Univerzita Karlova
Pedagogická fakulta
Katedra anglického jazyka a literatury

BAKALÁŘSKÁ PRÁCE

Jazyková komplexnost knih J.K. Rowlingové o Harry Potterovi

Linguistic complexity of J.K. Rowling's Harry Potter series

Michal Gottwald

Vedoucí práce:	doc. PhDr. Markéta Malá, Ph.D.
Studijní program:	Specializace v pedagogice (B7507)
Studijní obor:	B AJ-NJ (7507R036, 7507R041)

2023

Odevzdáním této bakalářské práce na téma *Jazyková komplexnost knih J.K. Rowlingové o Harry Potterovi* potvrzuji, že jsem ji vypracoval pod vedením vedoucího práce samostatně za použití v práci uvedených pramenů a literatury. Dále potvrzuji, že tato práce nebyla využita k získání jiného nebo stejného titulu.

V Praze dne 3.12.2023

Tímto bych chtěl poděkovat paní doc. PhDr. Markétě Malé, PhD. za vedení mé bakalářské práce. Zejména za obrovskou vstřícnost a nespočet neocenitelných rad a návrhů, které výrazně přispěly k finální podobě této práce.

ABSTRAKT

Bakalářská práce se zabývá jazykovou složitostí prvního a posledního dílu knižní série J. K. Rowlingové o Harry Potterovi. Ověřuje hypotézu, že jazykové prostředky těchto knih, které ovlivňují srozumitelnost, jsou přizpůsobeny vyvíjejícím se čtenářským dovednostem a schopnostem čtenářů. Očekává se, že s postupujícím vývojem série se složitost bude zvyšovat, což reflektuje vyspělost jejího čtenářského publika. Tento výzkum zjišťuje, zda jsou pozdější díly série syntakticky a lexikálně složitější než díly dřívější, a to za pomoci využití korpusové stylistické metodologie. Teoretická část práce nastiňuje definici jazykové složitosti a její vztah ke srozumitelnosti, ale také pojednává o rozmanitých metodách analýzy literatury z lingvistického hlediska. Jazyková složitost je měřena pomocí několika kritérií: počtu tokenů, typů, kapitol, vět, rozsahu i typů větné podřadnosti a nedokonavých vět. Zohledňuje se také délka věty a poměr typů a tokenů. Metody dále zahrnují frekvenční seznamy a analýzu klíčových slov, doplněné podrobným zkoumáním textových ukázek.

KLÍČOVÁ SLOVA

jazyková komplexnost, literatura pro děti, stavba věty a souvětí

ABSTRACT

This bachelor thesis examines the language complexity in the initial and final volumes of J.K. Rowling's Harry Potter series, testing the hypothesis that the novels' linguistic features, impacting readability, are adjusted to the intended readers' evolving comprehension and reading skills. As the series progresses, the complexity is expected to increase, reflecting the maturation of its audience. The research investigates whether the series' later novels are syntactically and lexically more complex than the earlier ones, with the help of employing corpus stylistic methodology. The theoretical part of this thesis outlines the definition of linguistic complexity and its relation to readability and discusses various methods of analyzing literature from a linguistic perspective. The complexity is measured using various criteria: the number of tokens, types, chapters, sentences, the extent and types of clausal subordination and non-finite clauses. Sentence length and type-token ratio are also considered. The methods include frequency lists and keyword analysis, complemented by a detailed examination of text samples.

KEYWORDS

complexity, children’s literature, sentence structure

Contents

List of tables

List of figures

1 Introduction

This thesis delves into the linguistic complexity of J.K. Rowling's acclaimed Harry Potter series, specifically focusing on comparing the first and last volumes. What makes this topic particularly interesting is the unique opportunity to explore how Rowling's narrative style may have evolved to adapt to her audience's maturing comprehension skills over the series. This exploration is applicable due to the series' broad age range of readers and the significant developmental changes they undergo.

The central research question guiding this thesis is: Does Rowling adjust the linguistic complexity of her books for children depending on the age of the intended reader? Operationalizing this involves questioning whether the first book of the Harry Potter series, known for its lighter tone, exhibits less linguistic complexity compared to the last volume, which evolves into a darker narrative. The hypothesis stated here is affirmative; it is anticipated that the lexical complexity will increase along with specific lexical choices related to the themes of the selected volumes. This hypothesis is rooted in the assumption that as the series progresses, so does the age and, correspondingly, the reading proficiency of its core audience.

The thesis comprises two major parts. Firstly, it provides a theoretical background, outlining the definition of linguistic complexity and its relation to readability, as well as discussing various methods of analyzing literature through the lens of language. This section aims to establish a solid foundation of understanding for the subsequent analysis.

The second part of the thesis consists of both macro- and micro-analyses. The macro-analysis covers both lexical and grammatical complexity, involving examination of such factors as the number of tokens, types, chapters, sentences, or Type-Token Ratios (TTR), and lists of keywords and unique words, with a particular focus on comparing the first and last volumes of the Harry Potter series. The micro-analysis an provides in-depth analysis of sentences and clauses, including the number of clauses per sentence or its length. Moreover, the analysis examines dependent clauses (finite and non-finite) and their syntactic functions, both at the phrase and sentence level. Such a comprehensive approach enables a detailed and insightful analysis of Rowling's linguistic choices and their potential implications for the series' readability and audience engagement.

2 Harry Potter series

The Harry Potter series, penned by J.K. Rowling, stands as a monumental achievement in contemporary literature, particularly in the realm of fantasy and young adult fiction. This series spans seven books and narrates the journey of its central character, Harry Potter, through his years at Hogwarts School of Witchcraft and Wizardry. Rowling's narrative is rich in its exploration of themes such as friendship, courage, and the struggle between good and evil, deeply resonating with a diverse, global audience. Her storytelling is marked by a complex interplay of characters and an intricately designed magical world, which have not only attracted a vast audience but also significantly influenced popular culture. The series' impact extends beyond entertainment, contributing to academic discourse in areas such as literature or linguistics. Rowling's adept use of language and her ability to weave intricate plots and subplots have been widely acclaimed, making the Harry Potter books a subject of scholarly interest and a window into understanding narrative techniques and character development.

2.1 Harry Potter and the Philosopher's Stone (HP1)

'Harry Potter and the Philosopher's Stone' (or *'The Sorcerer's Stone'* in American Edition), the opening book of the series, serves as the gateway into the magical universe created using J.K. Rowling. This novel introduces readers to the young wizard, Harry Potter, and sets the stage for the epic saga that unfolds in the subsequent books. In the first volume, Rowling's narrative style is particularly noteworthy for its simplicity and clarity, aimed at capturing the imagination of a younger audience. The book establishes the foundational elements of the series, including key characters, the Hogwarts setting, and the central conflict involving the dark wizard Voldemort. Its language is characterized by a straightforward structure, with a focus on vivid descriptions and an engaging narrative pace. The use of imaginative elements, such as magical creatures and spells, adds to the book's appeal, making it a compelling read for children and adults alike. HP1's contribution to the series is crucial, as it not only introduces the major themes and characters but also sets the tone for the complexity and depth that develop in the later books.

The plot unfolds as Harry Potter, an orphan living a mundane life with his neglectful relatives, discovers his identity as a wizard on his eleventh birthday. Having arrived at

Hogwarts, a school for witchcraft and wizardry, Harry embarks on a fantastical journey of self-discovery and friendship. Alongside his friends Hermione and Ron, Harry encounters the wonders and dangers of the wizarding world. The story reaches its peak as they unravel a series of mysteries at Hogwarts, leading to the discovery of the Philosopher's Stone, an object capable of granting eternal life and unlimited wealth sought by the dark wizard Voldemort. The climax sees Harry confronting Professor Quirrell, who is revealed to be hosting Voldemort at the back of his head, setting the stage for their ongoing conflict.

