same collection, in contrast to what was seen in the cluster analysis (fig. 1-2), where it occupied a middle-position between the minimal style of English and Italian works. This implies that while "Where I find Myself" is neither fully in Lahiri's Italian nor fully in her English minimal style (i.e., function words and features among the 600 MFF, fig. 1-2), in terms of sentence length, it is more similar to English-language works.

Another important observation is that particularly when looking at the English-language analysis (fig. 5) with a larger corpus, we see that sentence-length of Lahiri's English writing is very much dependent on genre. Her English nonfiction employs significantly longer sentences than does her English fiction. Curiously, in Italian, there is no such difference between genres (fig. 5-6). As can also be seen, the English nonfiction "The Space

---

<sup>23</sup> A detailed table of mean sentence-length of each text and of the mean in L2 and L1 texts can be found at: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

between the Pictures” and “Rhode Island” have much longer sentences than her English fiction. While her English works differ according to whether she is writing fiction (shorter sentences) or nonfiction (longer sentences), in Italian, nonfiction and fiction are more similar in terms of sentence-length, again, before 2022. In fact, both Italian fiction and nonfiction are much more similar to Lahiri’s English fiction in terms of sentence-length. It indicates that while Lahiri has a relatively consistent style across genres in Italian, in comparison to her English fiction/nonfiction style, where sentence-length is more varied.

### Summing up Lahiri

It has been found that Lahiri’s writing style varies according to language of composition, both when looking at her minimal style (usage of function words and features among the 600 MFF), and when looking at a larger notion of style, considering vocabulary, lexical diversity and sentence-length. Firstly, cluster analysis suggested a clear division of Lahiri’s *oeuvre* along the language-line (fig. 1-2), with the anomaly of her latest work *Translating Myself and Others*, which seemed to occupy a middle-position between her English and Italian writing in terms of minimal style. Classification underpinned this stylistic discrepancy between Lahiri’s English and Italian writing, showing again that *Translating Myself and Others* stands out stylistically and is not easily classified as either belonging to her English or Italian style. That implies that Lahiri’s style changed when “returning” to English after having written in Italian. The examination of vocabulary, both of the function words and features within the 600 MFF that the NSC classifier identified as distinctive of English or Italian texts, as well as the total vocabulary, showed that Lahiri tends to prefer third-person pronouns, past tense, and to describe actions and people in her English writing. In contrast, in her L2 writing (Italian), Lahiri tends to prefer first-person pronouns, verbs of cognitive and emotional processes and the topic of writing (table 5-6, fig. 3-4). That Lahiri’s L2 writing should employ the first-person narrative perspective more supports parts of HP3, that L2 writing is personal in style.

It is important to highlight how this analysis suggests that Lahiri’s L2 writing should not be considered deficient in relation to her L1 writing. There is no significant variance in sentence length between L1 and L2 works (fig. 5-6). While Lahiri’s Italian works are in average <1 word per sentence shorter than English works, the difference in sentence-length between genres in her English texts is bigger than the difference in sentence-length between her L1 and L2 writing. Also, in terms of lexical diversity, it has been shown that Lahiri’s

texts originally written in Italian are actually more lexically diverse than her texts originally written in English. Findings in Lahiri's case therefore rebut HP1, that her L2 writing should be less lexically diverse and employ shorter sentences.

It should be noted that many of Lahiri's Italian works (*In Other Words*, *Translating Myself and Others*) are reflections upon learning and writing in Italian, and so would naturally contain more first-person perspective, certain content words, and possibly more quantifiers, merely because of the subject. On the one hand, Lahiri's subject of choice might be seen to bias the results. Yet, on the other hand, we might say that learning and writing in Italian prompted Lahiri to reflect on self, writing, and her writerly relation to language, in a way she would never have done had she written only in English. Even if subject matter presupposes a certain vocabulary and narrative perspective: what if the language presupposes the subject matter on a higher level? That Lahiri should prefer certain topics in her L2 in comparison to her L1 seems to support HP6, that authors should prefer certain subjects in one language in comparison to the other. The case of Beckett is particularly apt for further pursuing this question, since his literary output (as of yet) spans a wider period, where he switched frequently between his L1 and L2. If language indeed presupposes certain vocabulary, narrative perspective and subject matter, this should be clear when comparing all of Beckett's works written in English to all of his works written in French because of the considerably larger corpus, spanning a period from 1931 to 1981.

Moreover, the difference between Lahiri's English writing and her Italian writing as shown here, might also be construed as simply an evolution in style. A way to examine L2 writing and minimising the effect of time, would be to look at a single work composed in two languages and written around the same time. Yet, the only text of Lahiri's which we know to be composed in both languages, is "Where I find Myself". It would be opportune to examine this text for language-specific stylistic variance, but we do not know where in this text Lahiri switched language of composition, so that such an analysis would be mere guesswork. However, in the case of Beckett we know when changes of language of composition occur. That is why, in the following, I begin by examining Beckett's oeuvre in the same way as Lahiri's, but then turn to examine two dual-language compositions by close reading.

## 5 Samuel Beckett Analysis

How do we determine whether the difference in style between L1 and L2 writing as seen in the case of Lahiri is not simply to be attributed to the author's stylistic evolution over time? In other words, how do we untangle the effect of time and composition-language on style? One way to do this is to look at a much longer period of bilingual composition to see whether there is a clear evolution, such as is opportune in an *oeuvre* like Beckett's, who wrote in both his L1 and L2 over the span of 50 years. Another way to do it is to look at a very local example, such as within one work written in a shorter period of time. In the present analysis chapter, I do both.

Beckett is an interesting case because there is a great amount of secondary literature both about his style, its evolution across his *oeuvre*, and his bilingual writing (i.a., Beer, Coetzee, Slote, Hulle and Kestemont). As noted previously, Beckett said that he used L2 writing as a way to "impoverish" his style, implying that he actively sought to change his style. Others have pointed out that writing in two languages enabled his literary creativity and ability for artistic self-renewal, and that there is great stylistic evolution from his early works, such as *More Pricks than Kicks* to his later works such as *Worstward Ho*<sup>24</sup> (Beer, Slote). Would Beckett not have been as stylistically versatile if he had not written in both English and French, as Beer suggests (Beer 215)? Turning to studies of the chronological evolution of his style will be helpful in untangling the effects of language and time in Beckett's stylistic evolution, to examine how temporal stylistic variations relate to changes in composition language.

There have been various attempts at periodising Beckett's stylistic evolution (see Hulle and Kestemont's review). With their analysis by computational stylometry, Hulle and Kestemont show that time is major factor for a stylistic periodisation of Beckett's *oeuvre*, so that works close in time (by composition date) appear related in a stylometric dendrogram –

---

<sup>24</sup> The use the English titles of Beckett's works is purely arbitrary. In many cases, it is next to impossible to determine any precedence of either the French or English version. For example, consider a work like "Company", which was written in English but revised in French even during composition (Beer 219); where the French version was published first, and the English version subsequently revised according to the French (Slote 124).

something that might *seem* obvious but is not necessarily a given. Hulle and Kestemont found there to be three major stylistic time periods, divided by what they identified as major stylistic turning points (193). These differ slightly when considering Beckett's *oeuvre* in English and in French, insofar as *The Unnamable* (1949) and *From an Abandoned Work* (1955) are identified as major stylistic turning points (in both French and English), while "First Love" (1946) is only identified in the French, and *Watt* (1941) only in English language analysis (191). Importantly they note that there is a question as to whether *From an Abandoned Work* should be considered a turning point marking the beginning of Beckett's late-style period, or if it is, rather, a "re-turning point" (193). Thus, Hulle and Kestemont suggest that Beckett may have returned stylistically to his "old" style of 1931-1941 (197). Thus, their division of Beckett's *oeuvre* into stylistic periods clearly already considers language of composition as a factor. That this work should stand out is particularly interesting for the present study, since it was written in English during a period when Beckett wrote predominantly in French, and the first text composed in English since *Watt*, fourteen years prior. It suggests that Beckett had an early "English" style which he returned to when writing in English once more. As Hulle and Kestemont note, "[i]n this sense, the stylometric analyses also suggest... the effect of bilingual writing in Beckett's stylistic evolution" (Ibid.).

Yet, the effect of bilingualism in the stylistic periodisation of Hulle and Kestemont is only noted in passing, while it is the focus of the present analysis. There are various reasons to note especially Hulle and Kestemont's study as a point of departure for the present analysis. That is, since their approach is similar to mine, since their recent publication attempts to synthesise previous scholarship and because keeping their findings about the temporal evolution of style in mind, it is possible to ask whether there is any L2 effect on style discernible in Beckett's *oeuvre*, beyond what they found to be a chronological stylistic evolution.

Thus, for the analysis of Beckett's *oeuvre*, the following division by the stylistic turning point of Hulle and Kestemont is considered: an early style from 1931-1941 (which coincides with a period where Beckett exclusively wrote in English), a style from 1946-1949 (which coincides with a period where Beckett exclusively wrote in French), and a style from 1949 (*The Unnamable*) and forward. The latter stylistic period spans over the period where Beckett wrote in both English and French. Since I am interested in stylistic differences between texts composed in French and in English, it will be of significance whether the computational analysis suggests a division of the latter period according to the language of composition. Notwithstanding, it is not *a priori* assumed that a division in the latter period

should be discernible. It should also be noted that Hulle and Kestemont do not consider Edith Fournier’s translation of *More Pricks than Kicks* (*Bande et Sarabande*, 1995), whereby they exclude a key early work written in English, even if not translated into French by Beckett himself. Moreover, there are works included in their analysis of less than 4,000 words as single units (i.a., “Imagination Dead Imagine”), a setup that is, as previously discussed, not necessarily reliable. As such, the suggested stylistic periodisation of Beckett’s *oeuvre* by Hulle and Kestemont will be guiding in the following analysis, but room should be left for different results as the methodology, corpora, and parameters differ (in the following, 600 MFF rather than 300 MFW). Ultimately, the full corpus considered for in the present analysis – marked with the suggested stylistic turning points – is listed in Table 1. *From an Abandoned Work* is marked, even if not considered a turning point.

Title (EN)	Abbrieviated title (En)	Abbrieviated title (Fr)	Start date	Composition language	Wordcount (En)	Wordcount (Fr)
More Pricks than Kicks	Pricks	Bande	1931	en	63845	70291
Dream of Fair to Middling	Dream	NA	1932	en	84683	NA
Murphy	Murphy	Murphy	1935	en	68819	72640
Watt	Watt	Watt	1941	en	99232	99450
The End	End	Fin	1946	fr	10058	10055
The Expelled	Expelled	Expulsé	1946	fr	6459	6427
The Calmative	Calmative	Calmant	1946	fr	7481	7897
Mercier and Camier	Mercier	Mercier	1946	fr	34050	42047
First Love	Love	Premier	1946	fr	9712	10351
Molloy	Molloy	Molloy	1947	fr	90663	101097
Malone Dies	Malone	Malone	1948	fr	57016	56708
The Unnamable	Unnamable	Innomable	1949	fr	71351	70697
Texts for Nothing	Nothing	Textes	1951	fr	20379	20543
From an Abandoned Work	Abandoned	Abandonné	1955	en	4307	4480
How it is	How	Comment	1958	fr	36552	36208
Fizzles	Fizzles	Foirades	1960	fr (excl. "Still")	7175	7405
All Strange Away	AllStrange	NA	1963	en	3761	NA
Imagination Dead Imagine, Enough, Ping	Imag-Enough-P	Imag-Assez-Bir	1965, 1966	fr	4873	4907
Company	Company	Compagnie	1965	en	10664	11175
The Lost Ones	Lost	Depeupleur	1970	fr	8661	8472
Ill Seen Ill Said	IllSeen	MalVu	1976	fr	9201	9345
Wortsward Ho	Worstward	CapAuPire	1981	en	5661	6566

Table 7: Beckett’s oeuvre (chronological by year of start of composition). The stylistic turning points as identified by Hulle and Kestemont are in grey. Note that the start dates of composition are approximations and are those supplied in Hulle and Kestemont (180). Note that *Dream of Fair to Middling Women* and “All Strange Away” only exist in English (marked NA) and will only be considered in the English-language analysis.

A few of Beckett's prose works have been excluded here, such as "Lessness" (1969), and *Stirrings Still* (1983) since these pieces are shorter than 4,000 words (cf. methodology). Note that I have chosen to treat "Imagination Dead Imagine", "Enough", and "Ping" jointly because they are short, but also seeing they were published together in a collection (*Têtes-Mortes*) with their date of composition close in time ("Imagination Dead Imagine" written in 1965, "Enough" and "Ping" in 1966). Note also, that for the same reason *Texts for Nothing* and the *Fizzles*, consisting of eight prose pieces, are treated as jointly. The piece "Still" may be supposed to stand out stylistically since it is written later than the rest (1972, not 1960). However, this piece also warrants a separate treatment in the end of this analysis because it was written in English in contrast to the rest of *Fizzles*.

In the following section, I look at stylistic variance both across the whole Beckett *oeuvre* according to a minimal and expanded notion of style, before tuning to look at the single texts "The End" and *Fizzles* and to examining whether a change of composition language from L1 to L2 is stylistically identifiable *within a work*. In the latter analysis of single texts, stylistic variance due to time should be minimal, especially in the case of "The End", because the English and French part of this story were written in immediate continuation of each other. Therefore, if there is any language effect, it should be isolated from the influence of time on style and stand out more.

## 5.1. Distant reading

### Cluster analysis

As noted previously, this analysis considers only the defined function words and features within the 600 MFF. That leaves 337 words for the French language analysis and 259 for the English.<sup>25</sup> Having thus cleared out a very Beckettian vocabulary of content words (i.a., *yeux, vie, seul*), the cluster analysis of function words in Beckett's *oeuvre* in French and in English shows the following clusters (Figures 4-5):

---

<sup>25</sup> See full lists of analysed and excluded features at: [https://github.com/PascaleFMoreira/Bilingual\\_writing](https://github.com/PascaleFMoreira/Bilingual_writing)

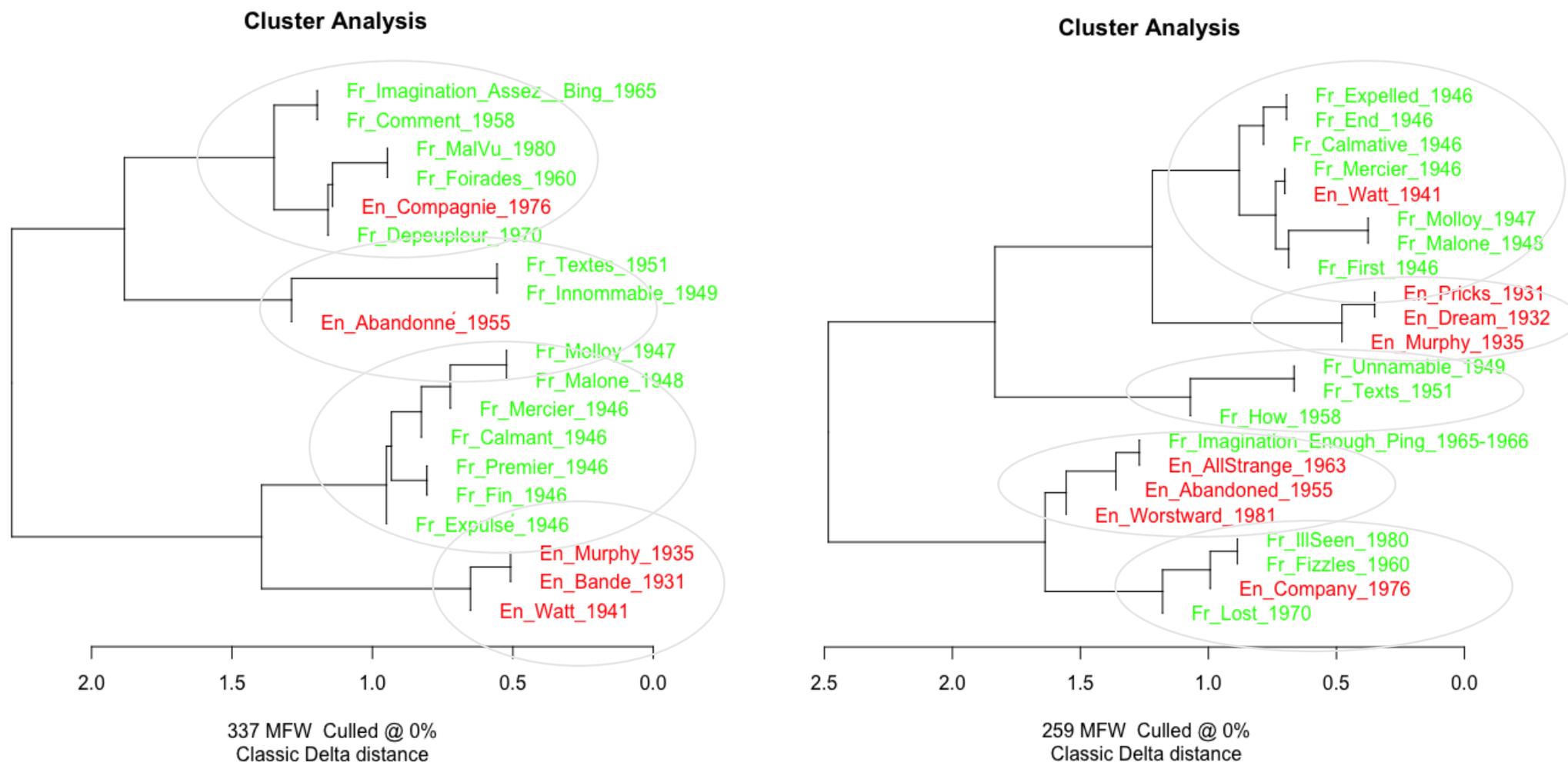


Figure 7-8: Cluster Analysis of the English (right) and French-language corpus (left) visualized in dendrograms. Reading the dendrogram from right to left, the most similar works are to the rightmost side and decreasing toward the left, so that, e.g., *Worstward Ho* and *Stirrings Still* are the most dissimilar to the cluster of *More Pricks than Kicks*, *Dream* and *Murphy* in the English language analysis. The language of composition of works is indicated by either “Fr” or “En”, and colours also refer to language of composition (green, French and red, English). Note that the number of function words and features among the 600 MFF were 337 in the French-language analysis and 259 in the English-language analysis.