2.2 Harry Potter and the Deathly Hallows (HP7)

In contrast, *'Harry Potter and the Deathly Hallows'*, the series' final instalment, marks a significant shift in tone and complexity from the earlier books. This novel is characterized by darker themes, more complex narrative structures, and a deeper exploration of the characters' psychological dimensions. Rowling's writing in HP7 reflects the maturation of both the characters and her target audience, with more sophisticated language, intricate plot developments, and a focus on themes such as mortality, sacrifice, and the nuances of moral choices. The book brings the long-standing conflict between Harry Potter and Voldemort to a climactic resolution, weaving together many threads laid out in the previous books. HP7's narrative is notable for its emotional depth, the culmination of character arcs, and the resolution of key mysteries, making it a dense and compelling read. The evolution in Rowling's writing style from HP1 to HP7 offers a rich linguistic and literary analysis field, highlighting her adaptability and growth as an author.

The story reaches its climax as Harry, now older and more seasoned, faces the ultimate challenge against the dark forces led by Voldemort. The narrative diverges from the familiar setting of Hogwarts, depicting Harry, Hermione, and Ron on a perilous quest to find and destroy Voldemort's Horcruxes – objects containing fragments of his soul. Their journey is fraught with danger, sacrifice, and loss, reflecting the grim reality of the magical war they are engaged in. As the plot unfolds, the trio uncovers the legend of the Deathly Hallows, ancient magical artefacts of immense power which play a crucial role in the final battle between Harry and Voldemort. The climax of the book is set in a fiercely fought battle at Hogwarts, where characters from previous books unite to defend the wizarding world against Voldemort and his followers. The resolution of the battle brings not only an end to

Voldemort's reign of terror but also signifies a new era of peace and rebuilding for the wizarding world. The book concludes with a touching scene at King's Cross Station, where a grown-up Harry and his friends watch their children embark on their journey to Hogwarts. This scene symbolizes the ongoing cycle of life and the enduring legacy of their experiences. It is a fitting end to the series, providing closure to the epic tale while leaving the door open to the imagination for future adventures.

3 Linguistic complexity

3.1 Definition

Linguistic complexity is a challenging concept to define, and there is still no universally accepted definition. Complexity can refer to various linguistic properties, such as syntax, morphology, and phonology. However, measuring these properties is complex, and different scholars have used various approaches, resulting in conflicting outcomes (Bult & Housen 2012:22). Additionally, what one linguist considers complex might be viewed as something other than complex by another. For instance, some may view complex syntax as more complex than complex morphology, while others may disagree. Ultimately, a universally accepted definition of linguistic complexity is needed to compare findings across studies and to draw generalizations about language complexity. By the nature of this bachelor's thesis, two approaches to the definition of complexity are compared.

As a representative of the first approach, Haueng discusses in his '*Development of linguistic complexity in Harry Potter*' (2012) linguistic complexity and its various definitions and features, as evaluated by different linguistic theories. The theories comprise traditional grammar, transformational-generative grammar, and functional grammar:

1. Traditional grammar, based on Greek and Latin languages, uses a prescriptive system to determine correct and incorrect grammar rules, with complexity measured by 'markedness'. The unmarked features of a clause are represented by main, declarative, affirmative, and active clauses. The marked features are conversely represented by subordination, interrogative, imperative, negative, and passive clauses.
2. Transformational-generative grammar, pioneered by Noam Chomsky, primarily focuses on syntactic structure. In this case, the complexity is determined by the length of sentences and tree structures that relate to complex interpretations of syntactic structures.
3. Functional grammar, developed by Michael Halliday, uses six categories to measure complexity, including lexical density and nominalization.

Lexical density measures the proportion of content words (nouns, verbs, adjectives and often also adverbs) and function words to the total number of words in a text. As far as the content

words are concerned, they belong to an open set that can be expanded to include also adverbs created from adjectives.

The other approach is represented by Bulté & Housen in their '*Defining and operationalising L2 complexity*' (2012). The reader may be met with the fact that a significant number of L2 studies examining 'complexity' either fail to clearly define the term or, when a definition is provided, it is often general, imprecise, or circular. This argument is supported by the following examples cited in their work:

1. '[complexity is the] use of more challenging and difficult language … Complexity is the extent to which learners produce elaborated language' (R. Ellis & Barkhuizen 2005: 139)[1]
2. 'Grammatical and lexical complexity mean that a wide variety of both basic and sophisticated structures and words are available to the learner'. (Wolfe-Quintero, Inagaki, & Kim 1998: 69, 101)[2] (Bulté & Housen 2012:22)

However, Bulté & Housen (2012) advocate that in linguistics, language complexity can be divided into two categories; relative and absolute complexity. Relative complexity, also called cognitive complexity, is defined in terms of the mental effort or resources that language users and learners have to invest in processing or internalizing a language feature or system of features. This approach views a language feature as complex if it is "costly or taxing for language users and learners" (Ibid.: 23). It should be noted that cognitive complexity is, to some extent, subjective and learner-dependent. However, there are also objective factors that can contribute to ease or difficulty, such as perceptual salience and frequency of occurrence in the input.

Conversely, absolute complexity defines language complexity based on objective, measurable terms such as the number of distinct elements and connections in a language feature or system. The inherent complexity of a language feature is just one of the factors that could influence the difficulty of learning or processing it. Moreover, there is not always

[1] Ellis, R., & Barkhuizen, G. 2005. *Analysing learner language.* Oxford: Oxford University Press.
[2] Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. 1998. *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawaii, Second Language Teaching & Curriculum Center.

a direct correlation between inherent complexity and the level of difficulty in processing or learning.

3.2 Readability

In '*Syntax, Readability and Ideology in Children's Literature*' (1998), Puurtinen suggests a link between 'complexity' and 'readability'. Together, they offer a nuanced understanding of how linguistic complexity can affect a reader's engagement with a text. The key idea is that there is a balance to be struck: a book should be complex enough to be engaging and challenging, but not so complex as to become inaccessible or frustrating for the intended audience. 'Readability, or ease of reading and understanding determined by linguistic difficulty, is one aspect of comprehensibility. Presently the concept is also understood to cover speakability (the term from Snell-Hornby 1988: 35[3] cited in Puurtinen 1998), i.e. the suitability of a text to be read aloud fluently, which is one of the most important qualities of children's books as children's books are read aloud not only by adults but also by children themselves." (Puurtinen 1998: 526)

3.3 Linguistic complexity in fiction

Linguistic complexity in literature can manifest differently due to a variety of factors related to the author's intentions, the genre and style of the work, the intended audience, and the thematic content. The next chapter focuses on how linguistic complexity is utilized within children's literature and popular fiction. The genres are chosen due to the comprehensive linguistic researches underpinning the principle. Furthermore, the investigation into different genres lays the groundwork for understanding the varied intentions behind linguistic complexity in literature, setting the stage for an in-depth analysis of J.K. Rowling's work, particularly the initial and final volumes of the Harry Potter series. The former is categorized as a children's book, while the latter may be classified as popular fiction.

[3] Snell-Hornby, M. 1988. *Translation Studies. An Integrated Approach*. Amsterdam/Philadelphia: John Benjamins.

3.3.1 Literature for children

Children's literature, though often overlooked in scholarly research, serves an essential role in education, society, and ideology. It not only entertains and helps develop reading skills but also conveys crucial knowledge, values, and behaviour. Didacticism, the idea of providing beneficial content, is a core component of children's books, whether it is clearly stated or subtly implied. This educational aspect is balanced with the need to make the content understandable, adapting the language and themes to fit the reading and comprehension levels of young readers. (Puurtinen 1998) This view is shared by Durão et al. (2018), who agree that "the language skills of children and adolescents are still being built, mostly depending on the formal level of education in which they are enrolled" (Durão et al. 2018: 29).

3.3.2 Popular fiction

Popular fiction is characterized by its accessible and reader-friendly linguistic and syntactic features "which separates genre fiction from Literature with a capital 'L'."(Montoro 2018: 63). It often utilizes simplified language, conversational style, and direct dialogue to engage a wide readership. This genre tends to underuse complex syntactic structures like post-modifying prepositional *of*-phrases, adjectives, and the *wh*-determiners (Montoro & McIntyre 2019). Despite its linguistic simplicity, popular fiction excels in creating engaging narratives, employing vivid imagery, and addressing universal themes, making it a beloved and relatable form of literature for a diverse audience. (Ibid.)

4 Methods of analyzing literature through language

Linguistic complexity, a key aspect of language analysis, manifests uniquely across various genres and registers, underscoring the diverse nature of language use. This complexity, often gauged through structural, lexical, and syntactic dimensions, differs significantly from one context to another, influenced by factors such as purpose, audience, and cultural norms. To demonstrate, as found for example by Biber et al. (2011), the types of complexity typically found in academic writing significantly differ from those prevalent in conversational language. "Thus, complexity is not a single unified construct, and it is therefore not reasonable to suppose that any single measure will adequately represent this construct" (Biber et al. 2011: 29). This distinction is essential and sets the stage for a deeper exploration of various methods used to assess linguistic complexity. The following paragraphs delve into several theoretical approaches and methods supported by studies that illuminate how complexity can be measured and interpreted differently across diverse linguistic contexts.

4.1 Lexical diversity, repetition, and density

When considering the complexity of language, vocabulary is often the first aspect that comes to mind. This is typically assessed by examining two main factors: the variety of words used (lexical diversity) and the frequency with which they are repeated (lexical repetition). Consequently, analyzing how often words appear in a text is a fundamental initial step in this evaluation. This examination is conducted using corpus linguistics (CL) software, which offers users lists of words along with their frequency rates. Subsequently, additional CL tools utilize this data to compute Type-Token Ratios (TTR) and Standard Type-Token Ratios (STTR), which gauge the levels of lexical repetition and diversity within texts in proportion to the overall word count (tokens).

An alternative approach to assessing the vocabulary level of texts, mentioned by Durão et al. (2018), involves evaluating their lexical density. This differs from lexical diversity as it incorporates an assessment of word complexity, rather than merely counting the number of unique words in texts without regard for their complexity. In essence, a text can exhibit lexical complexity without being lexically diverse, or vice versa. This means that while lexical diversity focuses solely on the quantity of different words present, lexical density provides a deeper analysis by considering the intricacy of the words used. To determine the

lexical density of texts CL software is employed. It sorts the vocabulary within a text into categories based on the complexity of the words. The criteria for complexity are derived from lists of the most frequently used words in language corpora, which are considered as statistically representative of the language as a whole. The underlying principle of this measure is that the more frequently a word is used, the less complex it is considered to be, since such words are commonly recognized and understood by speakers of the language.

4.2 Sentence length

The analysis of the language level of texts extends beyond mere vocabulary examination; it also includes the examination of sentence length. This is because the longer the sentences are, the more difficult they become to read and process, as their complex structure requires higher levels of reading and comprehension skills compared to shorter sentences. This kind of analysis is typically conducted with the help of CL software. "many word counters provide their users with numbers of sentences containing, say, 2, 3 4, 5 words per sentence and calculate average length of sentences, in a way that allows users to overview levels of complexity of texts" (Durão et al. 2018: 31).

Implicit in this approach is the concept of text readability, which can be quantitatively assessed using formulas like the Gunning Fog Index, developed in 1968 by Gunning. (Ibid.) This index employs mathematical relationships involving the average sentence length and the number of complex words to provide a measure of text readability.

4.3 Keywords

The investigation of two texts of different genres involves comparing the frequency of lexical items and/or morphosyntactic structures within a text against their occurrence rates in larger corpora (reference corpora), which are regarded statistically representative of a specific language or textual genre. In this context, CL software proves invaluable, providing comprehensive lists of keywords for analysis. Keywords are 'the words of a text whose frequencies are unusually high or low in relation to a reference corpus" (Durão et al. 2018: 32).

When constructing reference corpora, it is crucial to include texts that are contemporary with the texts being studied. Additionally, it is important to ensure that the size of the two corpora

is comparable, and as suggested by Berber Sardinha (2004)[4], the reference corpus should be at least five times larger than the texts under investigation (Durão et al. 2018: 33).

4.4 Subordinating conjunctions

Montoro (2018) refers to the work of other scholars, specifically Ramet (1999)[5] and Moretti (2010)[6], who have explored the distinction between popular fiction and serious fiction by examining the linguistic differences between the two. However, the author critiques these attempts, suggesting that while they have made efforts to distinguish between popular and serious fiction based on their language and style, their analyses haven't always provided a thorough and precise explanation of how exactly the two types of texts differ linguistically. Essentially, the author is acknowledging the attempts to categorize these genres based on language use, but also pointing out the limitations or lack of detail in these existing studies. 'This, of course, does not imply that these critics draw erroneous conclusions but simply that they do not draw upon replicable and/or falsifiable tenets." (Montoro, 2018: 64)

Montoro and McIntyre (2019) conducted a study to examine the idea that the use of subordination (a grammatical technique involving the use of subordinate clauses to create complex sentences) contributes to the perception of complexity in both popular and serious fiction. Their research aimed to test whether serious fiction is indeed more syntactically complex than popular fiction. Following the implementation of the log-likelihood test (Dunning 1993)[7] and the application of part-of-speech tagging to one of their corpora (The Chick Lit Corpus, Montoro 2012)[8], the researchers discovered an overuse of subordinating conjunctions in both serious and popular prose fiction, compared to the British National Corpus (BNC Sampler Written). This finding was further corroborated by Montoro (2018) through a similar analysis using a different corpus (Modern Vampire Corpus – MVC, Montoro 2015)[9], which also revealed a similar overuse of subordinating conjunctions. This

[4] Berber Sardinha, T. 2004. *Linguística de corpus.* São Paulo: Manole.
[5] Ramet, C. 1999. *Ken Follett: The transformation of a writer*. Bowling Green, OH: Popular Press.
[6] Moretti, F. 2010. "History of the novel, theory of the novel." *Novel: A Forum on Fiction*, 43 (1), pp. 1–10.
[7] Dunning, T. 1993 "Accurate methods for the statistics of surprise and coincidence", *Computational Linguistics*, 19 (1), pp. 61–74.
[8] Montoro, R. 2012. *Chick lit: The stylistics of cappuccino fiction*. London: Bloomsbury.
[9] Montoro, R. 2015. "Style in popular literature, " in Sotirova, V. (ed.) *The Bloomsbury companion to stylistics*. London: Bloomsbury, pp. 673–689.

evidence, focusing solely on the part-of-speech tag of subordinating conjunctions, does not support the notion that popular fiction is syntactically less complex than serious literature. These results suggest that if syntactic simplicity is a defining characteristic of popular genres, it cannot be attributed to a lack of subordinating conjunctions. 'These results […] also underscore the usefulness of a corpus stylistics methodology to test literary critical claims." (Montoro 2018: 68) This conclusion contrasts with what Moretti (2010) suggested, challenging the notion that the syntactic complexity, at least in terms of subordination, distinguishes serious fiction from its popular counterpart. For the purposes of this thesis, a comparative analysis of the levels of subordination present within the genres of children's literature and popular fiction will be undertaken. It should be noted that the results of this comparison will contribute only a portion of the overall conclusion pertaining to the linguistic complexity of these genres in Harry Potter's books.

Montoro and McIntyre (2019) explore additional keyness findings that not only demonstrate that simplicity is characteristic of popular fiction, but also shed light on the specific level at which this simplicity manifests. Their analysis focuses on the underuse of specific parts of speech, concentrating on three categories that are under-represented: prepositions, adjectives, and the *wh*-determiner (whose), all of which are fundamental elements of the noun phrase.

4.5 Noun Phrase

Montoro and McIntyre (2019) directed their attention to the analysis of one specific grammatical category, "which can account for the simplicity often associated with popular fiction in an insightful manner" (Montoro 2018: 64), the Noun Phrase (NP). Moreover, Montoro noticed a striking contrast between the increasing use of NP in written English of the twentieth century (Leech et al. 2009: 206)[10] and the lack of nouns he observed in his corpora. It could be argued that the limited use of nouns in popular fiction gives it a more informal tone compared to serious fiction, which ultimately results in a greater degree of simplicity. According to Biber and Clark (Biber & Clark 2002: 44)[11], they contend that non-

[10] Leech, G., Hundt, M., Mair, Ch. and Smith, N. 2009. *Change in contemporary English*: *A grammatical study.* Cambridge: Cambridge University Press.
[11] Biber, D., & Clark, V., 2002. *Historical shifts in modification patterns with complex noun phrase structures*.

clausal modification is notably common in academic writing but appears to be nearly non-existent in everyday spoken language and informal conversation. Consequently, one could posit that the fewer nouns are present in the corpus, the simpler and more readable it becomes from a syntactic perspective.