In both dendrograms there are *two main branches* with four or five sub-branches (indicated by my circles). *The first branch* holds early-period English texts, middle-period French texts. In the English language analysis (fig. 8), texts from 1949 to 1958 also belong to this branch, even if at a removed position. *The second branch* holds a later group of texts. In the French analysis (fig. 7), texts from 1958 and forward, and in the English analysis (fig. 8), texts from 1949 and forward. In the English language analysis, the three texts written in French from 1949 to 1958 belong to a cluster for themselves, but not so in the French language analysis (which has only four subgroups).

From this structure it is clear that the stylistic turning points identified by Hulle and Kestemont are important to understanding Beckett's stylistic evolution. That is, there is firstly, a cluster of early style texts from 1931-1941 (*Pricks to Murphy*, and *Bande to Murphy*), which were composed exclusively in English, and a cluster of middle-period style texts from 1946-1949 ("First" to *Malone* and "Premier" to *Malone*), composed exclusively in French. As also noted by Hulle and Kestemont, *Watt* clearly stands out, and is grouped with the middle-period in the English language analysis (fig. 8), yet with middle-period French texts in the French analysis (fig. 7). *The Unnamable (L'Innommable)* from 1949 is also indicated as a stylistic turning point here (initiating new clusters). For Hulle and Kestemont initiates this work initiates Beckett's "late style" from 1949 and forward (191).

Yet in the above dendrogram, the picture is different. There is a cluster from 1949 to 1958 in the English language dendrogram (fig. 8), and a cluster from 1949 to 1955 in the French language dendrogram (fig. 7). *The Unnamable* thus does not initiate a "late style", but a pre-late style cluster. The late-style cluster is rather from 1960 to 1980, starting with *Fizzles/Foirades. From an Abandoned Work* (1955), which was the first text Beckett wrote in English after having written in French from 1946 is, like *Watt* before, an anomaly here. It is grouped with the late-style texts when analysed in English, but with the pre-late style texts in when analysed in French, although at a removed position. The anomaly of *From an Abandoned Work* may have to do with the much later translation of this work into French (in 1967). It does, however, stand out stylistically and is not closely grouped with any other work. Here it seems that writing *From an Abandoned Work* in English was not a "return" to his old English style for Beckett, instead it suggests stylistic change also in comparison to previous French compositions.

For the purposes of the present study, the division along language line of the early (1931-41) and middle style (1946-49) are interesting, a division that is very clear in the

bottom cluster in fig. 7 and top cluster in fig. 8, though with the anomaly of *Watt* in the English language analysis.

From the cluster analysis it is also notable that English-language compositions consistently stand out from clusters in both the French and English-language dendrograms. Both *Company* and *From an Abandoned Work*, which were composed in English, are not coupled closely with any other texts, suggesting stylistic variation from their style-period-clusters. The two most distinct style-groupings, the early works in English and middle-period French works, may suggest that Beckett had a language-specific style and warrant further examination.

### Classification of language-style groupings

As noted (cf. methodology), a way to examine the strength of these groupings is by training a classifier and looking at the accuracy of classification. However, the groupings in Beckett's bilingual *oeuvre* are more complex than in the case of Lahiri, where we saw a clear division along the language-line. Beckett's stylistic versatility and the dramatic changes in his style as seen in fig. 7 and 8 makes a simple division along the language line improbable, if not impossible.

However, there are two distinct language-style clusters. That is, firstly, the cluster of early style composed in English consisting of *More Pricks than Kicks* and *Murphy* (excluding *Watt* here for the present moment) which form a cluster in both dendrograms, with the untranslated *Dream of Fair to Middling Women* included in the cluster in the English analysis. Secondly, there is the productive middle style texts composed entirely in French 1946-1948 (again excluding *Watt* for the present) from "First Love" ("Premier Amour") to *Malone Dies*, which forms a cluster of its own in both the French and English language dendrogram. Thus, an early-period group of texts composed in English and a middle group of texts composed in French form our language-groups of which the accuracy-strength is tested in the following analysis by classification.

As described (cf. methodology), classification in Stylo trains a model or prototype based on word-frequencies for each class, against which texts are compared. The model then "guesses" at which class a text belongs to, based on these prototypes. This analysis was performed in English and in French separately. Average *F-scores* were not as high as in Lahiri's case though still high, 0.8 for the English language analysis, and 0.81 in the French,

which would indicate that the division along the language line is not as strong in Beckett's case as in Lahiri's.

Table 8: Confusion matrix for classification in English

	Early English texts	Middle French texts
Early English texts	100	0
Middle French texts	5	295

Table 9: Confusion matrix for classification in French

	Early English texts	Middle French texts
Early English texts	100	0
Middle French texts	0	300

The confusion matrices show the total attempts at classification for each class, how many times a text was misclassified (confused) and if, what it was misclassified as. For example, early texts composed in English were misclassified 5 times in the English language analysis (table 8) as belonging to the middle-period of texts composed in French, while they were attributed to their proper class 100 times.

As is apparent from the matrices, the two classes are very clearly distinct in terms of *minimal style*. In the following section, I look at what features were identified as distinctive in distinguishing the early-period English and middle-period French texts. The words are listed with distinctiveness descending (down and to the right) in tables 10 and 11.

her (En)	: (En)	his (En)	but (Fr)	myself (En)	himself (En)
he (En)	have (Fr)	my (Fr)	me (Fr)	if (Fr)	it's (Fr)
she (En)	I (Fr)	( (En)	an (En)	perhaps (Fr)	be (En)
; (En)	of (En)	) (En)	" (En)	, (Fr)	them (Fr)
- (En)	it (Fr)	into (En)	' (En)	little (Fr)	must (En)

Table 10 and 11: Topmost 30 distinctive words for Lahiri's English and Italian fiction. In the table above, the words were taken from the NSC classification performed in English, and below in French. Words that are frequent in early-period English but infrequent in middle-period French texts are indicated by (En). Words frequent in middle-period French but infrequent in early-period English texts are indicated by (Fr).

" (En)	ou (Fr)	; (En)	mais (Fr)	même (Fr)	j'aurais (Fr)
je (Fr)	pas (Fr)	mes (Fr)	ma (Fr)	j'avais (Fr)	en (En)
me (Fr)	ne (Fr)	... (En)	lui (En)	être (Fr)	ce (Fr)
' (En)	avait (En)	moi (Fr)	plus (En)	! (En)	: (En)
- (En)	était (En)	mon (Fr)	, (Fr)	. (Fr)	eût (En)

In comparison to Beckett's early-period English texts, it is clear that first person pronouns (I, my, me, myself, *je, me, mes, moi, mon, ma*) are more distinctive for middle-period French texts, while third-person pronouns are more distinctive for early-period English texts (her, he, she, his, himself, *lui*). Just like in Lahiri's case, it indicates that Beckett writes more from the third-person perspective or about other people in English, and from a first-person perspective, about himself or, rather, the narrative self in French. Verbs are also predominantly conjugated in the third-person singular (*avait, était, eût*) in early-period English texts, and in the first-person singular (*j'avais, j'aurais*) in middle-period French texts. It is also notable that markers of dialogue (") are more distinctive for Beckett's English texts, like, it seems, all other punctuation marks than comma and period, which are more distinctive of Beckett's French texts.

In sum, the differences in use of function words and features when looking just at Beckett's use of the 600 MFF, what I have called the minimal style, differs a lot across his *oeuvre*, with interesting differences particularly between his early-period English texts and his middle-period French texts. I now move on to examine style in the broader sense, looking at vocabulary, lexical diversity and sentence-length and finally to the close reading of style in Beckett's "The End" and *Fizzles*.

## Vocabulary

Let us again zoom in on these two distinct groupings, Beckett's early-period English texts and middle-period French texts, but looking at them with a wider notion of style in mind (going beyond function words and features). What distinguishes early-period English texts and middle-period French texts? Here, I perform Zeta comparing only these two groups and not the whole corpus, as in the analysis of Lahiri's works, because Beckett's *oeuvre* is clearly more stylistically varied.

Analysing these groups with Zeta gives an insight into the distinctive vocabulary of each group, as discussed (cf. methodology). The Zeta-score is, as noted, an indicator of the consistent use of a given term in either of the groups of comparison. A Zeta-score of 0 indicates that a word had similar frequency in each of the groups, while a score of, for example, 0,7 indicates that a word appeared more consistently in one group compared to the other. Most distinctive words are shown from the top and decreasing. On the left of both

graphs are the preferred words of early-period group of texts written in English, and on the right the preferred words of middle-period group of texts written in French.

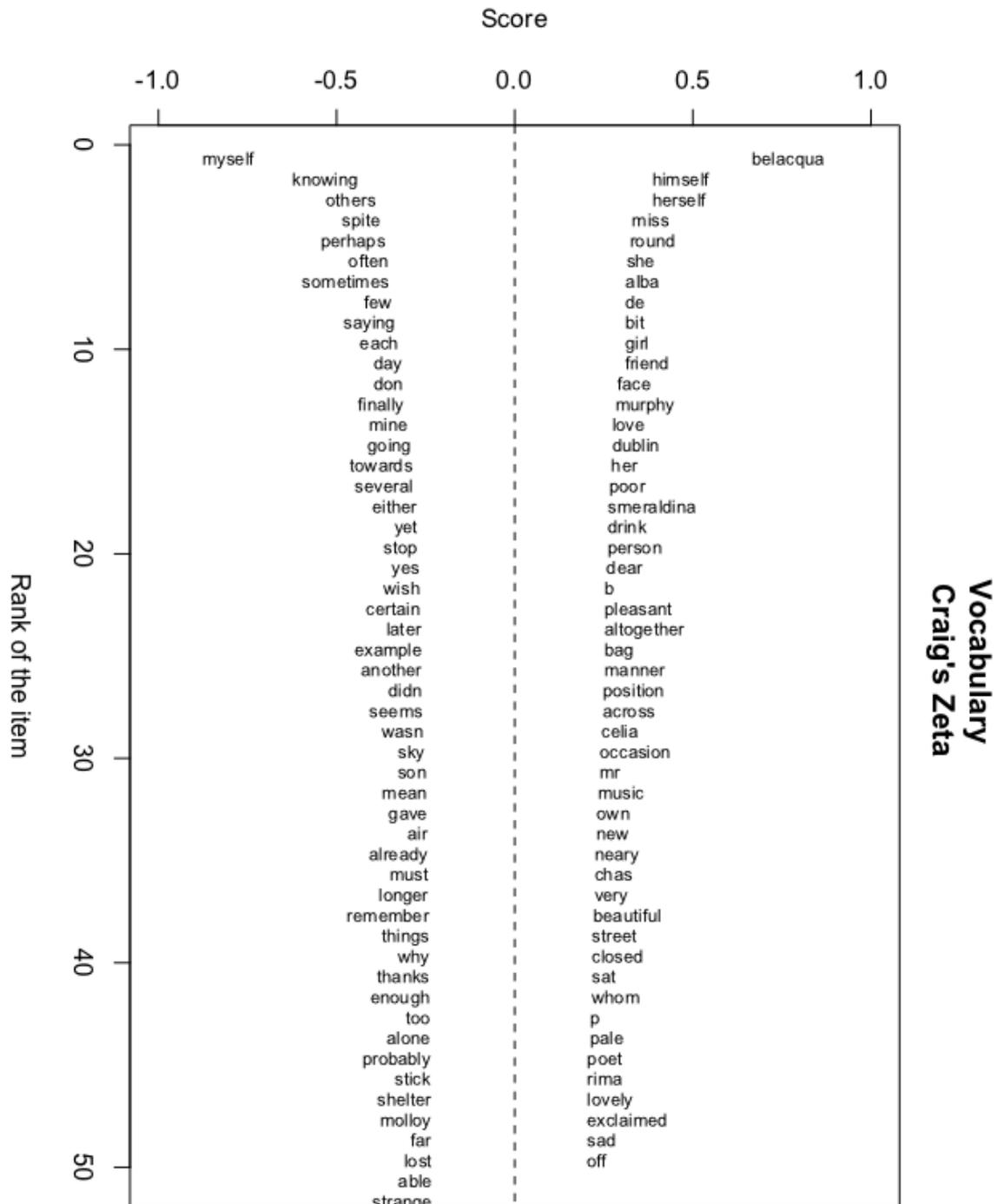
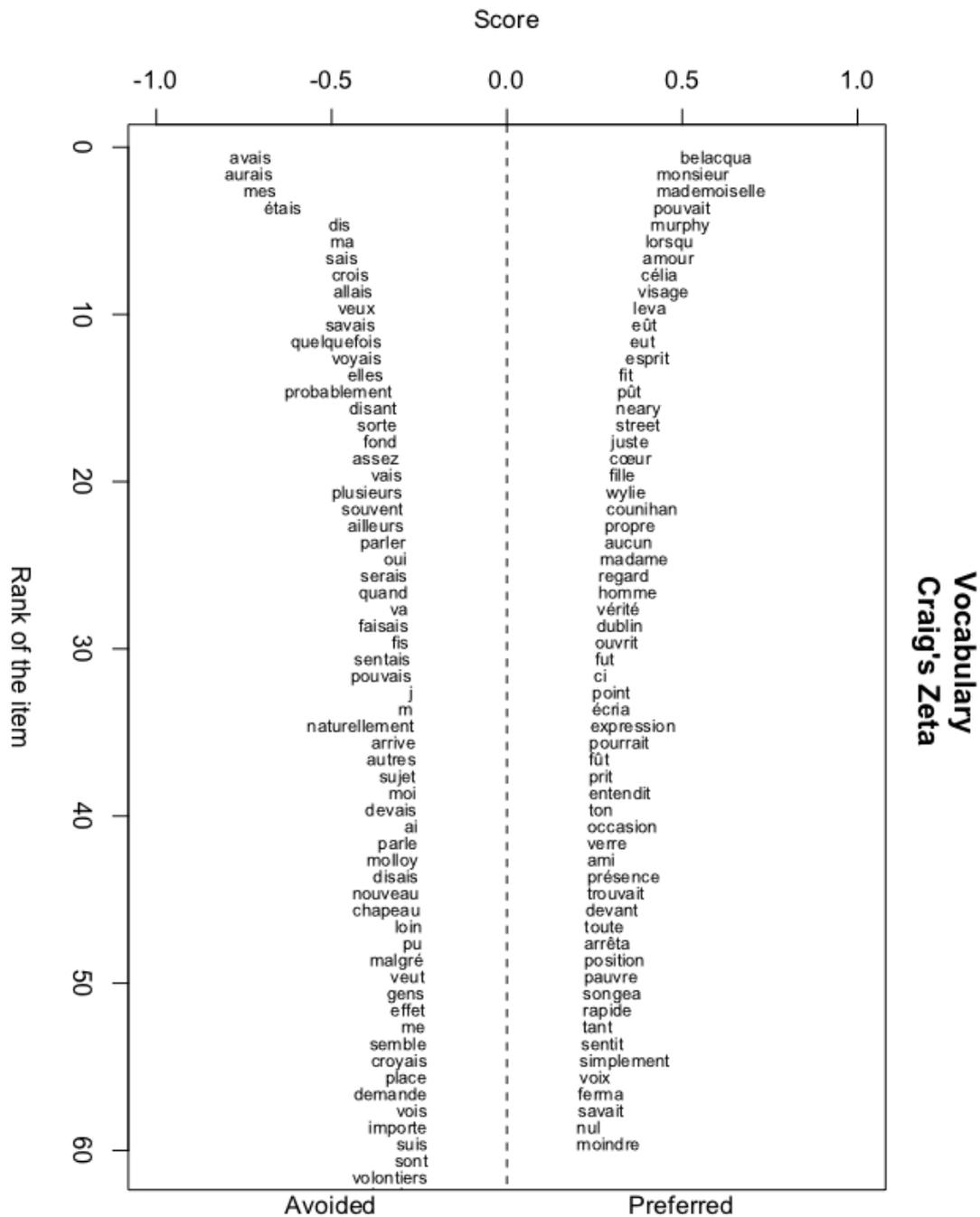


Figure 9-10: Distinctive words listed by Zeta-score. Analysis in English (top graph), and analysis in French (bottom graph). Each graph shows the words preferred by the early texts written in English (on the right) and those preferred by the middle-period texts written in French (on the left). The closer a word is to the middle line (0), the less distinctive it is. The further it is from the middle line, the more distinctive it is for the group of texts.