4.5.1 *Of*-preposition

Montoro and McIntyre (2019) emphasize that popular fiction showcases a form of syntactic simplicity, which operates on the clausal level rather than the phrasal level. This particular aspect of syntactic simplicity involves the post-modification of the NP referred to as the '*of*-phrase.' More precisely, they explore whether one of the distinctive characteristics of genre fiction that leads critics and readers to "describe popular literature as easy to read" and which "tends to be equated with a kind of syntactic simplicity" (Montoro 2018: 63) is the reduced occurrence of post-modification through the use of the *of*-phrase.

Montoro (2018) delves into the concept of '*IO tags*' ('of' used as a preposition) as initially proposed by the findings in Montoro and McIntyre (2019). Upon closer examination, these tags are employed to introduce prepositional phrases that post-modify nouns within NP. To investigate the use of the *of*-phrase in more depth, Montoro (2018) utilizes the Modern Vampire Corpus (MVC) and Chick Lit Corpus and subjects both corpora to analysis using the '*AntConc software*' (Anthony, 2014)[12]. This software allows for a more sophisticated discernment of the various functions of the *of*-phrase than previous tools. This approach offers a re-evaluation of the original findings discussed in Montoro and McIntyre (2019).

One aspect is, the Chick Lit novels exhibit reduced usage of both nouns and prepositions, suggesting that noun phrases in Chick Lit may be less frequent and less syntactically complex compared to those in the BNC Written Imaginative corpus, typically associated with more sophisticated writing. However, while the MVC employs fewer nouns, it surprisingly overuses the *IO tag*. This overuse contradicts the norm observed by Biber et al. (2011), who noted that such overrepresentation prevails in academic writing rather than fiction.

[12] Anthony, L., 2014 *AntConc* (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net Accessed 1 Nov. 2023.

4.5.2 Multifunctional *of*-preposition

Examining the concordances of the *IO tag* in both corpora in greater depth uncovered its multifunctionality. In other words, it is used in contexts beyond nominal post-modification. The manual analysis of the *IO* concordances indicates that, apart from nominal post-modification, the *IO tag* appears in various syntactic environments. Some of the context are complements of verbs, adjectives and adverbs. 'Moreover, this manual approach confirms that the *IO* (of) *tag* has the same six functions besides that of nominal post-modification in the Chick Lit Corpus and the MVC respectively. (Montoro 2018: 72). This method established an opportunity to distinguish cases of nominal post-modification and discern them from non-nominal functions, ultimately facilitating a precise and separate quantification of each occurrence.

Based on the newly acquired outcomes, it appears that Montoro's two-stage methodology has significantly polished the original findings. Firstly, by incorporating a second software tool, the specific function of the *IO tag* was narrowed. Secondly, this analysis confirmed not only the statistical underuse of the *IO tag* in the Chick Lit Corpus but also, crucially, highlights the statistical underrepresentation of the post-nominal *of*-phrase in the MVC. This finding challenges the initial interpretation, where all functions of the *IO tag* were considered as one entity.

In conclusion, this study emphasizes that the perceived simplicity in popular fiction can be attributed to both the reduced occurrence of nouns and the diminished frequency of post-modifying prepositional *of*-phrases. Such a statement holds significant value when evaluating the outcomes of this thesis.

4.6 Grammatical structures

This chapter aims to provide a comprehensive understanding of how grammatical structures are employed across different contexts and how they shape the complexity of linguistic expressions, from finite complement clauses preceded by common verbs to appositive noun phrases. Moreover, when assessing the complexity of grammatical structure, it is insufficient to rely only on sentence length or the amount of clausal subordination as indicators.

4.6.1 Single measures

As Biber et al. (2011) suggest, there is a strong connection recognized between the use of subordinate clauses and grammatical complexity. This relationship has become deeply embedded in the fundamental understanding of how complex grammar is constructed and analyzed. Similarly, numerous studies employ the T-unit length (sentence length) as a metric for evaluating linguistic complexity. "The logic here is straightforward: as we add more structures to a simple clause, that clause becomes increasingly complex, and the T-unit length becomes longer" (Ibid. :12). In the same research where the registers of advanced writing and conversation are compared, it is explained why relying on these metrics as the sole indicator of complexity is flawed. In this case, two sample sentences are used for explanation.

1. *"Well, since he got so upset, I just didn't think we would want to wait for Tina to come back.*
2. *This may be part of the reason for the statistical link between schizophrenia and membership in the lower socioeconomic classes."* (Biber et al. 2011 :14)

It is evident upon examination that both sentences are composed of 20 words each. However, *sentence 1* comprises 4 dependent clauses, while *sentence 2* comprises none.

Regarding sentence length - numerous linguistic elements contribute to the length of a T-unit, such as incorporating dependent clauses, integrating embedded phrases, or adding supplementary adjectives and adverbs. These components effectively expand the T-unit, making it more complex in structure. If we were to assess complexity solely based on the length of the T-unit, we might conclude that *sentences 1 and 2* possess an equivalent level of complexity. However, this approach overlooks the fundamentally distinct structures that each sentence may exhibit. Such an analysis based merely on length fails to account for the unique syntactic and grammatical nuances that differentiate one sentence from another, despite their similar lengths.

As to dependent clauses – when applying this metric, *sentence 2* is deemed to be quite simplistic, in stark contrast to *sentence 1*, which is classified as highly complex. However, this assessment encounters two significant issues. Firstly, the English language encompasses a wide array of dependent clauses, each showing various degrees of complexity. This

diversity in dependent clause types implies that not all of them contribute equally to the overall complexity of a sentence. Secondly, the complexity inherent in non-clausal embeddings (phrases and constructions) must also be taken into consideration. “In fact, we argue below that certain types of non-clausal embedding represent higher orders of complexity than dependent clauses” (Ibid. :15). Therefore, a comprehensive evaluation of sentence complexity should account for both the variety of dependent clauses and the intricacy of non-clausal embeddings to accurately reflect the true complexity of a sentence.

4.6.2 Developmental stages

Grammatical structures which are quickly learned and commonly utilized in everyday speech by native speakers inherently lack difficulty. Consequently, these patterns do not exemplify a high level of complexity in linguistic production. Conversely, intricate forms of phrasal embedding are often reserved for the specialized contexts of formal writing. Such discourse styles are not naturally acquired, and a significant number of English native speakers seldom, if ever, engage in this type of language use. Additionally, these stages of linguistic acquisition, when they do occur, they usually take place later in life, often during adulthood. Given these considerations, it seems plausible to suggest these grammatical structures embody a much greater level of production complexity compared to the complexity features found in everyday conversation. Thus, as Biber et al. (2011: 29,30) found in their research, the stages generally progress from finite dependent clauses functioning as constituents in other clauses. This progression continues through intermediate stages of non-finite dependent clauses and phrases acting as components in other clauses, and eventually reaches the final stage, characterized by a dense use of phrasal (non-clausal) dependent structures within noun phrases. Table 1 outlines a detailed hypothesis regarding the potential composition of these developmental stages. As highlighted in their research, the proposed developmental stages require empirical investigation through studies of second language development. “However, this framework has the advantage of being based on the empirical study of actual language production, rather than being posited purely on intuitive or theoretical grounds” (Ibid. 31). Based on the findings of developmental stages, many years later, Biber et al. (2023) found that syntactic differences hold greater significance than structural differences. “… although syntactic distinctions are more important than structural

distinctions, considering both together clearly provides the best account of textual co-occurrence patterns". (Biber et al. 2023: 25)