As can be expected, proper names rank very high on the list – predominantly in the early English texts – yet there are other important differences between these groups. When looking at vocabulary, it is clear that Beckett’s early-period English texts are much more focused on outward description and activity, of people and things. In his early English texts, Beckett uses many adjectives to describe people and things (poor, dear, new, beautiful, pale, lovely, sad)(right fig. 9), which is different from the adjectives in Beckett’s middle-French period works (alone, far, lost, strange) which can be said to be more descriptive of the self

and own experiences (left fig. 9). When looking at the corpus in French (fig. 10), adjectives are not as frequent overall. That is perhaps because *Dream of Fair to Middling Women* does not exist in French (we are looking at a smaller corpus), and because the translations of the early-period English works originally written in English, one of which is by Édith Fournier and the other by Beckett himself, perhaps use more synonyms, so that any one adjective does not rank as high in distinctiveness.

Still, it seems that early English texts are more outwardly oriented, and middle-period French texts more inwardly when comparing various word-classes between them. In both language analyses, what is distinctive for middle-period French prose is words that indicate the kind of questioning and answering one might have with oneself, such as perhaps, yet, seems, mean, remember, why, and probably, *probablement, ailleurs, malgré, semble, naturellement*, and *(n')importe* (fig. 9-10). Moreover, distinctive nouns in the early English texts also support the idea that the early English texts are more oriented to describing people and things. Nouns such as face, girl, friend, drink, bag, music, street, poet, *visage, fille, homme, ami, voix, expression* suggest a prose full of activity, characters and cityscape (fig. 9-10). In contrast, Beckett's middle-period French texts contain much fewer nouns. These are: day, wish, sky, air, shelter, *gens, fond* and *place*, which suggest much more general descriptions of setting, ambience and the elements. We could say, then, that Beckett's early English writing is focused on outward and subjective description of the social, people and activities, while his French middle-period writing is much more introspective and describing of the objective world.

The difference in orientation of Beckett's early English and middle-period French texts is further emphasised by his use of pronouns. Just like when looking at the distinctive features identified by the NSC classifiers, clearly, first person pronouns (my, me, I, *mes, ma, j('), m('), moi, me*, as well as verbs conjugated in the first-person singular) are predominant in the earlier texts written in English. In contrast, third person pronouns (her, she, his, he, himself, herself, *ton*, and verbs conjugated in the third-person singular) are distinctive for the middle-period texts written in French. Moreover, proper nouns and titles (miss, mr, *monsieur, mademoiselle, madame*) predominate the early texts in English in comparison to the French.

In fact, when considering Beckett's entire corpus, there is a similar pattern for all texts written in English in comparison to all texts written in French (figure 11-12). The scores are here closer to 0, which can be expected, since this analysis of the total corpus is forcing texts that belong to various style-periods and where vocabulary is more varied internally into two groups (texts originally written in English and texts originally written French). Still, like in

Lahiri’s case, we see predominantly first-person pronouns (I, me, mine, *mes*, *ma*, and verbs in the first-person singular) in the texts originally in French, and third-person pronouns (he, she, his, hers, and verbs in the third-person singular) in Beckett’s texts originally in English.

When (physically) looking into the works themselves, it is apparent that of Beckett’s works treated in this study, 10 out of 15 of texts composed in French are first person narrative perspective, while this is only true for 2 out of 9 of the texts composed in English. While of Beckett’s English writing, 6 out of 9 take a third-person narrative perspective, of his French writing, only one is clearly third-person narrative perspective (*Ill Seen, Ill Said*). There are, however, works where the narrative perspective is not specified or obscured, one particular case, that of “Still”, will be treated in detail in the close reading part of this chapter.

Thus, there is a clear predominance of first-person narrative perspective in Beckett’s L2 writing overall, like in Lahiri’s L2 writing. This is important, because it approximates the HP4 of this study based on the findings of Hanauer, that L2 writing tends to take a first-person perspective and write in a more personal manner. The present findings seem to replicate the effect of writing to L2 that Hanauer has suggested. Yet Hanauer also suggested that L2 writing should be more direct and descriptive, where apparently, the opposite is the case when looking at Beckett’s vocabulary. Beckett’s L2 writing is focused on the (narrative) self and objective world, rather than the social world, where Beckett’s L1 writing is more descriptive of people, things, and actions.

I now turn to look at quantitative features of style, lexical diversity and sentence-length, to examine whether the L2 deficiency idea is descriptive of Beckett’s writing.

### Lexical diversity and sentence length

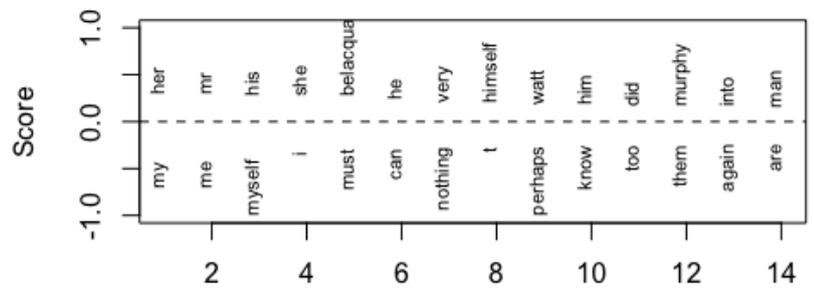
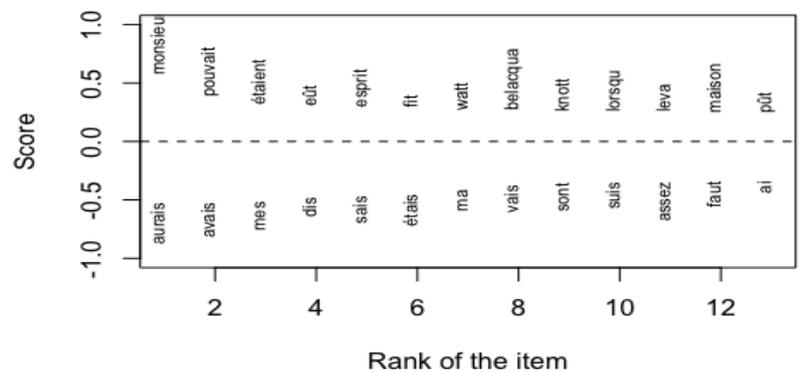


Figure 11-12: distinctive words in Beckett’s complete corpus, with texts originally written in English on the top, and texts originally written in French on the bottom, analysed in English and French.



Here again, lexical diversity is calculated for texts in L1 and L2 and compared. The MSTTR of Beckett's texts are displayed in table 6.

	Analysis in French	Analysis in English
	French texts	English texts
mean	0,71	0,66
median	0,71	0,69
max	0,74	0,72
min	0,67	0,51
standard deviation	0,02	0,07

Table 12: MSTTR-100 values for Beckett's English and French texts analysed in their original language. Max. and min. indicate the highest and lowest MSTTR of texts in a group. Standard deviation indicates the dispersion of values from the mean in each group.

As in the case of Lahiri, L2 texts seem to have a higher MSTTR. Yet, a very low mean of the English texts may be misleading in Beckett's case, since it depends on a single text, *Worstward Ho*, having a very low MSTTR (the minimum 0.51).<sup>26</sup> That one or few texts are biasing the mean is also suggested by the higher standard deviation for texts originally written in English (0.07). Therefore, the median MSTTR is more representative for the MSTTR in the groups of original English and French texts here. As can be seen, the median MSTTR between French and English original texts varies only by 0.02.

While lexical diversity is higher in Beckett's French texts, the findings in the case of Beckett are not as strong as in the case of Lahiri, where we saw a significantly higher MSTTR in her L2 texts. Although the median MSTTR for Beckett's texts originally composed in his L2 French are slightly higher, they are only so: slightly. Still, it is not lower, as has been suggested for L2 writing (HP1). Moreover, it should also be noted that both maximum and minimum MSTTR is higher in the original French texts, than it is in the original English. As already noted, Beckett's *oeuvre* spans a longer period than Lahiri's, wherefore we may expect more variance in style and thus in lexical richness. Gontarski has noted, for example, that Beckett's style changed significantly after *How it is* from which point on it was marked by minimalism, that is, to "write 'less'" and to remove ornamentation (xv), which surely had an impact on lexical diversity. Standard deviance is also higher than in the case of Lahiri (where it was in all instances <0.02) in Beckett's English writing. It suggests that Beckett's lexical diversity in English went through a stylistic evolution, and more so than in his French writing, where lexical diversity is consistently slightly higher.

---

<sup>26</sup> For the detailed table of MSTTR for each text, see: <https://github.com/PascaleFMoreira/bilingualwriting.git>

Apart from less lexical diversity, L2 writing is supposed to display shorter sentences (HP1). For Beckett's corpus, the mean sentence length (in words) is computed using a simple script.<sup>27</sup> Figures 13-14 show Beckett's *oeuvre* as arranged by the mean words per sentence and by start-year of composition.

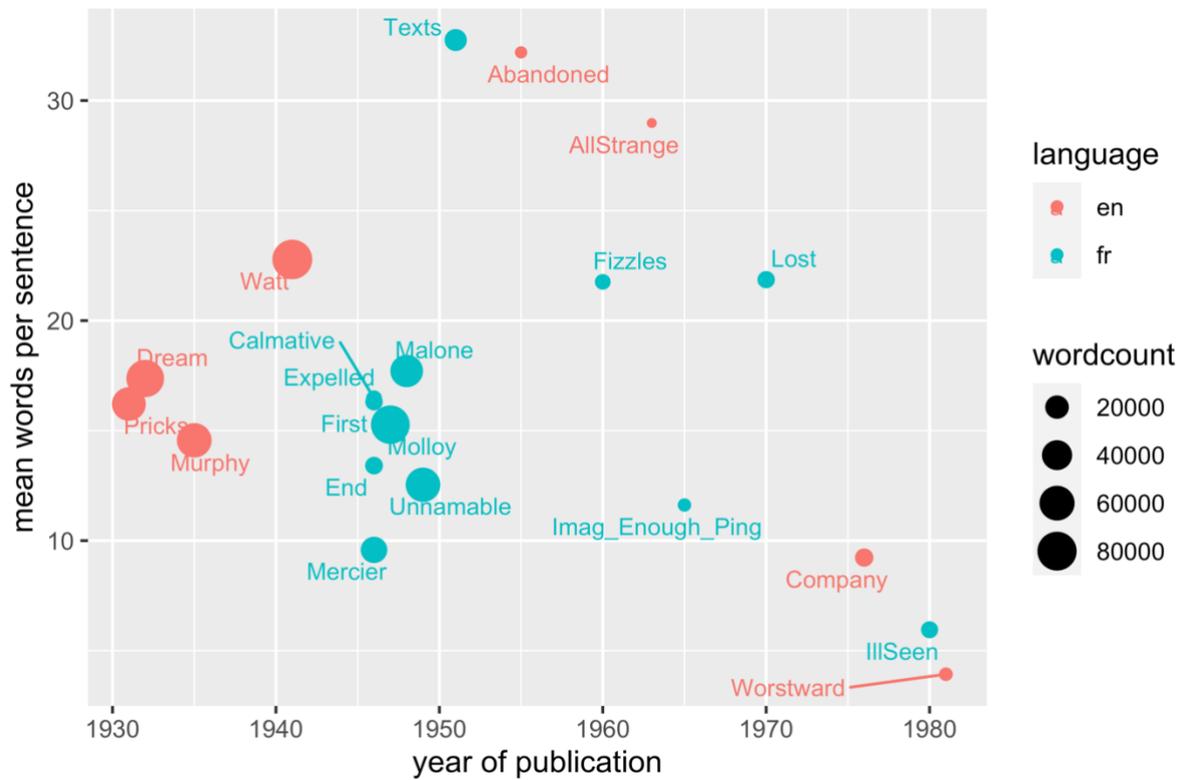
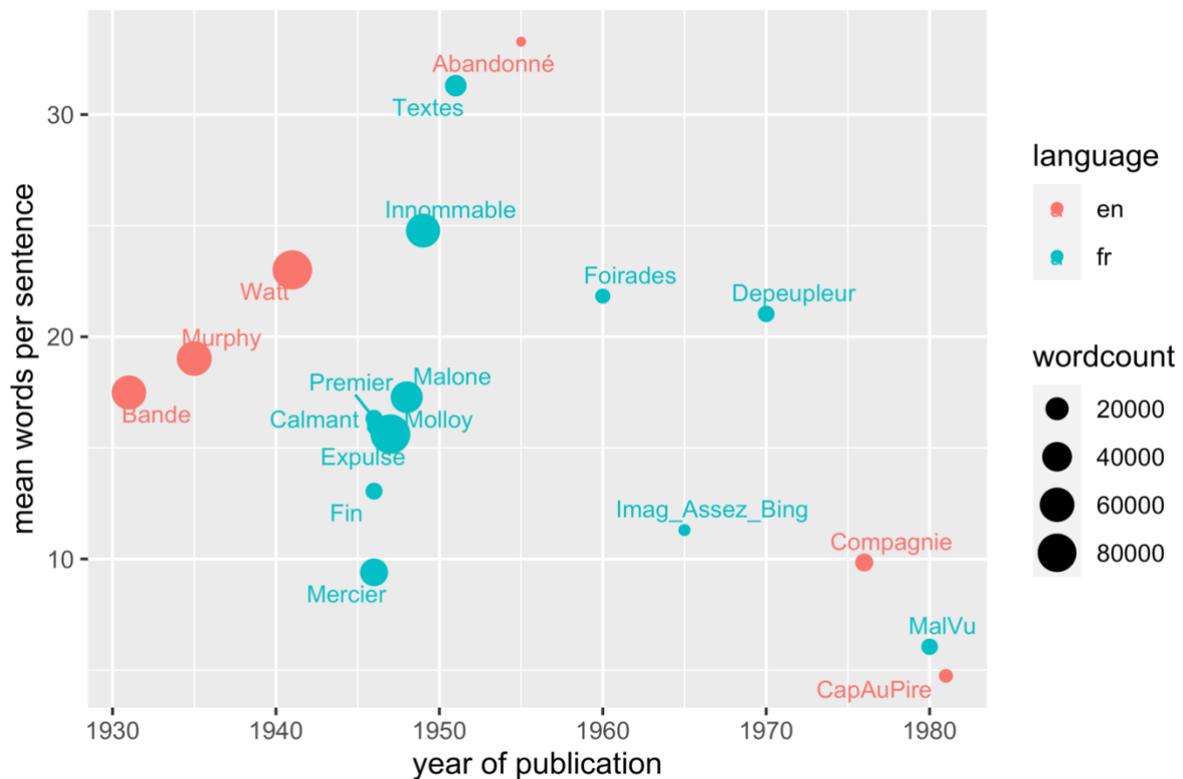


Figure 13-14: Scatterplots of mean sentence length per text (y-axis) and year of composition-start (x-axis). The analysis of works in English is set above and in French below. Colours indicate the original language of composition, and size indicates total length of text (in words).

<sup>27</sup> The scripts used in this thesis can all be found, as noted, at: <https://github.com/PascaleFMoreira/bilingualwriting.git>



As can be seen in these plots, sentence-lengths in the corpora in each language differ only slightly. One difference is the significantly longer sentences of *L'Innommable* in contrast to its English counterpart, which is interesting since the English version was published only five years after the French. Texts that have a significantly longer sentence length (>1 word more per sentence) in English than in French are: *The Expelled* and *Texts for Nothing*. Texts with significantly longer sentences in French than in English are: *L'Innomable*, *Bande et Sarabande*, *Cap au Pire*, and *Murphy*, the latter three of which were originally written in English. Yet works in their original language and in translation tend to have the same sentence-lengths, despite what may be assumed about the language difference between French and English(!). Moreover, there is no clear chronological trend here beyond the early and middle French period. Early-period works written in English and middle-period works written in French form neat clusters with comparable sentence-lengths also in this analysis (*Watt*, again, occupying an outlier-position), while later works, from *The Unnamable*, 1949, and forward are more varied.

If we zoom in on the differences between works originally written in French (green) in contrast to works originally written in English (red), we see that *L'Innomable* in French and *Murphy* in English are the only works with significantly longer sentences when we read them in their *original language* instead of in translation (24,8 contra 12,5 words per sentence,

and 19 contra 14,5 words per sentence). However, when seen overall, *the mean sentence-length of Beckett's works originally written in English in contrast to the mean sentence-length of his works originally written in French does not differ significantly* (the difference is on average < 0,5 words per sentence).<sup>28</sup> *L'Innomable* may be considered an outlier with significantly longer sentences in Beckett's L2 French.

Moving on from computational analysis to the close reading of style in Beckett's texts, these findings should be kept in mind. In the end of my examination of Beckett, I will shortly summarise findings from both types of analyses.

## 5 .1. Close reading "The End" and *Fizzles*

As noted, Beckett is especially interesting as a case-study of L2 writing because the genesis of his works is so well documented. Thus, several of his texts are recorded to have been begun in one language and continued in another, perhaps, as noted previously, as a writing-strategy to move forward when feeling stuck. As such we might suppose that these works of mixed composition language will stand out stylistically. The works of which we know for sure that Beckett began in English and continued in French, or where he switched between languages throughout, are (in chronological order): "The End", the collection *Fizzles*, "Imagination Dead Imagine", *Company* and *Stirrings Still*. In fact, it is probable that Beckett changed languages to some degree in the composition of all his works – perhaps excluding the earliest texts written in English, which would explain their distinctive position in Hulle and Kestemont stylistic periodisation (2012) and in the dendrograms in the present study (fig. 7-8). Of these, "The End" and *Fizzles* are examined more closely here, since the exact points where Beckett changed language of composition in these texts is known.

As noted before, "The End" is a key work for examining the change in style between Beckett's L1 and L2 writing at the time when he first started writing in French. Therefore, the focus in close reading "The End" is a comparison of L1 and L2 parts of the text. *Fizzles*, however, is a much later work. Here, most attention is given to examining whether his L1 writing displays cross-linguistic influence of French, because Beckett wrote in French for an extended period of time (1946-1955) before "returning" to English and to writing *Fizzles*.

---

<sup>28</sup> For the detailed table on sentence-lengths of individual works and standard deviation, see:

### “The End”/“La Fin”

“The End” was written in 1946 during Beckett’s productive middle-period of French writing. It was begun in English as “Suite” but continued in French after 29 pages in the manuscript (114). Bolin notes the exact point where Beckett switched to French in the manuscript, from where we could expect to see stylistic variance if the hypothesis of language-specific styles holds true (Bolin 106). That point is approximately when the narrator describes how he made a black cloth to cover his face (approx. the middle of the story). The first half of “The End” written in English, was subsequently translated by Beckett himself into French for the first publication of the story in France. Likewise, he later translated the latter half of the story into English for its English publication. In this close reading, I read the first half in its original English, and the second in its original French, looking at stylistic differences between the first and second half, with a focus on literary devices, language-use, and tropes. Still, I compare certain turns of phrase in the original language to how they were translated by Beckett (either into English or French).

The short story “The End” narrates the misfortunes of a vagrant and is focalized through him. The vagrant recounts his exclusion from a charitable institution, to which he would “gladly have turned back” (Beckett, ‘The End’ 81).<sup>29</sup> However, for reasons unknown, he cannot, and so the story begins with the harsh “end” of his life, meaning both the final period and the actual end of his life. During this period, he lives in turn in a basement, a cave, and a shed, and survives on the money he was given by the institution “to get me started” and by begging on the street (78).

Importantly, there is a shift in subject matter at the middle of the story, coincident with the point at which Beckett switches from English to French. While in the English part, the vagrant is on the go, travelling, meeting, and having conversations with people, in the latter French part, he is alone and in dialogue only with himself. Looking at markers of dialogue illustrates the difference between parts clearly. There are 22 question-marks in the story, of which 13 indicate a question directed at someone to the vagrant or from him to someone, and are followed by, “I said” or “he/she said”. All 13 of these occur in the first, English part of the story. The remaining 9 question marks follow questions which the vagrant

---

<sup>29</sup> In this section on “The End”, I will henceforth refer to “The End” by page numbers only, and to the French version of the story by “Fin” and the page number. All other sources will be clearly marked as such.

directs to himself. Of these, 6 occur in the latter part of the story composed in French. Thus, while the first half composed in English is more focused on people and on conversation, the latter half contains no conversation except for with the self. That the second half of the story, composed in Beckett's L2, is more introspective is in line with the previous findings in differences of vocabulary (Zeta) between Beckett's English and French, but was also seen in the way markers of dialogue (") were more distinctive for early-period English texts in the distinctive features identified by the NSC classifier (table 10 and 11). In general, Beckett's L1 writing is more outwardly descriptive of people, actions, and things, while L2 writing is more introspective.

Moreover, as noted, both Lahiri and Beckett tend to use the first-person singular narrative perspective in their L2, while they prefer the third-person perspective in L1. While there is no clear shift in narrative voice between the first and second part of "The End", there is a split in the character himself. Throughout the story, the narrator alludes to a relatively comfortable former life and to his philosophical education in stark contrast to his life as a vagrant. The split between the vagrant's former and later life is echoed by a similar split in narrative voice. There is a certain distance between the narrator and his recounted self, as he wanders through towns, suburbs, and the countryside. As Bolin notes, Beckett took a new approach to narrative voice from 1946 (starting with "The End"), where "[r]ather than ironically commenting upon the suffering of a protagonist in the past tense, [the] voice speaks from and of its *own* ongoing decrepitude and ignorance" (Bolin 98, italics in original). As such, the narrator is both the experiencing I and narrated I in "The End" – which gives rise to certain discrepancies. For example, the narrator frequently double-guesses and corrects his own statements, as in: "[t]o see nothing at all, no, that's too much" (97). The discrepancy is here between the experiencing I, who exclaims, almost as an interjection: "to see nothing at all" and the commentary of the narrating I, correcting: "no, that's too much". Or similarly, as in: "I liked doing little odd jobs, no, not particularly, I didn't mind" (96). The narrator's double-guessing of the narrated I is here marked by the word "no" and the correction: "I didn't mind".

The discrepancy between the narrating and experiencing I is also underlined by Beckett's choice of words. For example, in the latter half of "The End" originally written in French, the vagrant speaks of whether he remembers having visions in his childhood, and notes: "Mon mythe le veut ainsi" ("Fin" 120). In Beckett's English translation, that is: "My myth will have it so" (98). While in French, this turn of phrase underlines the possible misconception or fictionalisation of his former life, it does so even more in English. In fact,

“my myth” is unidiomatic in English but not necessarily in French. While “myth” is considered a collective fantasy, it may be used in the personal sense in French. For example, you could say: “il a peu à peu construit son mythe” in French. If “my myth” thus stands out as a Gallicism, a calques, or at least a peculiar language in English, that strangeness is reinforced by the direct translation of the French “le veut ainsi” to “will have it so” in English. The effect of the calques and literal translation is a defamiliarization of language, where the reader may stall. Moreover, by using the word “myth” with its allusions of fictionality, (instead of, for example, memory) the disorder in succession between the vagrant and his former self is underscored, as well as between the narrated and the narrator.

### Language use and polysemy in “The End”

The strangeness and Gallicism of “mon mythe” can be ascribed to this part being written in French, whereby the language-use in the English translation tends to become unidiomatic. In fact, it is notable that each half of the text is idiomatic for itself when looking at the halves in their original language, even if they are not so in translation. Each of the halves contain expressions idiomatic in their original language of composition. From beginning of “The End” to the point where Beckett switches from English to French, is peppered with English idioms, i.a. : “it was many a long day”, “nothing to write home about”, “not so old[/sick] as all that”, “a pack of lies”, “get a grip on yourself”, “bald as a coot”, “for hours on end” (78, 79, 80, 82, 86, 87, 89)<sup>30</sup>. In fact, in just the paragraph before Beckett switches to French, there are the idioms: “my hour was not yet come”, “the day came when”, and “the old haunts” (91)<sup>31</sup>. Idioms in the English first half of the story are not translated into French idioms in the French version of the text. For example, “it was *many a long day* before I could button it at the neck” in the English beginning, is translated: “*pendant longtemps* je ne pouvais fermer le col” in the French version (78, “Fin”, my italics). Moreover, there are no English idioms in the latter half of the story composed in French. However, there are French idioms in the second half originally written in French, such as, i.a.: “tom[be] sous la main”, “un mal de chien”, “c’est tout dire”, “toujours le même morceau” (“Fin” 108, 110, 116,

---

<sup>30</sup> I have used the google n-grams tool to determine whether a phrase may be considered idiomatic in a language. N-grams will offer a sense of whether a given word or phrase is frequent and thus common in the very large google books corpus in English or French: <http://tinyurl.com/n-gram-idioms>

<sup>31</sup> <http://tinyurl.com/n-gram-idioms-2>

119)<sup>32</sup>. Like English idioms are not translated into French idioms, French idioms are not transposed into similar ones in the English version of the text. For example, the pain the vagrant describes as “un mal de chien” is translated in English as “atrocious” (93). Thus, there is, on the one hand, an English version of “The End” with the first half full of English idioms, and on the other hand, a French version of “The End” with the second half full of French idioms. The distribution of idiomatic expressions either in English or French clearly bears witness to which language each part was composed in.

While language affects style so that the first and second part of “The End” use English or French idioms respectively, lexical particularities of the French language also make the French version of “The End” different from the English version. As noted, the story was begun as “Suite”, but first published as “La Fin” in French. In fact, when reading it in French, the title is echoed throughout the text. There are various uses of the word “fin” in French, such as in the adverbs *finale* or *enfin*, the verb *finir*, and the noun *la fin*, while there is no corresponding variety of “end” in English. While *finir* as a verb occurs frequently in the French version of the story, it is often rendered by the word *done* in English, such as in: “une fois mon travail fini”, translated as “when my work was done” in English (95, “Fin” 112). Another example is as the narrator reflects on his childhood as he lays dying, and he cuts the reveries short saying: “pour en finir avec les images”, rendered in English as: “to have done with these visions” (“Fin” 121).

Another example of how the “end” (*fin*) echoes across the French text, also underscores the way Beckett uses the polysemy of words stylistically. That is his frequent use of the word *enfin* in French, particularly in the latter part of the story originally composed in French. The term *enfin* is a French adverb that could be translated variously in English. It means both “finally”, “at last”, and “after all”, but can also mean “however” or “nevertheless”, or even “that is to say”. In fact, in the French version of “The End”, it is often used in the latter sense in the mentioned instances when the narrator double-guesses the narrated, as in: “[j]e compris alors que ce serait bientôt fini, *enfin*, assez bientôt” (“I understood then that the end was near, at least fairly near”)(78; “Fin” 78, my italics,). Here, *enfin* means “or rather”, instead of “finally”. Yet, in this case, because *fini* and *enfin* are set side by side the literal meaning of *enfin* is enforced (literally “in the end”), in French, the sentence may be read as: ‘it would soon be ended, in the end’.

---

<sup>32</sup> <http://tinyurl.com/n-grams-idioms-3>

There are more instances where these senses of the word *enfin* begin to blend. For example, when speaking of how he might stand “portant un vieux kepi... saluant à la militaire”, the narrator corrects himself, saying: “non, faux, *enfin*, je ne sais pas, j’avais mon chapeau *à la fin*” (“Fin” 87, my italics). Again, in French, this may be read as, ‘that is to say, I had my hat at the end’, or, ‘*in the end*, I don’t know, I had my hat *at the end*’. In this way, the word “fin”, the title and subject matter, echo across the text in French. In English, there is no similar effect, where the same sentence is rendered as: “wearing a kepi and saluting in military fashion, no, that must be wrong, I don’t know, I had my hat at the end.” (83). Thus, use of *enfin* in French activates both its literal and figurative meanings and draws attention to them in all instances of the word in the story. Moreover, *enfin* here both signals double-guessing, but also alludes to the title and subject matter of the story, “the end” of the vagrant’s life. This “end” is much like the word *enfin*: an “end” to be understood both in the figurative and literal sense: as period of time (the ending) and as a limit (the end as such). That is, the story clearly signals from the very first page that it is “bientôt fini, enfin, assez bientôt”, yet the story is built on the suspense of that end (“Fin” 78). The blending of the senses of temporal succession and limit inherent to the word *enfin* illustrates on the sentence level how the story works at large, with its constant postponement of this “end” of life, which then, *enfin, à la fin*, concludes it.

Thus, in the French version of the story, Beckett effectively makes use of the polysemy of *enfin*, as well as the variety of the French *fin/finir* as a stylistic device to blend literal and figurative senses (as also seen in the use of “go on” in English). It is a stylistic and defamiliarizing device in the sense of Mukařovský, as it activates the awareness of polysemy in the reader and instils a certain strangeness, drawing attention to language itself; we may never be certain if two meanings are or are not intended. The effect is that, as a reader, we find ourselves in a similar position to the narrator: Double guessing at the discrepancy between language and what it is supposed to describe. Moreover, the echo of *fin* across the French version of the text is a clear example of how Beckett plays with polysemy in his L2 French, and therefore an example of language-play in L2 writing, as suggested in HP2 of this study.

### Tropes, doubt and repetition in narrating and writing

The example of *fin* and *enfin* is a type of restatement that Beckett uses as a stylistic device particularly in French. Yet it should be noted that the restatement of a word or phrase is characteristic of Beckett's style in "The End" in general (in both English and French), where it frequently has the effect of activating multiple meanings of a word. Stylistically, this is known as *antanaclasis* (sometimes *polyptoton*), a literary and rhetorical trope, where a word or phrase is repeated, but is used to mean something different the second time than it did the first time (Brogan). Antanaclasis occurs from the outset of the story in the repetition of "go on", in the first paragraph. When reflecting on the shoes the vagrant has been given, he notes: "I would have to get them mended, or get myself another pair, or go on barefoot, if I wanted to go on" (78). First, "go on" is used in the sense "to walk", and then it may be used figuratively as in "to continue living". The effect is that the reader becomes aware of the polysemy of the phrase and may begin to read all instances of "go on" throughout the story as possibly both literal and figurative. Not least in the closing phrase: "without the courage to end or the strength to go on" (99).

Another type of restatement in "The End" often occurs, as noted, in the form of a correction when the narrator enters into dialogue with himself and with language, assessing the words used to describe experience. Instances of double-guessing in "The End" tend to occur more in the second part composed in French than in the first part composed in English. In fact, right after the point where Beckett switches to French as the language of composition, there is such double guessing. The vagrant describes how he got a rag to cover his face, explaining that he took it: "from the lining of my greatcoat, no, I had no greatcoat now, of my coat then." (92). It is not clear here whether the vagrant has misremembered, or if he has simply used the wrong word "greatcoat" rather than "coat". In fact, it could be both. There are similar occasions throughout the French part of the story, such as: "In any case, I slept very little at this period, I wasn't sleepy, or I was too sleepy, I don't know, or I was afraid, I don't know" (96). Overall, it is not clear if the constant double-guessing is due to experience misremembered, or experience misstated. Double guessing, but also repetitions and statement of doubt in the narrative voice increase in the French part of the story. In only two pages of the French half of the text, repetitions and statements of doubt abound, such as *si j'ose dire* ("if I may say so"), *comment dire* ("how shall I say"), *si j'avais bonne mémoire* ("if my memory serves me right"), and *je ne sais pas* ("I don't know") ("I don't know" is repeated 4 times in the course of two pages)(96-97). Moreover, the double-guessing and this linguistic

hesitation in the narrative voice, particularly in the French part of the story, has its parallel in the subject matter. Actually, in the first, English part of “The End”, there are continuous allusions to the vagrant having a strange way of speaking. Reflecting on it, he notes that he has “a way of assimilating the vowels and omitting the consonants” (83). If one would indeed omit the consonants and “assimilate” the vowels, that is, make vowels acoustically similar to nearby vowels, the effect would be an incomprehensible progression of vowel sounds. Yet, these allusions to his strange way of speaking cease in the French part of the story, where the vagrant does not speak to anyone. Instead, language becomes one of the themes of the vagrant’s dialogue with himself. Here, he notes that “[o]n est là toujours entre les deux rumeurs” (“there you are always between two murmurs”)(97, “Fin” 119). This statement may of course be interpreted in different ways. For the vagrant, the two murmurs between which he finds himself may well be the “murmur” of the outside, social world and the incessant internal “murmur”. It is a division that is also represented in his biography: a former (social) and latter (lonely) life, replicated in the narrative structure of the story: a first, social half and a second, lonely half. Yet considering the bilingual composition of this text, it seems to apply equally well to the murmurs of French and English, in-between which Beckett the author wrote the text. In fact, when looking at the narrative voice in “The End”, it is as if the double-guessing and reflection on the textual construction of meaning inherent to the process of writing in L2, was transposed into the text itself. That is, the narrator is continuously assessing the appropriateness of words in the French part of the story, similar to how Beckett may have been assessing the appropriateness of French words when writing. The text exists within these two murmurs.