Table 1: Hypothesized Developmental Stages for Complexity Features (Biber et al. 2011: 30-31)

Stage	Grammatical structure(s)	Example(s)
1	Finite complement clauses (that and WH) controlled by extremely common verbs (e.g., think, know, say)	1a we never quite **know** what to make of him (conv)
		1b just **think** that he didn't pay attention
2	Finite complement clauses controlled by a wider set of verbs	2a I'd **forgotten** that he had just testified on that one (conv)
	Finite adverbial clauses	2b If you're sitting next to me and you want ninety degrees, and I want sixty degrees, we're just gonna be battling each other... (conv)
		2c I'm assuming I gained weight because things are a little tighter than they used to be (conv)
	Non-finite complement clauses, controlled by common verbs (especially want)	2d I don't **want** to fight with them about it (conv)
		2e I **hate** watching the people interact (conv)
	Phrasal embedding in the clause: adverbs as adverbials	2f He's so confused anyway (conv)
	Simple phrasal embedding in the noun phrase: attributive adjectives	2g It certainly has a nice flavor (conv)
		2h Tom Jones is apparently a real name (conv)
3	Phrasal embedding in the clause: prepositional phrases as adverbials	3a He seems to have been hit on the head (fict)
	Finite complement clauses controlled by adjectives	3b It seemed quite **clear** that no one was at home (fict)
		3c I was **sure** that I could smooth over our little misunderstanding (fict)
	Non-finite complement clauses con- trolled by a wider set of verbs	3d The snow began to fall again (fict)
	That relative clauses, especially with animate head nouns	3e ...the guy that made that call (fict)

	Simple phrasal embedding in the noun phrase: nouns as pre-modifiers	3f ...some really obscure cable channel (fict)
	Possessive nouns as pre-modifiers	3g Tobie's voice (fict)
	Of phrases as post-modifiers	3h **editor** of the food section (fict)
	Simple PPs as post-modifiers, especially with prepositions other than of when they have concrete/ locative meanings	3i **house** in the suburbs (fict)
4	Non.finite complement clauses controlled by adjectives	4a These will not be **easy** to obtain (acad)
	Extraposed complement clauses	4b **It is clear** that much remains to be learned... (acad)
		4c In that case **it is useful** to phrase sustainability in terms of... (acad)
	Non-finite relative clauses	4d ...the **method** used here should suffice... (acad)
		4e **Studies** employing electrophysiological measures (acad)
	More phrasal embedding in the NP 5 attributive adjectives, nouns as pre-modifiers	4f The prevalence of airway obstruction and self-reported disease status (acad)
		4g Positive propagule size effects have been demonstrated for both plant and animal systems
	Simple PPs as post-modifiers, especially with prepositions other than of when they have abstract meanings	4h with half of the subjects in each age/ instructional condition receiving each form (acad)
		4i The specific growth rate at small population sizes....
5	Preposition + non-finite complement clause	5a The **idea of** using a Monte Carlo approach (acad)
	Complement clauses controlled by nouns	5b The **hypothesis** that female body weight was more variable (acad)
	Appositive noun phrases	5c The **CTBS** (the fourth edition of the test) was administered in 1997–1998 (acad)
	Extensive phrasal embedding in the NP: multiple prepositional phrases as post-modifiers, with levels of embedding	5d The [presence of layered [[structures] at the [[[borderline]] of cell territories]]] (acad)

5 Material and method

The chapter forms the backbone of the research presented in this thesis. It outlines the materials used and the methodologies employed to conduct the study. Moreover, this section is crucial as it ensures a transparent and replicable framework for the research, which is a key aspect of scientific inquiry.

5.1 Material

The thesis analyzes and compares two texts – the first (HP1) and the last (HP7) volume of the series of books about Harry Potter by J.K. Rowling. In order to carry out the quantitative macro-analysis (Chapter 6), the pdf texts of the books were converted to plain text format (.txt) using the programme '*AntFileConverter*'[13].

5.2 Methods of macro-analysis

5.2.1 Corpus analysis software

For the macro analysis of both corpora, the '*CQPweb*'[14] tool was used. It is a web-based system which is "compatible with any corpus, but is especially useful for large corpora, corpora with word-level annotation (such as part-of-speech tagging), and corpora with rich text-level metadata"[15].

5.2.2 Installing both corpora

After creating a new account in the web software, the two corpora, HP1 and HP7, were installed. The *CQPweb* platform is capable of handling corpora with up to 2,000,000 tokens, which comfortably exceeds the size requirements for our datasets. In addition, the corpora undergo morphological tagging via '*TreeTagger*,' a tool which prepares them for in-depth linguistic examination. This includes the application of part-of-speech (POS) tagging and lemmatization. *TreeTagger* systematically assigns POS labels (such as noun, verb, adjective, etc.) to each word in the texts. Lemmatization involves consolidating various inflected forms

[13] Anthony, L., 2022. AntFileConverter (Version 2.0.2) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software Accessed 20 Nov. 2023.
[14] Hardie, A, 2012. CQPweb - combining power, flexibility and usability in a corpus analysis tool. International Journal of Corpus Linguistics 17 (3): 380–409. https://cqpweb.lancs.ac.uk Accessed 20 Nov. 2023.
[15] https://www.research.lancs.ac.uk/portal/en/upmprojects/cqpweb(855e22e7-97fb-4863-9101-9b245729ccbd).html Accessed 20 Nov. 2023.

of a word into one group for analysis as a single entity, based on the word's lemma, or its base form as found in a dictionary.[16] Consequently, this approach of tagging and lemmatizing the corpora provides a more structured and standardized dataset, which is crucial for the objectives of this thesis.

5.2.3 Number of tokens, types and TTR

Upon uploading the corpora, information regarding the corpus metadata can be accessed under the section titled “corpus info”. To view the metadata for both corpora, it is necessary to switch between the uploaded datasets, as the system is designed to display the metadata of only one corpus at a time.

5.2.4 Number of chapters and sentences

Since the corpora are uploaded as text documents without details of pages or chapters, it is necessary to count the chapters in each book manually. Regarding the number of sentences, a standard query needs to be created. Deconstructing the query for counting the number of sentences is necessary because, at first glance, the query appears more complex than it is. By breaking it down, one can gain a clearer understanding of how it functions, making it easier to comprehend and potentially modify.

1. *<s>*: This is a tag used to denote the start of a sentence in many corpus processing tools.
2. *word=”...”*: This is the structure used to query specific words in the corpus. It is saying, "Look for occurrences where the word matches the following criteria."
3. .+: Represents a regular expression that matches any character (represented by ‘.’) one or more times (represented by ‘+’). So,’.+’ will match any word of at least one character in length. This is a way to ensure that the query counts all words, regardless of what they are.

When combined, *‘<s>[word=".+"]’* is essentially a query that says, "For each sentence in the corpus, find all instances where there is at least one word." Since this criterion will be

[16] Cambridge Dictionary, https://www.dictionary.cambridge.org/dictionary/english/lemmatization Accessed 1 Dec. 2023.

met by every sentence, the query will effectively count the number of sentences in the corpus. The last step prior to running the query is setting the query mode to '*CQP syntax*'.

5.2.5 Keywords sorted according to word forms tag

The next step was to compare the target corpus (HP1) with the reference corpus (HP7) to generate a list of keywords. The tool was essentially identifying words that are characteristic of HP1 in comparison to HP7. This process involved a detailed statistical analysis where the frequency of words in both corpora were compared. Firstly, frequency list 1 (target corpus) and frequency list 2 (reference corpus) needed to be chosen. The '*log-likelihood*' filter was set, as it helps to identify words that are significantly more or less common in one corpus compared to the other. Lastly, the minimum frequency of a token was set to 1.

As a result, the software created a list of keywords, which included both positive and negative keywords. Positive keywords are those words that are significantly more frequent in the target corpus HP1 than in the reference corpus HP7. These words are essential in the sense that they highlight specific themes, topics, or linguistic styles that are more prominent in HP1. On the other hand, negative keywords are words that are significantly less frequent in HP1 compared to HP7. These words point out the themes, topics or characters that are underrepresented or less emphasized in HP1. To serve the objectives of this thesis, only the first 20 keywords were included and analyzed.