Overall, “The End” clearly differs stylistically according to language of composition, both in terms of turns of phrase (idioms), subject matter, as well as regarding the play with words, which is more striking in the French version of the story. Furthermore, we have seen that Beckett’s writing contains unidiomatic usages of words and calques, albeit in a surprising way, since it is not his French that contains Anglicisms, but his English that contains Gallicisms. These work effectively to defamiliarize it to English readers, who may in this way “notice the artist at work” but also experience linguistic defamiliarization (Sommer 30). Beckett’s French is not to be considered deficient here (after all he uses French idioms as well), rather, there are clearly differences between the part of “The End” composed in English and that composed in French. It is a difference in style, seen as both language use (idioms), literary tropes, narrative voice and not least in subject matter, conferring upon the

story its two-part structure of moving from social interaction and outward description to soliloquy and introspection.

### *Fizzles/Foirades*

*Fizzles* was first published in 1976 as a collection of eight pieces by *Grove Press*. Because Beckett did not specify the order of the pieces, each subsequent edition has adopted a different order (Rabinovitz 308). Another version of *Fizzles/Foirades* was also published in 1977 as a collection of five pieces in collaboration with illustrator Jasper Johns, and as a bilingual edition with Beckett's own translations. For the purposes of the present study, I examine the Grove edition in English and the first French edition by *Minuit* (1976), since these, in contrast to the version with Johns, contain the only piece originally written in English, "Still". "Still" was written in 1972 and translated into French for the French publication, while the remainder of *Fizzles* were written in 1960 and later translated into English for the English publication (Gontarski 289).

It should be noted that Beckett's text is surely different when read in French or English. Not just because Beckett took liberties in translation both on the lexical level, grammatical level, and in content matter, but also because of "the radical incommensurability of languages" (Butler 115). That is, while there is of course a difference between idiomatic expressions in English and French (for example, the French "soupe" should in certain contexts be translated to "Tea" in English), there are also more structural differences between the languages, for example, that French uses more animism, tending to personify inanimate things by human attributes (Butler 118). Another example is how French may be more concise, not needing as many prepositions as English, so that "j'arrive" must be dynamized with a preposition and adverb when translating into, for example, "I'll be right over" (Ibid.). Butler gives various examples of calques and transposition in Beckett's translations, but also how the "animism" of French is reflected in his English translations of French, so that we might say that the structure and characteristics of French are apparent in Beckett's English translations (119-120).

The aim in the present analysis, however, is different from comparing originals and translations. It is rather to ask whether there is a stylistic difference between Beckett's *original* English style and *original* French style, such as is the general inquiry of the present study. Moreover, whether an influence of having written in French may be seen in Beckett's

“return” to English. “Still” is therefore examined in particular because it is the only piece in the collection originally written in English. The aim is to see whether “Still” displays signs of heightened linguistic consciousness (according to HP3 in the present study). But also, whether “Still” displays cross-linguistic influence of French, according to HP5, that bilinguals may seek to express structures, phrases or words from one of their languages by altering the standard language-use of the other. Butler’s description of some characteristics of French may be helpful here. If Beckett transposed French language-use to his English writing, we may see such characteristics in “Still”. After having written in French from 1946 to 1955, Beckett began writing in both English and French. Thus, if the period of French writing stylistically changed Beckett’s English writing, “Still” as a work written in 1972 should reflect that.

*Fizzles* consists of eight short prose pieces or “fizzles”, with a length of two to five pages each. Although the pieces seem unconnected, echoes are nevertheless heard between them, both of words, phrases, and themes. Moreover, characters, places, motifs, and themes from Beckett’s earlier works reappear in *Fizzles*, making it a bricolage of Beckett’s *oeuvre*. The first fizzle, for example, contains allusions to the protagonists of other works, Murphy and Molloy (Rabinovitz 308). Moreover, the prominent theme of the journey in *Fizzles* is reminiscent of Beckett’s wandering protagonists, as in *Molloy*, *The Unnamable* or in “The End”. In *Fizzles*, the prominent theme of travelling or journeys, to Rabinovitz, suggests “the progress of life, of thought, or of making one’s way - as author or reader - through the work” (309). Yet it should be noted that journeying in *Fizzles* often stands in stark contrast to its opposite, immobility – a contrast which is central to the collection. Several of the fizzles narrate characters moving or journeying, even if they do so reluctantly, through a form of labyrinth (fizzle no. 1), landscape (no. 3), unspecified space (no. 4), and desert (no. 8).<sup>33</sup> Other fizzles (no. 2, 6, and 7) describe characters in closed spaces (rooms) that seek to “simply stay still” even if they do move off and on, for example to a window and back (Beckett, ‘Fizzles’ 239).<sup>34</sup> Fizzle number 5 does both. Also, in contrast to the other fizzles, it does not describe a character, but a “closed place” which, however, seems to expand and shrink at will while being described, and which is filled with “millions” both “wandering and still” (236). Fizzle 5 thus sets up and highlights the opposing themes of open or closed spaces

---

<sup>33</sup> I refer to the order of *Fizzles* in the Grove Edition here.

<sup>34</sup> In this section on *Fizzles*, I will henceforth refer to the English *Fizzles* with page numbers only, I refer to the French version by *Foirades* and page number. All other sources will be clearly stated as such.

in the whole collection, and of reluctant wandering and impossible stillness, whether that is taken to be of life, mind, or the author and reader in the text.

The fizzles vary in the way they are told, but when read in continuation of each other, they do suggest a pattern in narrative voice. Rabinovitz suggests that the narrative voice and the characters in *Fizzles* are but “different aspects of a single self” (Rabinovitz 311). In both fizzles 3 and 4, the narrative “I” describes how at birth, “it was he, I’m inside... it was he who wailed, he who saw the light, I didn’t wail, I didn’t see the light” (232, 234); implying that the “I” is inside of “he”, but at deeper level than that of experience. Moreover, the “I” describes how “he”: “seeks a voice for me... I have none” (232). In this way, the dynamic between the narrative “I” and narrated “he” may be interpreted as a dynamic between layers of self. In fizzles 3 and 4, it is an “I” who speaks about a “he”, while in fizzle 6, the “I” speaks about a “you”. Because of this form of narration in some of the fizzles, whenever the narrative voice speaks only about a “he”, even if not using first person pronouns (in fizzles 1 and 8), it is nevertheless implied that the narrator is again the “I” that is narrating; a version of the narrated character himself. However, fizzle 7, “Still”, stands out in this pattern because it omits pronouns altogether, as shall be described in detail shortly. At present, it is important to note that the use of narrative voice in the seven fizzles originally written in French, excluding the English fizzle “Still”, again confirms the pattern seen in the examination of vocabulary of both Beckett and Lahiri’s *oeuvre*: that Beckett’s L2 writing (here, the seven fizzles originally written in French) tend to employ a first-person narrative perspective.

### “Still”

The piece “Still” in *Fizzles*, as the only one originally written in English, stands out variously, both from this pattern of narrative perspective and in its style, when including both narrative devices, tropes, and language use in an expanded notion of style. Importantly, fizzle 7 is one of the few fizzles that bears a title: “Still”, implying the thematic centrality of the piece in the collection. In the Minuit edition, this piece is also placed last as a form of closing statement (Beckett, *Foirades* 42–47). In “Still”, an unnamed character is described to move on and off in a room with windows, watching the sunset and nightfall. In the beginning, the character sits in a wicker chair by a south window, then turns to look southwest through a window there, gets up to stand at a western window, then sits back down at the south window, then looks through the west and the east window and finally puts the head to rest in the right hand. So much for the plot. The piece is narrated *as if* from the first-person

perspective and with internal focalization, interspersing outward description with internal thought, such as in: “staring out at nothing just failing light quite still till quite dark *though of course no such thing just less light*” (241, my italics). As in this example, where internal descriptions end and internal begin is not signalled even by a comma in “Still”. This internal/external narration is reminiscent of both the narrative mode of “The End”, and of the narrative device commonly known as stream of consciousness, where the inside and outside melt into one.

Still, I say narrated *as if* in the first person, because all pronouns are entirely absent from the piece. Rabinovitz notes that “Still” tells of a man, but that can in no way be inferred from the text itself (315). In “Still”, Beckett seems to write without subject or narrative perspective. An example of how Beckett writes without subject can be seen as the narrator describes looking through a window of the room: “[n]ormally turn head now ninety degrees to watch sun if already gone then fading afterglow” (240). The conjugation of “turn” in this sentence would imply a first-person perspective – or that Beckett is using the infinitive. Elsewhere, Beckett strategically uses the present participle instead of a conjugated verb (sitting, trembling, staring) or obscures the subject by letting body parts be the subjects of verbs and actions, such as in “[e]yes stare out unseeing” or “[t]he right hand... leaves the armrest taking with it the whole forearm complete with elbow” (240-241), where it is brought to a point of absurdity. These strategies are strained to omit the subject at all costs. As noted previously, the third-person narrative perspective is predominant in Beckett’s L1 English writing, but he does not employ it in “Still”. However, neither does he employ the first-person perspective, but seems to choose a middle way. Perhaps being versed in both ways of writing, Beckett sees alternative possibilities of narrative perspective.

“Still” omits not only pronouns, but prepositions, conjunctions, and, as noted, punctuation marks. To get a sense of such condensed prose, consider the following example: “Even get up certain moods and go stand by the western window till quite dark and even some evenings some reason long after” (*Fizzles* 240). Here it is clear that pronouns are omitted (who or what “get up”?) as well as prepositions (“even [on] some evenings [for] some reason”). Prepositions are not entirely absent, but very sparse in “Still”. The prose of “Still” may be considered an example of how Beckett was influenced cross-linguistically by the conciseness of French and the lesser need in French to supply prepositions into his English writing. As noted before, a form of such condensed or “minimalist” prose is characteristic especially of Beckett’s later works (after 1966)(Gontarski xv). Gontarski suggests that it was the outcome of adopting the minimalist aesthetics of architecture from

Mies van den Rohe and Adolf Loos (Ibid.). I would suggest that we (also) consider the role Beckett's close to ten-year period of writing in French played in his stylistic development toward a condensing his prose.

Another cross-linguistic influence can be seen in the words Beckett does use. Studies of Beckett's translations into French are often focused on calques or Anglicisms (Germoni; Sardin-Damestoy). That is, finding the places where Beckett is exposed as a native speaker of English and not of French. For all that, studies rarely address calques in his English, that is, we could say, where Beckett is "exposed" as an L2 speaker of French. There are in fact various Gallicisms in "Still" (written in English from the outset). One example is where Beckett writes that "the sun shines out and goes down" (240). The sun does not commonly "shine out" in English, but the sun may "brille dehors" in French. Another example of Gallicism in his English is how Beckett writes "leave it so", in: "Leave it so all quite still" (241). While the phrase is awkward in English, to say "laisse-le là" is common in French. That is also what Beckett translates the phrase as in the French translation of "Still": "Le laisser là tout immobile" (Foirades 47). That is to say that Beckett's English is here clearly influenced both by the grammatical conciseness and vocabulary of French. It should, however, not be reduced to a flaw or (unintended) calques. Rather, these influences of French in "Still" contribute to an effect of linguistic defamiliarization – for English readers in the case of Gallicisms, and for any reader in the case of Beckett's creative omissions of narrative perspective and prepositions.

It should be noted that the example above of the sun "shining out" may be seen as an instance of language-use that supports HP5. That is, that authors may use language creatively to express a certain nuance of a word or phrase in one language, which does not have an equivalent in the other language. In French, it is common to say both "la soleil brille" and "la soleil brille dehors", but in the latter case, the strength of the sun is emphasised. The phrase "la soleil brille *dehors*" in French brings perceptual connotations with it, the heat, the sharpness of the light, etc. Thus, we might say that Beckett coins the phrase "the sun shines out" in order to express a certain perceptual nuance that the standard English phrase "the sun shines" cannot express. It is thus both an example of (perhaps wilful) calques, and of how Beckett expresses particular perceptual experiences connected to French words and phrases by changing the English language.

Moreover, "Still" plays creatively with language in other ways. Both by transforming English idiomatic phrases, by employing tropes like antanaclasis, and by drawing attention to the graphic similarity of words. I will treat these three characteristics in turn. Consider, for

example, the first sentence of “Still”: “Bright at last close of a dark day the sun shines out at last and goes down” (*Fizzles* 240). A couple of things characteristic for the style of “Still” are happening in this sentence. Firstly, Beckett does employ an English idiomatic phrase here: “at close of day”, as a native speaker is said to do (Hinkel 247).<sup>35</sup> Yet he reshapes it creatively, interposing the adjectives “last” “a dark”, so that what would have been an idiomatic use, becomes a strange language-use. Without the adjectives, the phrase “bright at close of day the sun shines” would not be odd in English, and “bright” would be clearly understood to describe the sun. However, the addition of “last” in Beckett’s sentence, makes the phrase “bright at last” read as an adverbial phrase, that it is: “bright at last(!)”, rather than “bright” being a modifier describing the sun. Breaking up the idiomatic phrase “at close of day” does not make it ungrammatical, although it does slow down reading and draw attention to the way meaning is constructed in the sentence.

Secondly, the repetition of “at last” in this sentence is used in a particular way. It is again the trope known as antanaclasis, where two words that look lexically identical are repeated close to each other but are used to mean different things. In the first instance, “last close... of day” may mean *at the last of days*, where “at” is a preposition and “last” a modifier of “day”. In the second instance of “at last”, it is an adverbial phrase in the sense of *finally the sun shines*. The trope antanaclasis is particularly interesting in L2 writing, because it implies an awareness and wilful use of the polysemy of words. It is an instance of the sort of language-play suggested characteristic of L2 writing (HP2). As noted in the close reading of “The End”, Beckett used antanaclasis and polysemy particularly in his French L2 writing – here, we see it again in an originally English text.

There are various examples of antanaclasis in “Still”. Like the word *fin* in “The End”, Beckett frequently employs one particular word in different senses throughout this text, and like in “The End”, that word is in the title: “still”. The word “still” may be taken both as an adjective meaning *immobile* (being “not still at all but trembling”) or *quiet* (being “still... listening to the sounds”) or as an adverb, meaning *even now* (sitting “still head in hand”) or *until now* (there being “less light still”) (all *Fizzles* 241). Beckett uses it frequently and in all these senses throughout the short piece, or on just one page. The word “still” is also used antanaclastically, as in: “Normally watch the night fall... from this narrow chair or standing by the western window *quite still* either case. *Quite still* namely staring at some one thing

---

<sup>35</sup> <http://tinyurl.com/n-gram-idioms-4>

*alone* such as a tree or bush a detail *alone* if near if *far* the whole if *far* enough till it goes” (241, my italics). In this example, still is first used in the sense of *immobile*, specifying that he stays in a certain position (sitting or standing) and shortly after “still” is used again to mean either *immobile* or *quiet*, specifying that he is focused on looking (not making a sound). In this sentence, Beckett also creates antanaclasis with “far”. In the first instance, “far” takes the spatial meaning (whether the bush or tree is *far away in space*), and in the second instance of “far”, it takes a temporal meaning (that he will see it if there is *far enough of time* before the landscape darkens). Similarly, the repetition of “alone” in this sentence may be interpreted as an antanaclasis, if we take the first “alone” to mean that he is there staring *on his own*, and the second to mean that he is *only* focusing on one detail. However, it may also be that word “alone” is used synonymously here, with no antanaclasis, but the fact that I am in doubt is a good example of how frequent antanaclasis opens the readers eyes to polysemy and encourages a reading that weighs every word carefully.

A third characteristic of the language-use of “Still” is that Beckett engages in frequent wordplay, where he both plays with polysemy and the graphic similarity of words. One example is his frequent use of the word “till”, especially as an echo of and combined with the word “still”, such as in: “failing light quite still till quite dark” (*Fizzles* 241). Apposition of “still” and “till” occurs six times in this short piece of 835 words. Moreover, Beckett also uses “quite” and “quiet” apposed in the same way, as in: “quite quiet” (*Fizzles* 241). These words are very frequent in the short text. “Still” or “till” occur 34 times and “quite” or “quiet” occur 21 times. Because “still” also frequently occurs with “quite”, as in: “leave it so all quite still or try listening to the sounds” (*Fizzles* 242), and because “quite” is so graphically similar to “quiet” the effect is that the more the combination is repeated, the more “quite still” begins to read like “quiet still” or “silence silent”. Thus, the graphic similarity of words is effectively used for halting the reader and encouraging a reading for polysemy.

### An exercise in not having a voice

This language play is not to be considered simply a stylization of language but is inherently linked to the themes of stillness and wandering of *Fizzles* as a whole, themes that Beckett brings to echo across the fizzes. Like in “The End”, language is thematised in *Fizzles*, not only in the explicit reference to “finding a voice” as mentioned in the beginning of this reading, but also in the way Beckett uses the narrative perspective and language of the pieces. In stark contrast to the omission of narrative perspective in “Still”, fizzle 3 and 4

employ pronouns very frequently in their original French. In French, pronouns do not only refer to people and things, but may be used in impersonal verbal constructions, such as in “il faut” (it is necessary). Beckett employs pronouns to excess in these two fizzes by interspersing personal and impersonal uses, for example as in: “il ne bouge pas, il me cherche une voix, il est impossible que j’aie une voix” (“he is still, he seeks a voice for me, it’s impossible I should have a voice”) (232; *Foirades* 36). This constant repetition of *il* occurs throughout fizzle 3 and 4, in contrast to constant omissions of pronouns in “Still”. While in fizzle 3, a voice is sought for the narrative I, in “Still”, we are asked, almost in an imperative form (or, again, infinitive) to “[I]eave it so all still or try listening to the sounds” (241). In fact, considering the omission of narrative perspective, and the scaled down language of “Still”, the piece may be considered a literary exercise in *not having a voice* – which does not mean that it does not have style.

In sum, Beckett plays with the impersonal construction in French, employing it frequently and interspersing it with the personal form to enforce its contrast to the omission of subject in “Still”, as well as the contrast between presence and absence, voice and stillness in all fizzes. His use of both lexical and grammatical polysemy (*il*) in *Fizzles* supports HP2, that L2 writers play with language in L2. But so too does he play with English. As shown here, Beckett’s play with antanaclasis, apposition, calques, and English idioms effectively defamiliarizes language and draws attention to the polysemy of language. Thus, the case of *Fizzles* also supports HP3, that bilingual authors display an active and involved awareness of language in their L2 *and* in their L1. Moreover, when taken together, the way Beckett omits narrative perspective, his condensed prose, and the Gallicisms in his English may be ascribed to an influence his L2 French writing upon his L1 English writing. It supports HP5, which suggests that L2 writing should have a cross-linguistic on L1 when authors “return” to write in L1.

Ultimately, Beckett does not merely stylize language, but uses it to strengthen and engage with the subject matter and form of “The End”, “Still” and *Fizzles* as a whole. As shown in the close reading of “The End”, Beckett’s English and French writing is clearly dissimilar stylistically, yet the change in language of composition was also seen to contribute to the two-part structure of “The End”. Similarly, “Still”, the only piece written in English, intricately engages with the themes of presence and absence, levels of self, and finding voice and stillness in the way it plays with language on the sentence-level.

## Summing up Beckett

In summary, there are clearly differences between Beckett's L1 and L2 writing, both when looking at an expanded and minimal notion of style. The cluster analysis of Beckett's *oeuvre*, looking at minimal style suggested both temporal divisions (largely aligned with those of Hulle and Kestemont) and divisions along the language line, particularly between his early-period English style and middle-period French style (fig. 7-8). The more varied picture in later works suggests that, with time, Beckett's (minimal) style became less stylistically divided in respect to which language he composed in. Yet, English-language compositions consistently stood out from clusters in both the French and English-language analyses. These results very much echo those of Lahiri, where we saw a strong division along the language line initially, while the latest text, after Lahiri "returned" to English, was more varied in style.

One clear indication of language-specific style in the case of Beckett is found in looking at vocabulary both through the classifier (vocabulary of minimal style) and Zeta (full vocabulary). It seems that Beckett generally writes in a first-person narrative perspective in French, while preferring a third-person perspective in English (fig. 9-10). The same was seen in the close reading of "The End". Where the English (first) part of the text had a social focus (people, dialogue, travelling), the second, French part, was more introspective. Importantly, both the division along the language lines in terms of narrative perspective, but also the division of early English style and middle-period French style may have to do with Bolin's claim, that Beckett significantly altered his narrative voice beginning with "The End" – which is coincidentally when he started to write in French. Hanauer's considerations imply that this shift may have to do with the fact that he started to write in French, so that the shift in style may have been an effect of L2 writing and its tendency toward a personal tone (HP3).

Moreover, we saw that early English-language compositions and middle-period French compositions formed neat groups of comparable sentence length (fig. 12-13). Yet the examination of sentence length, like the cluster analysis, showed more variation in later style (from 1949 and forward). The examination of sentence length showed results in opposition to part of the hypothesis (HP1), that L2 writing (here, Beckett's French texts) should generally employ shorter sentences. Yet neither is the opposite the case, that Beckett's French works should employ longer sentences. In general, it should be said that Beckett has a very varied style regarding sentence length, sometimes as low as 3 words per sentence in average in a text, sometimes as high as 22 (fig. 12-13).

There were no clear indications of cross-linguistic influence in terms of sentence length. That is, that Beckett should have imported features (such as shorter sentences in French) into his English style. Yet in the close reading, it was shown how Beckett imported French turns of phrase into his English with a defamiliarizing effect. Importantly, in both “The End” and “Still”, it is not Beckett’s L2 French that contains calques, but his L1 English. That supports the idea (HP2) that L2 authors should play creatively with language and, for example, use calques intentionally. Thus, Beckett draws on the interpretive frames and vocabulary of one language when writing in another. Beckett was also seen to change idiomatic expressions, transpose French words and phrases into English and stretch the norms of standard-language in the close-reading of “The End” and “Still”.

## 6 Two authors' left-hand style

### Synthesis

#### Deficiency in L2 writing

Importantly, in both the case of Lahiri and Beckett examining their style by sentence-length and lexical diversity displayed results to the contrary of part of HP1 that L2 writing should be less lexically diverse and employ shorter sentences. Actually, when looking at lexical diversity of Beckett and Lahiri's texts, it seems that L2 writing is more lexically diverse, suggesting that Beckett and Lahiri use more synonyms and a bigger vocabulary in their L2 writing (although Beckett's L2 texts are only slightly more lexically diverse). It is important to note that it is so *despite the difference in L2 proficiency* between Beckett, who learned French at an early age, and Lahiri, who learned Italian at an adult age. Here, it should be kept in mind that high lexical diversity score does not automatically imply a language-master or indicate proficiency. As Laufer and Nation note, proficiency should not be determined by how big our vocabulary is, but by how well we use it (310). Nevertheless, the aim of this study is not to assess proficiency, but to compare L1 and L2 writing. As such, the significantly higher MSTTR of Lahiri's L2 compared to her L1 writing, and somewhat higher MSTTR of Beckett's L2 writing compared to his L1 writing surely underpins the idea of language-specific style, that is, a more varied vocabulary in the two authors' L2 writing. Furthermore, it rebuts the hypothesis that L2 writing should be less lexically diverse (part of HP1). Also in contrast to HP1, sentence-length was not shown to be either significantly higher or lower in Beckett and Lahiri's L2 writing than in their L1 writing. Sentence-length does however display the diversity of especially Beckett's style, who wrote texts with very long sentences and very short. In general, the case of Beckett and Lahiri's L2 writing disproves the idea that L2 writing should be deficient L1 writing (HP1). Still, their L2 writing is very *different* from their L1 writing.

#### Difference and language specificity in L2 writing

Going methodologically beyond quantitative measures such as lexical diversity and sentence-length helps us understand in what way L1 and L2 writing differ in Beckett and

Lahiri's case. In terms of minimal style (use of function words and features), cluster analysis of Beckett and Lahiri's *oeuvres* show a different minimal style in L1 compared to L2 writing.

When looking at vocabulary (both of minimal and extended style) it was found that particularly Beckett and Lahiri's use of first- or third-person pronoun or narrative perspective differs between L1 and L2 texts. A predominance of first-person perspective in L2 texts was true both for Lahiri, but also for Beckett's *oeuvre*, which is significant because Beckett's *oeuvre* spans a longer period of time and might as such be subject to more stylistic change. As Hanauer has suggested (HP4), L2 writing seems to be more personal than L1 writing. In the case of Beckett and Lahiri, both authors tend to use first-person narrative perspective in the L2 writing, and third-person narrative perspective in their L1 writing. This preference was consistent across time also in Beckett's *oeuvre*. In general, Hanauer's consideration (HP3), that L2 writing should employ a more personal and introspective style, is affirmed in the case of these two writers.

The agreement of the findings with parts of HP4 should be highlighted. The idea that L2 writing should be more personal stands in contrast to the idea which Yildiz considered characteristic of the monolingual paradigm: That L1 should be the only language in which one may properly "think, feel, and express oneself" (7). In fact, the opposite seems to be the case when looking at the bilingual *oeuvres* of Lahiri and Beckett. Both seem to write more personally (first-person perspective) and more introspectively in their L2. When looking at vocabulary, the introspective tendency of Beckett and Lahiri's L2 texts is also seen in the use of verbs, adjectives, and nouns, which are more oriented toward the (narrative) self, thoughts and feelings. Meanwhile Beckett and Lahiri's L1 texts are full of nouns, action-verbs, and adjectives describing the social world: people, places, actions, and things. In sum, it seems that it is the L2 that is "the language of emotions" and introspection (Pavlenko *Emotions* 30), while writing in the "mother tongue" displays an outwardly descriptive and less personal style in Beckett and Lahiri's case. But does that mean that Beckett and Lahiri each have two language-specific styles? And, as suggested by HP7 of this thesis, that L2 writing essentially uses a more personal or introspective style than L1 writing? The answer of the present study leans toward a no.

When looking at style in Beckett and Lahiri's *oeuvres*, the binary division between a "mother tongue" and an "acquired language" seems to be reflected, even if the characteristics of L2 style is the opposite of what has been claimed in the monolingual paradigm. I say "seems", because especially the case of Lahiri does not conform to a neat division into "mother tongue" and acquired language writing. While Italian can indeed be considered an

“acquired” language for Lahiri, her English and Bengali may both be her first languages, which muddles and complicates a L1 and L2 division whereupon HP7 rests. Why this predominance of the first-person perspective and introspection in L2 texts? As HP7 suggests, a L2 detachment effect might mean that authors feel freer to express their thoughts, emotions, and, as Dorfman put it “to show myself” (212). Yet it should be remembered that the findings of the present study point to a predominance of a certain *narrative perspective* and (internally focalized) narrator. Moreover, the L2 detachment hypothesis does not suggest that L2 writing should deal less with characters, actions, and outward description. With that in mind, the L2 detachment effect seems insufficient to me for explaining the characteristics of L2 writing in the case of Beckett and Lahiri.

Moreover, that is also why it is important to emphasise the ways in which the findings of the present study *do not* support a clear-cut and essentialist distinction between L1 and L2 style. As has been shown in the case of Beckett and Lahiri, if seen chronologically, there is a strong early division between L1 and L2 style both when studied by cluster analysis and classification. In Beckett’s case, this division corresponded neatly with the findings of Hulle and Kestemont, so that we may speak of an early-period English style and a middle-period French style. In Lahiri’s case, the division of her style along the language line was also palpable prior to 2022. Both Beckett and Lahiri’s minimal style changes after they “return” to their L1. For Lahiri, the 2022 work *Translating Myself and Others* which was in part written in English stands out stylistically, both in terms of sentence-length, and minimal style (in the cluster analysis and classification alike). For Beckett, it was seen that the style of *From an Abandoned Work* is not, as suggested by Hulle and Kestemont a “return” to his English style, but different from Beckett’s early-period English texts. Nevertheless, in Beckett’s case, works written in English consistently stand out from clusters of texts written in French. That is, Beckett’s texts written in English after the early and middle-period *are* different from texts written in French, but they are not different in a way similarly enough to say that Beckett has a consistent “English style” across his *oeuvre*. Likewise for Lahiri, while she has an “English-” and “Italian style” before 2022, her “return” to English with essays in *Translating Myself and Others* cannot be easily classified as part of her old “English style”.

It is also notable that there has been little support for HP6, that authors prefer certain words, concepts, or topics in one language in comparison to the other. Lahiri was seen to write more about writing, but there were no language-specific topics seen in the case of Beckett. Yet that is perhaps because methods in the present study were not optimized to pay attention to topics, such as in topic modelling.

While in both Beckett and Lahiri's case it was shown that L1 and L2 *minimal style* was different when they started to write in L2 and it then became less specific, anomalies to this pattern should also be addressed. In Beckett's case, there is the exception of *Watt* (1931), which is written in English, but nevertheless tends to group differently (with early-period English texts or with middle-period French texts) depending on whether we look at the French or English text. As Hulle and Kestemont note, the oscillation of *Watt* in a stylometric dendrogram may have to do with the fact that the French translation was made 28 years after it was written (1969), so that the translation is somewhere else stylistically (Hulle and Kestemont 195). I think we should also consider the language factor. That is, the fact that *Watt* oscillates between early-period English texts and middle-period French texts depending on whether we examine it in English or French, is because *Watt* occupies a middle position between them stylistically. The English of *Watt* is perhaps, as Slote argues, an "English inflected by French" (Slote 118). As Kager has noted, *Watt* was written while Beckett was living in France and speaking French daily. Beckett was also seen to edit and comment on the manuscript of *Watt* in French (75). While the language of composition of *Watt* was English, it was nevertheless revised and rethought in French. Thus, the case of *Watt* also points to the artificial analytical division between L1 and L2, that are in reality much more intermingled in the bilingual, who is, as Derrida argued, entrenched in language(s). I think it is important to highlight even if working with graphs and numbers that methods are only as good as our analytical categories.

Still, Beckett and Lahiri's L2 texts do share characteristics which distant and close reading methods can capture. Clearly something is happening when these authors change their language of composition. The tendency of an outwardly descriptive style in L1 writing and personal and introspective style in L2 was particularly reflected in close reading "The End", where these tendencies were also reflected in the narrative structure. Because Beckett wrote the first part in English and the second part in French, "The End" has a two-part structure both in terms of content matter and style in the expanded sense. "The End" as the exact point in Beckett's creative evolution where he chooses to switch to his L2 French for literary composition thus reflects the significance of that choice.

Yet, in the case of *Fizzles*, the tendency for first-person narrative perspective and introspection of L2 writing, and third-person narrative perspective and outward description of L1 writing was not as present. While the originally French texts in *Fizzles* do employ first-person perspective to some extent, the originally English-language text "Still" plays with omitting narrative perspective altogether. "Still" is also not as direct and outwardly

descriptive as was noted for other L1 texts but is actually more similar to the style of Beckett's L2 texts in its introspective tone (hence its minimal plot). Moreover, in *Fizzles*, the text originally in English and those originally in French speak to each other in terms of themes and language-use. Thus, while "Still" stands out in terms of narrative perspective, *Fizzles* is not as palpably divided according to language of composition as "The End".