The list of unique words was created by the same steps that were undertaken to create a list of keywords, as the *CQPweb* offers in the same section also the option of creating a list of unique words while comparing HP1 to HP7 and vice versa.

5.2.6 Table 5: Frequency of *death*

The list of the 20 keywords that is enclosed in this thesis shows only the frequency of the word *death* per million words. However, the *CQPweb* tool also shows absolute frequency which shows the total number of appearances in the given corpora. The concordance lines of the search term *death* were sorted according to the right context. The position '-1 Right' was set and specified to show words starting with *Eater*, so that the phrase *Death-Eater* is included in its singular and plural form. After manual examination of the outcome, this number was subtracted from the total count.

$$\text{Frequency per Million} = \left(\frac{Total\ number\ of tokens}{Frequency\ of\ 'death'}\right) \times\ 1{,}000{,}000$$

Plugging in the numbers:

$$\text{Frequency per Million} = \left(\frac{105}{253{,}303} \times\ 1{,}000{,}000\right)\ \approx 414.52$$

The frequency per million was calculated by dividing the frequency of the word 'death' by the total number of tokens, and then multiplying the result by 1,000,000. This standardized the frequency to a rate per million tokens, allowing for easier comparison across texts of different lengths.

5.2.7 Sentence length

This calculation provides a measure of how dense or complex the text is on average, with higher numbers indicating longer sentences on average. To calculate the average sentence length given the number of tokens and the number of sentences in each corpus, the following formula was used:

$$\text{Average sentence length} = \left(\frac{Total\ number\ of\ tokens}{Total\ number\ of\ sentences}\right)$$

Plugging in the numbers:

$$\text{Average sentence length for HP1} = \left(\frac{99{,}538}{7{,}282}\right) \approx 13.7$$

$$\text{Average sentence length for HP7} = \left(\frac{253{,}303}{17{,}546}\right) \approx 14.4$$

The methods used to obtain the total count of tokens and sentences are detailed in earlier chapters, specifically in sections 5.1.3 and 5.1.4.

5.2.8 Keywords sorted according to part-of-speech tag

The process of generating a keyword list based on parts of speech differs from that based on word forms only in the selection of the part-of-speech tag as the basis of comparison between the two corpora. This leads to creating a keyword list sorted by the frequency of various parts of speech, such as common nouns, adjectives, adverbs, etc. Rather than labelling with

terms like *'common noun'*, the list employs *'English Penn TreeBank P.O.S. tags'*[17], corresponding here to '*NN*'.

5.3 Methods of micro-analysis

For the detailed micro-analysis, 50 sentences were chosen from the books – specifically, the first 25 sentences of Chapter 3 in HP1 and the first 25 sentences of Chapter 3 in HP7. Each of these sentences was subjected to an in-depth examination, where they were broken down into various categories (see a list below, which includes examples). Additionally, the analysis involved counting the number of words and clauses, as well as the tally of coordinated main clauses. The corresponding tables for each text sample are available in Appendix 3 and Appendix 4. These tables served as the basis for creating Tables 10-12 and Figures 1-2.

The applied categories are namely:

A. Finite clauses functioning as noun modifiers (ex. 1)
B. Non-finite clauses functioning as noun modifiers (ex. 2)
C. Finite clauses functioning as other phrase complements/modifiers (ex. 3)
D. Non-finite clauses functioning as other phrase complements/modifiers (ex. 4)
E. Finite clauses functioning as clause-level constituents (ex. 5)
F. Non-finite and verbless clauses functioning as clause-level constituents (ex. 6)

Examples of each category from text used in the micro-analysis:

1. *there was no escaping Dudley's gang,* ***who visited the house every single day***
2. *pausing to add the piece of broken mirror to the rucksack* ***filled with things*** *he would be taking with him*
3. *Harry was glad* ***school was over***
4. *The rest of them were all quite happy* ***to join in Dudley's favourite sport: Harry Hunting.***
5. ***When he reached the living room*** *he found all three Dursleys.*
6. *Harry strolled downstairs,* ***his hands deep in his jeans pockets.***

[17] https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html Accessed 28 Nov. 2023.

To determine the count of direct speech segments, *CQPweb* software was used. A query was entered in both corpora to generate the number of final quotation marks. For more information regarding query inputs, refer to Chapter 5.2.4.

6 Macro-analysis

This chapter focuses on a quantitative analysis of the first (HP1) and last (HP7) volumes of the Harry Potter series. The analysis is divided into two main categories: lexical and grammatical complexity.

The lexical analysis encompasses various metrics such as the total number of word forms (tokens), chapters and sentences, standardized type/token ratio (TTR), and provides a list of keywords, including word forms and unique words. The grammatical analysis emphasizes the study of sentence length and part-of-speech patterns.

6.1 Analysis of lexical complexity

6.1.1 Number of tokens, types and TTR

Table 2: General data regarding lexical complexity

	HP1	HP7
Number of tokens	99,538	253,303
Number of types	5,928	11,348
Standardized TTR (1,000-token basis)	0.3665	0.3706
Non-standardized TTR	0.0596	0.0448

The data shows a marked increase in both the total number of tokens (2,5 times) and types (1,9 times) from HP1 to HP7. This suggests an increase in lexical diversity and complexity in the later book. The standardized TTR, which is a measure of lexical diversity adjusted for text length, is slightly higher in HP7 than in HP1. This indicates a greater variety of vocabulary in HP7, reducing the impact of the length of the volume. The non-standardized TTR, which is the ratio of unique words to total words, is lower in HP7. This is expected as longer texts tend to repeat words more, thus lowering this ratio. Therefore, because of the aim of this thesis, it is definitely more reasonable to follow the indicator of standardized type/token ratio, which is calculated on the basis of 1,000-token segments of the text.

6.1.2 Number of chapters and sentences

Table 3: Number of chapters and sentences

	HP1	HP7
Number of chapters	17	36 + epilogue
Number of sentences	7282	17546

The last volume is more extensive in terms of both chapters (2,2 times) and sentences (2,4 times) compared to the first volume. These observations highlight the growth in narrative complexity and depth as the Harry Potter series progresses from its beginning to its conclusion.

6.1.3 Keywords sorted according to word forms tag

Table 4: List of top 20 key word forms, HP1 compared to HP7 (punctuation excluded, frequency per million words - pmw); the word forms highlighted in green are typical of HP7, those without highlighting of HP1

Number	Word	HP1 frequency (pmw)	HP7 frequency (pmw)	Log-likelihood
1	Hagrid	3,717.17	679.03	373.79
2	Quirrell	1,125.20	3.95	272.77
3	yeh	1,185.48	71.06	204.37
4	n't	8,408.85	4,437.37	188.23
5	Dudley	1,406.50	193.44	170.6
6	Professor	1,778.22	367.15	162.04
7	'd	2,712.53	884.32	151.54
8	Uncle	1,225.66	189.50	138.32
9	ter	783.62	43.43	138.17
10	death	110.51	1,223.83	137.58
11	her	1,798.31	4,236.03	137.28
12	wand	622.88	2,313.43	136.69
13	Voldemort	381.76	1,756.79	125.13
14	yer	602.78	15.79	124.61
15	Vernon	1,165.38	197.39	123.66
16	Not	2,350.86	4,721.62	110.87

17	upon	20.09	746.14	108.17
18	very	1,727.98	529.01	104.78
19	points	442.04	3.95	102.45
20	Filch	542.51	27.63	97.86

As shown by the data, a substantial part of the most frequent words in HP1 consist of names (*Hagrid, Professor Quirrell, Dudley, Uncle Vernon, Filch)*, as the story is more centred around Harry's introduction to the wizarding world and characters. What is also prominent are the dialect word forms, such as *yeh, ter* and *yer* (examples 1-3), which are typical of Hagrid, who is a more central character in HP1, significantly decreasing in prominence by HP7.