### Linguistic consciousness and L2 writing

As noted, neither "Still" nor *From an Abandoned Work* seem to really return to a L1 style. The stylistic variance across Beckett and Lahiri's *oeuvres*, and the fact that these two authors do not "return" to a fixed L1 style after having written in L2, suggests looking at L2 writing as a creative resource rather than a fixed (pro- or deficient) style. In fact, looking at L2 writing as a creative resource helps us understand the stylistic evolution and choices of authors, also because the role of language for literary style tends to be understated. Bolin did note that starting with "The End", Beckett took a new approach to narrative voice, yet he does not consider language as a factor in that change (98). Likewise, Gontarski has connected Beckett's minimalist style from 1966 and onwards to minimalist aesthetics in architecture, but not to the fact that Beckett wrote in French for a long period of time, which has a structural tendency for conciseness (xv; Butler 118). Still, I would refrain from declaring L2 writing the cause of a certain fixed (narrative) style. The variance of Beckett and Lahiri's *oeuvres* in minimal style as seen in the cluster analyses also disputes such an idea. Any stylistic choices that writing in an L2 effected and influenced in the case of Beckett and Lahiri should not be seen as permanent changes, but only as a part of the stylistic versatility which L2 writing helped facilitate. And it is perhaps a

In general, looking at Beckett and Lahiri's L2 writing has supported HP3, that an engagement and awareness of language is palpable in L2 writing and in L1 writing when author's "return" to it. The first part of HP3 is supported by the findings from the examination of lexical diversity in L1 and L2 writing in both Beckett and Lahiri's case. Higher lexical diversity indicates that Beckett and Lahiri are more active, more involved with their choice of words and with the lexical palette of the L2 available. Moreover, by close reading Beckett's "The End" and *Fizzles*, I have shown various ways in which Beckett's defamiliarizes and plays with language in French, but also in English, and where French has been a creative resource for his English writing. That is, how Beckett plays with polysemy

(such as through antanaclasses), with the graphic similarity of words, with standard-language norms and idiomatic expression, and with transposing French words and phrases into his English. Thus, the close-readings support the HP3, that L2 writing should heighten linguistic awareness and dispose toward language-play, and HP2, that L2 writing defamiliarizes language, but also HP5. That is, that language is altered and used creatively to express a certain concept from one language that does not have an equivalent in the one the writer seeks to express it in. It was seen how Beckett introduced calques and non-standard phrases and word-order to do just that (particularly the example of “my myth” in “The End”). Yet, Beckett was also shown to change standard English, not only to express semantic meaning, but also perceptual, sensori-motor features of French words (as in the example of the “sun shines out”), which underscores the embodied nature of bilingual language creativity.

Moreover, it was also shown how Beckett’s language-play is prompted by the particular qualities of the language itself (such as the similarity of the word “quite” and “quiet”, the polysemy of the word still in English, the possible personal meaning of the French *mythe*, the impersonal verbal constructions in French etc.). That is, creativity with a language is grounded in the particular linguistic features of that or another language. In the close reading it was not shown that Beckett is a language-master shaping language to his will, but rather that he is attentive to the opportunities that *multiple* languages offer him.

## 7 L2 style: Conclusions Analysis

The present study has examined the differences between Beckett and Lahiri's L1 and L2 writing by close- and distant methods. It was found that these two authors' L2 writing is generally more personal and introspective, and more lexically diverse than their L1 writing. Moreover, it was shown that L2 writing is not deficient in comparison to L2 writing when looking at lexical diversity and sentence-length. It should be highlighted that Beckett and Lahiri's L2 writing share the outlined characteristics despite the fact that Lahiri learned her L2 at a much later age than Beckett, which suggests that these characteristics of L2 writing are not overly dependent on level of proficiency.

Stylometric analyses by cluster analysis and classification showed that Beckett and Lahiri's L1 and L2 writing differ in minimal style, that function words are used differently, but that in returning to L1, Beckett and Lahiri write more stylistically varied than they did before. Close-reading of two mixed-language texts by Beckett showed that his L1 English and L2 French writing differs in an expanded notion of style especially regarding narrative perspective. Close reading also showed that Beckett employs ostentatious language-play in both languages, both to defamiliarize language, draw attention to the polysemy of words and phrases, to enforce and explore subject matter. In "The End", that was seen particularly in Beckett's L2 writing, but in "Still" also in his L1 writing. As such, it is suggested that a higher linguistic attentiveness and awareness both in L1 and L2 is a bilingual advantage. I also showed how Beckett played with language in and to express both semantic meaning and sensori-motor aspects of L2 words and phrases not inherent to existing words in L1, which suggests that the L1 and L2 as well as L1 and L2 style are not distinct in the bilingual mind, but that there is transfer between them.

The findings from close reading are important because they show how quantitative and qualitative analysis may complement each other in examining L2 style, by examining different notions of style. Combining distant reading, which looks at quantitative features and minimal style, and a stylistic close reading which looks at an expanded notion of style, can show how function-word use, language-use, form and content work together in generating L2 style. Moreover, combining methods can give us a much more elaborate perception of the innovative role of L2 writing in L2 author's style(s).

Distant reading is not essential to examining L2 writers' styles, but it is useful, just like close reading is not essential but useful. For example, we may have arrived at the finding that Beckett and Lahiri's L2 writing uses more first-person perspective by a different route than computational stylometry, such as strategically focusing a close reading on narrative perspective in L2 texts. But it is important to note that stylometry identified a pattern in this case, which I would otherwise not have looked for. Conversely, distant reading is blind to stylistic devices such as tropes, and the nuances of linguistic features such as word polysemy. The trope *antanaclasis* is a good example, because it is the repetition of a word in different senses, which computational stylometry based on word frequencies would see as one word. Without combining stylometry with close reading, I would not have seen nor looked for how Beckett plays with language. Therefore, combining these methods means combining different perspectives on style, what to see and how to see it, and thereby gaining a deeper understanding of L2 writing.

Beckett and Lahiri's L1 and L2 writing styles are different, and their L2 writing does have characteristics in common. Still, conclusions are limited to describing these two writers, who are, in fact, very particular kinds of L2 writers. Both Beckett and Lahiri are relatively proficient in their L2 and have followed a very similar trajectory: having their literary debut in their L1, living in the country of their L2, changing their language of literary composition to L2 and then returning to their L1. Findings in the case of Beckett and Lahiri are also limited to one language family (Latinate languages), a historical time, a specific (Western, 20<sup>th</sup> and 21<sup>st</sup> century) culture, literary tradition, and not least, to prose as a genre. L2 writers from a different time, or L2 writers who, for example, debut in their L2, who write in different scripts and genres – or across scripts and genres – may display different characteristics of L1 and L2 style or may not display any significant differences between L1 and L2 style at all.

That is also to say that there are a multitude of possibilities for a more varied and comprehensive study of L2 style, by looking at a bigger corpus of L2 writers, at different kinds of L2 writers, at a different genre or genres, or at other languages and scripts. This study has shown that a combination of close- and distant reading methods can be productive for further examining L2 writing style. Yet, there are a wealth of methods for assessing, for example, lexical diversity, not least a rapidly growing palette of methods in computational stylometry and stylistics, which may be more apt for assessing, for example, topics or parts of speech preferred in L2 as opposed to L1 writing.

I began this study by asking: what does this “writing with the left hand” or “poverty” of L2 style *look like*? In the case of Beckett and Lahiri, it does not look like a “poverty” at all. Rather, the findings of this study call for attentiveness to the *difference* between L1 and L2 style, and for considering the L2 as a resource for literary composition. Like the scientific and popular consensus has moved away from considering writing with the left hand a defect or a deficiency compared to using the right, so too should dominant beliefs about L2 “left-hand writing” be reconsidered. If we make room for L2 writing to be different rather than deficient, we are in a better position to understand the interaction between writer and language(s), as well as between language(s) and style(s).

Changing the attitude toward L2 writing may seem like a small feat, but the idea that literary L2 writing should be deficient or “styleless” is inherently linked to a predominant monolingual paradigm and privileging of the “mother tongue” in Western culture and society. Studying and encouraging L2 writing is one way of moving beyond it.

## Tools and packages

- Eder, Maciej ‘Stylometry with R: A Package for Computational Text Analysis’. *The R Journal*, vol. 8, no. 1, 2016, p. 107, <https://doi.org/10.32614/RJ-2016-007>.
- Michalke, Meik (2021). *koRpus: Text Analysis with Emphasis on POS Tagging, Readability, and Lexical Diversity*. 2021, Version 0.138, <https://reaktanz.de/?c=hacking&s=koRpus>.
- Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, 2016, <https://ggplot2.tidyverse.org>.

## Bibliography

- Aragno, Anna, and Peter J. Schlachet. ‘Accessibility of Early Experience Through the Language of Origin: A Theoretical Integration’. *Psychoanalytic Psychology*, vol. 13, no. 1, 1996, pp. 23–34.
- Arendt, Hannah. ‘Was Bleibt? Es Bleibt Die Muttersprache’. *Zur Person: Porträts in Frage Und Antwort*, by Günter Gaus, Feder Verlag, 1964, pp. 13–33.
- Arndt, Susan, et al., editors. *Exophonie: Anders-Sprachigkeit (in) der Literatur*. Kulturverlag Kadmos, 2007.
- Asher, James J., and Ramiro García. ‘The Optimal Age to Learn a Foreign Language’. *The Modern Language Journal*, vol. 53, no. 5, 1969, pp. 334–41.
- Attridge, Derek. *Peculiar Language*. Routledge, 1988.
- Bakhtin, Mihail Mihailovič. *The Dialogic Imagination: Four Essays*. Edited by Caryl Emerson and Michael Holquist, University of Texas press, 1994.
- Baldick, Chris. ‘Style’. *The Oxford Dictionary of Literary Terms*, Oxford University Press, 1 Jan. 2008. [www-oxfordreference-com.proxy.library.uu.nl](http://www.oxfordreference-com.proxy.library.uu.nl), <http://www.oxfordreference.com/view/10.1093/acref/9780199208272.001.0001/acref-9780199208272-e-1095>.
- . ‘Trope’. *The Oxford Dictionary of Literary Terms*, Oxford University Press, 2015. [www-oxfordreference-com.proxy.library.uu.nl](http://www.oxfordreference-com.proxy.library.uu.nl), <http://www.oxfordreference.com/view/10.1093/acref/9780198715443.001.0001/acref-9780198715443-e-1172>.
- Barone, Susan M., and Carrie Cargile. ‘L1 and L2 Writing Differences: From Understanding to Practice’. *English Teaching Forum*, vol. 58, no. 3, 2020, pp. 10–19.
- Beckett, Samuel. ‘Fizzles’. *The Complete Short Prose, 1929 - 1989*, edited by S. E. Gontarski, Grove Press, 2000, pp. 224–46.
- . *Pour Finir Encore Et Autres Foirades*. Éditions de Minuit, 1976. *Internet Archive*, <http://archive.org/details/pourfinirecoree00beck>.

- . 'The End'. *The Complete Short Prose, 1929 - 1989*, edited by S. E. Gontarski, Grove Press, 2000, pp. 78–99.
- Beer, Ann. 'Beckett's Bilingualism'. *The Cambridge Companion to Beckett*, edited by John Pilling, 1st ed., Cambridge University Press, 1994, pp. 209–21. *DOI.org (Crossref)*, <https://doi.org/10.1017/CCOL0521413664.012>.
- Bennett, Karen, and Rita Queiroz de Barros. *Hybrid Englishes and the Challenges of and for Translation: Identity, Mobility and Language Change*. 2020.
- Besemeres, Mary. 'Different Languages, Different Emotions? Perspectives from Autobiographical Literature'. *Journal of Multilingual and Multicultural Development*, vol. 25, no. 2–3, 2004, pp. 140–58, <https://doi.org/10.1080/01434630408666526>.
- Birdsong, David. 'Ultimate Attainment in Second Language Acquisition'. *Language*, vol. 68, no. 4, 1992, pp. 706–55.
- Blom, Elma, et al. 'Cognitive Advantages of Bilingual Children in Different Sociolinguistic Contexts'. *Frontiers in Psychology*, vol. 8, Apr. 2017, p. 552. *DOI.org (Crossref)*, <https://doi.org/10.3389/fpsyg.2017.00552>.
- Bolin, John. *Beckett and the Modern Novel*. Cambridge University Press, 2013.
- Boot, Peter, and Floor Naber. 'Exploring the Features of Naturalist Prose Using LIWC in Nederlab'. *Journal of Dutch Literature*, vol. 10, no. 1, 2019, pp. 50–70.
- Booth, Wayne C. *The Rhetoric of Fiction*. 2nd ed, University of Chicago Press, 1983.
- Brogan, Jaqueline V. 'Polyptoton'. *The Princeton Encyclopedia of Poetry and Poetics*, edited by Roland Greene et al., Fourth Edition, Princeton University Press, 2012, pp. 1086–87. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=913846>.
- Burrows, John. 'All the Way Through: Testing for Authorship in Different Frequency Strata'. *Literary and Linguistic Computing*, vol. 22, no. 1, Apr. 2007, pp. 27–47. *DOI.org (Crossref)*, <https://doi.org/10.1093/lc/fqi067>.
- . *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Clarendon Press ; Oxford University Press, 1987.
- Butler, Lance St. John. 'TWO DARKS: A Solution to the Problem of Beckett's Bilingualism'. *Samuel Beckett Today / Aujourd'hui*, vol. 3, 1994, pp. 115–35.
- Byford, Annette. 'Lost and Gained in Translation: The Impact of Bilingual Clients' Choice of Language in Psychotherapy'. *British Journal of Psychotherapy*, vol. 31, no. 3, 2015, pp. 333–47, <https://doi.org/10.1111/bjp.12148>.
- Chatman, Seymour, editor. *Literary Style: A Symposium*. Oxford University Press, 1971.
- Corver, Norbert, and Henk C. van Riemsdijk, editors. *Semi-Lexical Categories: The Function of Content Words and the Content of Function Words*. Mouton de Gruyter, 2001.
- Coulmas, Florian. 'Language Masters: Defying Linguistic Materialism'. *International Journal of the Sociology of Language*, vol. 1, no. 137, 1999, pp. 27–38.
- Craig, Hugh. 'Style, Statistics, and New Models of Authorship'. *Early Modern Literary Studies*, vol. 15, no. 1, 2009, <http://purl.oclc.org/emls/15-1/craistyl.htm>.
- Craig, Hugh, and Arthur F. Kinney. *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge University Press, 2009.

- Crossley, Scott A., and Danielle S. McNamara. 'Computational Assessment of Lexical Differences in L1 and L2 Writing'. *Journal of Second Language Writing*, vol. 18, no. 2, June 2009, pp. 119–35. *DOI.org (Crossref)*, <https://doi.org/10.1016/j.jslw.2009.02.002>.
- Daelemans, Walter. 'Explanation in Computational Stylometry'. *Computational Linguistics and Intelligent Text Processing*, edited by Alexander Gelbukh, vol. 7817, Springer Berlin Heidelberg, 2013, pp. 451–62. *DOI.org (Crossref)*, [https://doi.org/10.1007/978-3-642-37256-8\\_37](https://doi.org/10.1007/978-3-642-37256-8_37).
- Dijk, Marloes van, et al. 'Bilingualism and Creativity: Towards a Situated Cognition Approach'. *The Journal of Creative Behavior*, vol. 53, no. 2, June 2019, pp. 178–88, <https://doi.org/10.1002/jocb.238>.
- Disney, Dan. *Exploring Second Language Creative Writing: Beyond Babel*. John Benjamins Publishing Company, 2014. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=1715261>.
- Dolven, Jeffrey. 'Style'. *The Princeton Encyclopedia of Poetry and Poetics: Fourth Edition*, edited by Roland Greene et al., Princeton University Press, 2012, pp. 1369–70. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=913846>.
- Dorfman, Ariel. 'Footnotes to a Double Life'. *The Genius of Language: Fifteen Writers Reflect on Their Mother Tongues*, edited by Wendy Lesser, Pantheon Books, 2004, pp. 216–17.
- Eder, Maciej. 'Does Size Matter? Authorship Attribution, Small Samples, Big Problem'. *Digital Scholarship in the Humanities*, vol. 30, no. 2, June 2015, pp. 167–82. *DOI.org (Crossref)*, <https://doi.org/10.1093/lc/fqt066>.
- . 'Performance Measures in Supervised Classification'. *Computational Stylistics Group*, [https://computationalstylistics.github.io/blog/performance\\_measures/](https://computationalstylistics.github.io/blog/performance_measures/).
- . 'Style-Markers in Authorship Attribution A Cross-Language Study of the Authorial Fingerprint'. *Studies in Polish Linguistics*, vol. 6, 2011, pp. 101–16.
- Eder, Maciej, and Jan Rybicki. 'Do Birds of a Feather Really Flock Together, or How to Choose Training Samples for Authorship Attribution'. *Literary and Linguistic Computing*, vol. 28, no. 2, June 2013, pp. 229–36. *DOI.org (Crossref)*, <https://doi.org/10.1093/lc/fqs036>.
- Ervin, Susan. 'An Analysis of the Interaction of Language, Topic, and Listener.' *Readings in the Sociology of Language*, edited by Joshua A Fishman, 1968, <http://www.degruyter.com/doi/book/10.1515/9783110805376>.
- Fausey, Caitlin M., and Lera Boroditsky. 'Who Dunit? Cross-Linguistic Differences in Eye-Witness Memory'. *Psychonomic Bulletin & Review*, vol. 18, no. 1, Feb. 2011, pp. 150–57. *DOI.org (Crossref)*, <https://doi.org/10.3758/s13423-010-0021-5>.
- Ferré, Rosario. 'Bilingual in Puerto Rico'. *Switching Languages: Translingual Writers Reflect on Their Craft*, edited by Steven G. Kellman, University of Nebraska Press, 2003, pp. 135–38.
- Fialho, Olivia, and Sonia Zyngier. 'Quantitative Methodological Approaches to Stylistics'. *The Routledge Handbook of Stylistics*, edited by Michael Burke, Routledge, Taylor & Francis Group, 2014, pp. 329–45.