1. *"Got summat fer **yeh** here — I mighta sat on it at some point, but it'll taste all right."* (HP1)
2. *"Gotta get up **ter** town, get all **yer** books an' that."* (HP1)
3. *"Like I said, **yeh**'d be mad ter try an' rob it," said Hagrid.* (HP1)

At first sight, based on a mere comparison of data from the list above, the frequent use of *very* in HP1 might reflect a tendency to avoid less common adverbial intensifiers in children's books. However, upon further scrutiny, there appears to be another type of difference between the volumes. In HP1, the intensifier *very* is used more frequently in the narrative parts, most often modifying the adjectives and adverbs *good, well, important, last, difficult, fast, hard, long, angry, cold, first, old* and *strange* (examples 7 and 9). In HP7, the most frequent right collocates of *very* are similar *well, much, good, long, clever, close, end, least, little, nice, odd, same, sorry, valuable*. They are, however, more often parts of direct speech in HP7 (examples 8 and 10).

7. *The Potters knew **very** well what he and Petunia thought about them …* (HP1)
8. *"You have done **very** well, Severus. How long do you think I have?" Dumbledore's tone was conversational …* (HP7)
9. *He was in a **very** good mood.* (HP1)
10. *… said Scrimgeour, "but a Snitch would be a **very** good hiding place for a small object …"* (HP7)

In a similar vein, the increased frequency of *wand* in HP7 reflects the heightened magical activity and conflicts. In HP7, *wand* frequently co-occurs with dynamic verbs, e.g. *draw, thrust, withdraw, pull out, point, give, raise, seize* and *wave* (examples 13 and 14). Moreover, *wands* are rarely modified by adjectives in HP7. In HP1, on the contrary, dynamic actions involving a *wand* are rarer (and occur towards the end of the volume), and the noun *wand*, denoting a new object that is introduced to the reader as a part of the wizarding world, is pre-modified by a descriptive adjective more often, e.g. *magic, mahogany, single, best, powerful* and *battered-looking* (examples 11 and 12).

Several examples of this observation include:

11. *"... only place fer wands, Ollivanders, and yeh gotta have the **best** wand."*
12. *He rummaged around in his trunk and pulled out a very **battered-looking** wand.*
13. *In desperation Harry **pointed** his wand at the sidecar and shouted, "Wingardium Leviosa !"* (HP7)
14. *Hermione **waved** her wand to ignite the old gas lamps …* (HP7)

Table 5: Frequency of the noun *death*

	HP1 total frequency	HP1 frequency (pmw)	HP7 total frequency	HP7 frequency (pmw)
death	11	110.51	310	1,223.83
Death, excluding the phrase Death Eater	11	110.51	105	414.52

From the data presented in the table, the much higher occurrence of *death* in HP7 might indicate the darker, more serious tone of the final book. Upon closer examination, this word is used in the phrase *Death-Eater* 205 times out of the 310 occurrences. Even if only the remaining 105 occurrences were considered (as shown in Table 5), it can still be claimed that the word *death* is significantly more frequently used in HP7 than in HP1, and therefore it adds to the rising ominous tone of the final volume (examples 15-17).

15. *"Her nine-hundred-page book was completed a mere four weeks after Dumbledore 's mysterious **death** in June. I ask her how she managed this superfast feat."* (HP7)

16. *"Forgive me, but it was a very unusual move then, under imminent threat of **death**."* (HP7)
17. *"Kill me, then, Voldemort, I welcome **death**! But my **death** will not bring you what you seek. There is so much you do not understand."* (HP7)

6.1.4 Unique words

Unique words, in the context of comparing two different texts (or corpora), refer to word forms that occur exclusively in one text but not in the other.

Table 6: List of 20 unique words in HP1

Number	Word	Total frequency	Frequency (pmw)
1	gotten	20	200.93
2	Mom	19	190.88
3	Nimbus	16	160.74
4	Piers	15	150.70
5	mail	14	140.65
6	Flint	14	140.65
7	turban	14	140.65
8	Scabbers	14	140.65
9	Bludgers	13	130.60
10	Hooch	12	120.56
11	Keeper	10	100.46
12	jus'	9	90.42
13	Keys	9	90.42
14	exams	9	90.42
15	Halloween	9	90.42
16	Figg	8	80.37
17	Gotta	8	80.37
18	Chasers	8	80.37
19	chessmen	8	80.37
20	Remembrall	8	80.37

It is notable that the word *gotten* ranks first among the unique words not found in HP7. It might reflect a different linguistic choice or narrative style that J. K. Rowling later adjusted or moved away from. *Gotten* is more commonly used in American English than in British English.[18] Its usage in HP1 might have been part of an effort to appeal to a broader, international audience, including American readers.[19] Over time, as the series gained global popularity, and the target readers were expected to be older, such specific lexical choices might have been deemed less necessary. It must be emphasized that this assumption is made based on a mere comparison of data from the list above. Nonetheless, the rest of these words are specific to events, objects, or characters that are unique to the first book. To illustrate, *Nimbus* refers to a broomstick Harry receives in HP1 (example 18). *Piers* refers to Dudley's friend, who appears in the early chapters (example 19). *Scabbers* is the name of Ron's pet rat, prominent in the first book (example 20). Similarly, *turban* and *Halloween* are typical of HP1 (examples 21 and 22).

18. *"It contains your new **Nimbus** Two Thousand, but I don't want everybody knowing you've got a broomstick or they'll all want one."* (HP1)
19. *The Dursleys bought Dudley and **Piers** large chocolate ice creams at the entrance and then, because the smiling lady in the van had asked Harry what he wanted before they could hurry him away, they bought him a cheap lemon ice pop.* (HP1)
20. *"Percy got an owl from my dad for being made a prefect, but they couldn't aff— I mean, I got **Scabbers** instead."*
21. *Harry breathed in the funny smell that seemed to come from Quirrell's **turban**.* (HP1)
22. *"Snape's only got to say he doesn't know how the troll got in at **Halloween** and that he was nowhere near the third floor — who do you think they'll believe, him or us?"* (HP1)

[18] *Longman Dictionary of Contemporary English*, https://www.ldoceonline.com Accessed 24 Nov. 2023.
[19] Several distinctions are present between the Bloomsbury (the UK version) and Scholastic versions (the US version used for the analysis) of the HP series. Many of these alterations were implemented by the US editors to make the text more understandable for American audiences. For specific information, see https://www.hp-lexicon.org/differences-changes-text/. Accessed 1 Dec. 2023

Table 7: List of 20 unique words in HP7

Number	Word	Total frequency	Frequency (pmw)
1	Lupin	167	659.29
2	Eaters	142	560.59
3	Luna	140	552.70
4	Kreacher	140	552.70
5	Fleur	101	398.73
6	Bellatrix	98	386.89
7	Horcrux	93	367.15
8	Xenophilius	91	359.25
9	locket	85	335.57
10	Tonks	79	311.88
11	Aberforth	74	292.14
12	Greyback	72	284.24
13	elder	67	264.51
14	Hallows	67	264.51
15	Scrimgeour	65	256.61
16	Eater	63	248.71
17	Elf	59	232.92
18	Yaxley	55	217.13
19	Muriel	53	209.24
20	Ariana	52	205.29

All of these words represent specific objects or characters that are unique to the last book. The only words connected with dark magic, gloomy atmosphere and antagonists are: *Eater* and *Eaters* (see table 5), *Bellatrix*, *Horcrux*, *locket*, *Greyback*, *Scrimgeour* and *Yaxley* (examples 23-25).

23. *From above came **Bellatrix**'s voice. "You are lying, filthy Mudblood, and I know it!"* (HP7)

24. *"Can you imagine what he's going to do once he realizes the ring and the **locket** are gone? What if he moves the Hogwarts Horcrux, decides it isn't safe enough?"* (HP7)
25. *"NO!" shrieked Hermione, and with a deafening blast from her wand, Fenrir **Greyback** was thrown backward from the feebly stirring body of Lavender Brown.* (HP7)

6.2 Analysis of grammatical complexity

6.2.1 Sentence length

Table 8: Sentence length

	HP1	**HP7**
Number of tokens	99,538	253,303
Number of sentences	7,282	17,546
Average sentence length (tokens per sentence)	13.7	14.4

The data reveals that the average sentence length in HP7, measured by the number of words per sentence, is indeed higher in comparison with the figures for HP1, which increases the processing difficulty.