- Forster, Leonard. *The Poet's Tongues: Multilingualism in Literature. (1. Publ.)*. Cambridge UP, 1970.
- Foster, RoseMarie Perez. 'Psychoanalysis and the Bilingual Patient: Some Observations on the Influence of Language Choice on the Transference'. *Psychoanalytic Psychology*, vol. 9, no. 1, 1992, pp. 61–76.
- Genette, Gérard. *Narrative Discourse: An Essay in Method*. Cornell University Press, 1980.
- Germoni, Karine. 'Les Anglicismes De Beckett À L' Époque De La "French Frenzy": Le Cas De La Virgule "Anglaise", Du Sous-Commatisme Et Des Connecteurs Logiques'. *Samuel Beckett Today / Aujourd'hui*, vol. 30, no. 1, Mar. 2018, pp. 97–112. *brill.com*, <https://doi.org/10.1163/18757405-03001007>.
- Germoni, Karine, and Pascale Sardin. 'Tensions of the In-Between: Rhythm, Tonelessness and Lyricism In'. *Samuel Beckett Today / Aujourd'hui*, vol. 24, no. 1, Dec. 2012, pp. 335–51. *DOI.org (Crossref)*, <https://doi.org/10.1163/18757405-024001023>.
- Gontarski, S. E. 'Introduction'. *The Complete Short Prose, 1929 - 1989*, Grove Press, 2000, pp. xi–xxxii.
- Grosjean, François. *Bilingual: Life and Reality*. Harvard University Press, 2010. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=3301081>.
- Hall, Geoff. *Literature in Language Education*. Second edition, Palgrave Macmillan, 2015.
- Halliday, Michael Alexander Kirkwood. 'Linguistic Function and Literary Style: An Inquiry Into the Language of William Golding's *The Inheritors*.' *Literary Style: A Symposium*, edited by Seymour Chatman, Oxford University Press, 1971, pp. 330–68.
- Halteren, Hans van, et al. 'New Machine Learning Methods Demonstrate the Existence of a Human Stylome'. *Journal of Quantitative Linguistics*, vol. 12, no. 1, Apr. 2005, pp. 65–77. *DOI.org (Crossref)*, <https://doi.org/10.1080/09296170500055350>.
- Hamel, Rainer Enrique. 'The Dominance of English in the International Scientific Periodical Literature and the Future of Language Use in Science'. *AILA Review*, vol. 20, Dec. 2007, pp. 53–71. *DOI.org (Crossref)*, <https://doi.org/10.1075/aila.20.06ham>.
- Hanauer, David I. 'Appreciating the Beauty of Second Language Poetry Writing'. *Exploring Second Language Creative Writing: Beyond Babel*, edited by Dan Disney, John Benjamins Publishing Company, 2014, pp. 11–22.
- . 'Measuring Voice in Poetry Written by Second Language Learners'. *Written Communication*, vol. 32, no. 1, Jan. 2015, pp. 66–86. *SAGE Journals*, <https://doi.org/10.1177/0741088314563023>.
- Harris, Catherine L., et al. 'Taboo Words and Reprimands Elicit Greater Autonomic Reactivity in a First Language Than in a Second Language'. *Applied Psycholinguistics*, vol. 24, no. 4, Dec. 2003, pp. 561–79, <https://doi.org/10.1017/S0142716403000286>.
- Haverals, Wouter, et al. 'A Style for Every Age: A Stylometric Inquiry into Crosswriters for Children, Adolescents and Adults'. *Language and Literature*, Jan. 2022, p. 09639470211072163. *SAGE Journals*, <https://doi.org/10.1177/09639470211072163>.
- Helgesson, Stefan, and Christina Kullberg. 'Translingual Events: World Literature and the Making of Language'. *Journal of World Literature*, vol. 3, no. 2, 2018, pp. 136–52, <https://doi.org/10.1163/24056480-00302002>.

- Herrmann, Berenike J., et al. 'Revisiting Style, a Key Concept in Literary Studies'. *Journal of Literary Theory*, vol. 9, no. 1, 2015, <https://doi.org/10.1515/jlt-2015-0003>.
- Hirst, Graeme, and Vanessa Wei Feng. 'Changes in Style in Authors with Alzheimer's Disease'. *English Studies*, vol. 93, no. 3, May 2012, pp. 357–70. *DOI.org (Crossref)*, <https://doi.org/10.1080/0013838X.2012.668789>.
- Hoover, David. *Modes of Composition and the Durability of Style in Literature*. Routledge, 2020, <https://doi.org/10.4324/9780429348068>.
- . 'Zeta Revisited'. *Digital Scholarship in the Humanities*, Dec. 2021, p. fqab095. *DOI.org (Crossref)*, <https://doi.org/10.1093/llc/fqab095>.
- Hulle, Dirk Van, and Mike Kestemont. 'Periodizing Samuel Beckett's Works: A Stylochronometric Approach'. *Style*, vol. 50, no. 2, 2016, pp. 172–202.
- Jarvis, Scott. 'Capturing the Diversity in Lexical Diversity'. *Language Learning*, vol. 63, no. s1, 2013, pp. 87–106. *Wiley Online Library*, <https://doi.org/10.1111/j.1467-9922.2012.00739.x>.
- Javier, Rafael Art. 'Vicissitudes of Autobiographical Memories in a Bilingual Analysis'. *Psychoanalytic Psychology Summer 1995*, vol. 12, no. 3, 1995, pp. 429–38, <https://doi.org/10.1037/h0079703>.
- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013.
- Johnson, Jacqueline S., and Elissa L. Newport. 'Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language'. *Cognitive Psychology*, vol. 21, no. 1, 1989, pp. 60–99, [https://doi.org/10.1016/0010-0285\(89\)90003-0](https://doi.org/10.1016/0010-0285(89)90003-0).
- Kager, Maria. 'Comment Dire: A Neurolinguistic Approach to Beckett's Bilingual Writings'. *L2 Journal*, vol. 7, no. 1, Feb. 2015. *DOI.org (Crossref)*, <https://doi.org/10.5070/L27124380>.
- Kellman, Steven G., editor. *Switching Languages: Translingual Writers Reflect on Their Craft*. University of Nebraska Press, 2003.
- . *The Translingual Imagination*. Nebraska, 2000. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=1763078>.
- Kellman, Steven G., and Natasha Lvovich. 'Selective Bibliography of Translingual Literature'. *L2 Journal*, vol. 7, no. 1, Feb. 2015. *DOI.org (Crossref)*, <https://doi.org/10.5070/L27123971>.
- Kestemont, M., et al. 'Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux'. *Digital Scholarship in the Humanities*, vol. 30, no. 2, June 2015, pp. 199–224. *DOI.org (Crossref)*, <https://doi.org/10.1093/llc/fqt063>.
- Kharkhurin, Anatolij V. 'Bilingual Verbal and Nonverbal Creative Behavior'. *International Journal of Bilingualism*, vol. 14, no. 2, June 2010, pp. 211–26, <https://doi.org/10.1177/1367006910363060>.
- Kilgarriff, Adam. 'Comparing Corpora'. *International Journal of Corpus Linguistics*, vol. 6, no. 1, 2001, pp. 97–133, <https://doi.org/10.1075/ijcl.6.1.05kil>.
- Kosiński, Jerzy. *Conversations with Jerzy Kosinski*. University Press of Mississippi, 1993.

- Kremnitz, Georg. *Mehrsprachigkeit in der Literatur: Aus der Sicht der Soziologie der Kommunikation*. Edition Praesens, 2004.
- Lahiri, Jhumpa. 'I Am, in Italian, a Tougher, Freer Writer'. *The Observer*, 31 Jan. 2016. *The Guardian*, <https://www.theguardian.com/books/2016/jan/31/jhumpa-lahiri-in-other-words-italian-language>.
- . *In Other Words*. Bloomsbury, 2015.
- . Interview, *Whereabouts, with Howard Norman*. Interview by Howard Norman, video, 5 May 2021. *YouTube*, <https://www.youtube.com/watch?v=DwiFG9TuFIE>.
- . 'Where I Find Myself: On Self-Translation'. *Words Without Borders*, <https://wordswithoutborders.org/read/article/2021-04/april-2021-where-i-find-myself-on-self-translation-jhumpa-lahiri/>. Accessed 1 May 2021.
- Laufer, Batia, and Paul Nation. 'Vocabulary Size and Use: Lexical Richness in L2 Written Production'. *Applied Linguistics*, vol. 16, no. 3, Sept. 1995, pp. 307–22. *DOI.org (Crossref)*, <https://doi.org/10.1093/applin/16.3.307>.
- Leech, Geoffrey N., and Mick Short. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. 2nd ed, Pearson Longman, 2007.
- Leikin, Mark, et al. 'The Interplay of Bilingualism, Executive Functions and Creativity in Problem Solving Among Male University Students'. *Creativity Studies*, vol. 13, no. 2, May 2020, pp. 308–24, <https://doi.org/10.3846/cs.2020.10397>.
- Lissón, Paula, and Nicolas Ballier. 'Investigating Lexical Progression through Lexical Diversity Metrics in a Corpus of French L3'. *Discours*, no. 23, Dec. 2018. *DOI.org (Crossref)*, <https://doi.org/10.4000/discours.9950>.
- Marian, Viorica, and Margarita Kaushanskaya. 'Self-Construal and Emotion in Bicultural Bilinguals'. *Journal of Memory and Language*, vol. 51, no. 2, 2004, pp. 190–201, <https://doi.org/10.1016/j.jml.2004.04.003>.
- Mukařovský, Jan. 'Standard Language and Poetic Language'. *A Prague School Reader on Esthetics Literary Structure, and Style*, edited by Paul L. Garvin, 1932. Georgetown University Press, 1964, pp. 17–30, [https://monoskop.org/images/3/3a/A\\_Prague\\_School\\_Reader\\_on\\_Esthetics\\_Literary\\_Structure\\_and\\_Style\\_1964.pdf](https://monoskop.org/images/3/3a/A_Prague_School_Reader_on_Esthetics_Literary_Structure_and_Style_1964.pdf).
- Oyama, Susan. 'A Sensitive Period for the Acquisition of a Nonnative Phonological System'. *Journal of Psycholinguistic Research*, vol. 5, no. 3, 1976, pp. 261–83, <https://doi.org/10.1007/BF01067377>.
- . 'The Sensitive Period and Comprehension of Speech'. *NABE Journal*, vol. 3, no. 1, 1978, pp. 25–40, <https://doi.org/10.1080/08855072.1978.10668342>.
- Pavlenko, Aneta. 'Affective Processing in Bilingual Speakers: Disembodied Cognition?'. *International Journal of Psychology*, vol. 47, no. 6, 2012, pp. 405–28, <https://doi.org/10.1080/00207594.2012.743665>.
- . *Emotions and Multilingualism*. Cambridge University Press, 2006, <https://doi.org/10.1017/CBO9780511584305>.
- Peer, Willie van. *Stylistics and Psychology: Investigations of Foregrounding*. Croom Helm, 1986.

- Pennebaker, James W. 'How Authors' Forgettable Words Reveal Their Personality and Social Behaviors'. *Information Design Journal*, vol. 20, no. 3, Dec. 2013, pp. 267–72. DOI.org (Crossref), <https://doi.org/10.1075/idj.20.3.062pen>.
- . *The Secret Life of Pronouns: What Our Words Say About Us*. Paperback ed, Bloomsbury, 2013.
- Piller, Ingrid. 'Passing for a Native Speaker: Identity and Success in Second Language Learning'. *Journal of Sociolinguistics*, vol. 6, no. 2, 2002, pp. 179–208, <https://doi.org/10.1111/1467-9481.00184>.
- Rabinovitz, Rubin. "'Fizzles" and Samuel Beckett's Earlier Fiction'. *Contemporary Literature*, vol. 24, no. 3, 1983, pp. 307–21. JSTOR, <https://doi.org/10.2307/1208080>.
- Reid, Joy. 'A Computer Text Analysis of Four Cohesion Devices in English Discourse by Native and Nonnative Writers'. *Journal of Second Language Writing*, vol. 1, no. 2, May 1992, pp. 79–107. DOI.org (Crossref), [https://doi.org/10.1016/1060-3743\(92\)90010-M](https://doi.org/10.1016/1060-3743(92)90010-M).
- Rybicki, J., and M. Eder. 'Deeper Delta across Genres and Languages: Do We Really Need the Most Frequent Words?' *Literary and Linguistic Computing*, vol. 26, no. 3, Sept. 2011, pp. 315–21. DOI.org (Crossref), <https://doi.org/10.1093/lc/fqr031>.
- Sardin-Damestoy, Pascale. 'Chapitre IV: Un Art De L'inflexion (Ou Le Travail Sur L'axe Paradigmatique)'. *Samuel Beckett Auto-Traducteur Ou L'art De L'« Empêchement »*, Artois Presses Université, 2020, pp. 123–53. OpenEdition Books, <http://books.openedition.org/apu/5981>.
- Seliger, Herbert W., et al. 'Maturational Constraints in the Acquisition of Second Language Accent'. *Language Sciences*, 1975, pp. 20–22.
- Slote, Sam. 'Bilingual Beckett: Beyond the Linguistic Turn'. *The New Cambridge Companion to Samuel Beckett*, edited by Dirk Van Hulle, 1st ed., Cambridge University Press, 2015, pp. 114–25. DOI.org (Crossref), <https://doi.org/10.1017/CCO9781139871525.013>.
- Sommer, Doris. *Bilingual Aesthetics: A New Sentimental Education*. Duke University Press, 2004. ProQuest Ebook Central, <http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=1167990>.
- Stamatatos, Efstathios. 'A Survey of Modern Authorship Attribution Methods'. *Journal of the American Society for Information Science and Technology*, vol. 60, no. 3, 2009, pp. 538–56. Wiley Online Library, <https://doi.org/10.1002/asi.21001>.
- Tarlinskaja, Marina. *Shakespeare's Verse: Iambic Pentameter and the Poet's Idiosyncrasies*. Peter Lang, 1987.
- Tibshirani, Robert, et al. 'Class Prediction by Nearest Shrunken Centroids, with Applications to DNA Microarrays'. *Statistical Science*, vol. 18, no. 1, Feb. 2003. DOI.org (Crossref), <https://doi.org/10.1214/ss/1056397488>.
- Torruella, Joan, and Ramon Capsada. 'Lexical Statistics and Tipological Structures: A Measure of Lexical Richness'. *Procedia - Social and Behavioral Sciences*, vol. 95, Oct. 2013, pp. 447–54. DOI.org (Crossref), <https://doi.org/10.1016/j.sbspro.2013.10.668>.

- Vince, Gaia. *The Amazing Benefits of Being Bilingual*.  
<https://www.bbc.com/future/article/20160811-the-amazing-benefits-of-being-bilingual>. Accessed 7 Apr. 2022.
- Wilson, Rita. “‘Pens That Confound the Label of Citizenship’: Self-Translations and Literary Identities”. *Modern Italy*, vol. 25, no. 2, 2020, pp. 213–24,  
<https://doi.org/10.1017/mit.2019.73>.
- Yan, Jing, and Xiao-Yan Xu. ‘The Relationship between Syntactic Complexity and Writing Quality of Chinese EFL Learners’. *Proceedings of the 3rd Annual International Conference on Social Science and Contemporary Humanity Development*, Atlantis Press, 2017. *DOI.org (Crossref)*, <https://doi.org/10.2991/sschd-17.2017.65>.
- Yildiz, Yasemin. *Beyond the Mother Tongue: The Postmonolingual Condition*. Fordham University Press, 2012.
- Zyngier, S. ‘Stylistics: Pedagogical Applications’. *Encyclopedia of Language and Linguistics*, edited by Keith Brown, 2nd ed., Elsevier, pp. 226–32.