6.2.2 Keywords sorted according to part-of-speech tag

Table 9: List of 10 part-of-speech, HP1 compared to HP7 (frequency per million words)

Number	**Part-of-speech**	**HP1 frequency**	**HP7 frequency**	**Log-likelihood**
1	CD (Cardinal number)	6,138.36	3,766.24	85.43
2	,	56,902.89	63,769.48	59
3	NPS (Proper noun, plural)	703.25	181.60	51.87
4	IN (Preposition or subordinating conjunction)	76,453.21	82,655.95	37.36
5	VBN (Verb, past participle)	19,650.79	22,964.59	37.26

6	NNS (Common noun, plural)	27,959.17	24,350.28	36.93
7	PP$ (Possessive pronoun)	16,646.91	19,316.79	28.57
8	CC (Coordinating conjunction)	25,578.17	28,483.67	22.78
9	WP$ (Possessive wh-pronoun)	30.14	221.08	21
10	NN (Common noun, singular or mass	100,373.73	95,308.78	20.84

Upon examining and contrasting the information provided in the list, it is evident that HP1 exhibits a greater occurrence of cardinal numbers and proper nouns, a characteristic commonly associated with simplicity typical of its genre. Likewise, HP7 demonstrates an increased use of commas, prepositions, subordinating conjunctions, and possessive wh-pronouns, indicative of more complex sentence and phrase structures. The syntactic complexity marked by the use of conjunctions (the initial subordinate adverbial clause, coordinator *and* linking noun phrases), commas (a conjunct, coordinated adjectives), and prepositions (*with slits for nostrils*) can be illustrated by example 26 below. The clause also comprises a complex noun phrase with multiple pre-modification, and a post-modifying clause. In example 27, the commas are used to separate the clauses (including a participial one *disliking the disturbance*), accompanying the conjunctions. At the same time, the coordinator *and* is used to link phrases.

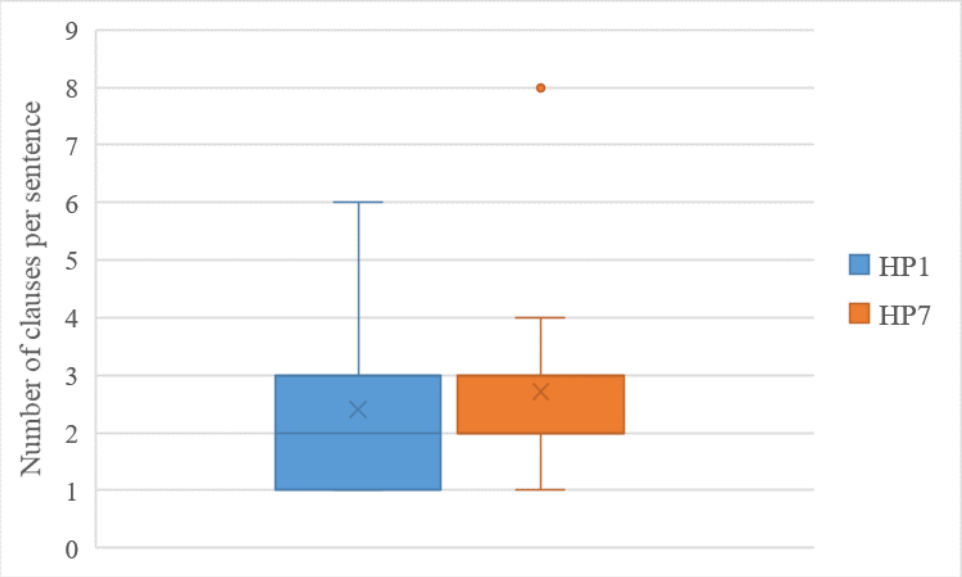
26. ***As*** *they drew nearer****,*** *however****,*** *his face shone through the gloom****,*** *hairless****,*** *snakelike****,*** ***with*** *slits* ***for*** *nostrils* ***and*** *gleaming red eyes whose pupils were vertical*. (HP7)
27. *The great snake****,*** *disliking the disturbance****,*** *opened its mouth wide* ***and*** *hissed angrily****,*** ***but*** *the Death Eaters did not hear it****,*** *so jubilant were they at Bellatrix* ***and*** *the Malfoys' humiliation*. (HP7)

7 Micro-analysis

This chapter focuses on an in-depth micro-analysis of 50 sentences from the HP1 and HP7. The selected sentences were analyzed using Biber et al.'s (2023) classification of syntactic functions and realization forms. Only the functions and types of clauses, and relations among them were considered. Moreover, apart from the mentioned classification, the analysis examines the number of words per sentence, the number of clauses, and the number of coordinated main clauses. Where this analysis differs from Biber et al.'s (2023) is not only in the concept of direct speech (integrated more loosely in the sentence), which is treated separately here (rather than as a finite clause constituent), but also in examining the modification of phrases, as there is only modification by clauses (finite and non-finite) considered.

7.1 Number of clauses per sentence

Figure 1: Number of clauses per sentence



As shown in Figure 1 and Table 10, the sentences in HP1 are composed of 1 – 6 clauses, while in HP7, the maximum number of clauses per sentence is higher, eight (example 28). Simple sentences are the prevalent type of sentence in HP1 (9 sentences, example 29); in HP7, simple sentences are rare (2 sentences).

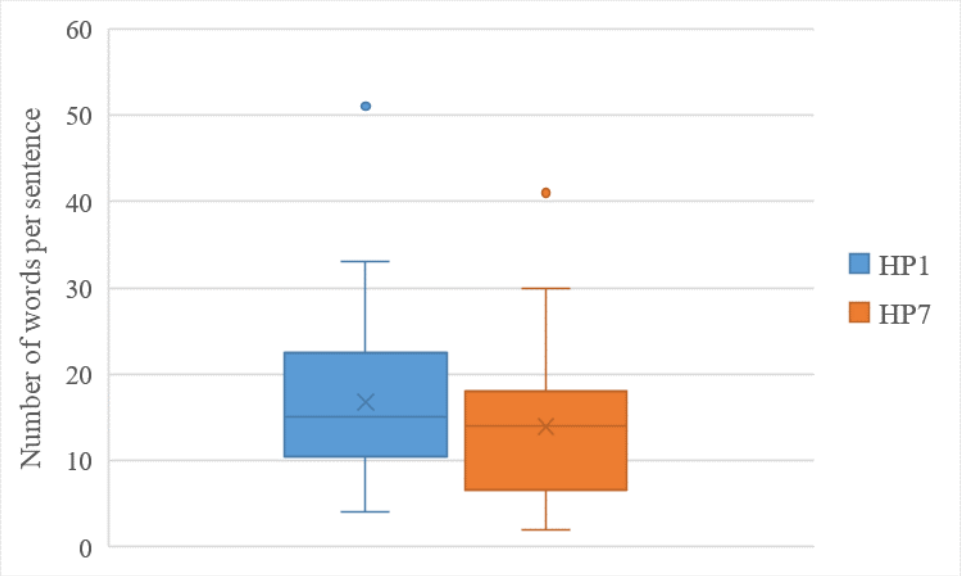
Table 10: Number of clauses per sentence

Number of clauses per sentence	HP1	HP7	total
1	9	2	11
2	5	12	17
3	6	6	12
4	3	4	7
5	1		1
6	1		1
8		1	1
Total	**25**	**25**	**50**

28. *It was not until his uncle bellowed, "BOY!" that Harry got slowly to his feet and headed for the bedroom door, pausing to add the piece of broken mirror to the rucksack filled with things he would be taking with him.* (HP7)
29. *The escape of the Brazilian boa constrictor earned Harry his longest-ever punishment.* (HP1)

7.2 Length of sentences

Figure 2: Length of sentences (in the number of words)



The length of sentences calculated in the number of words ranges from 4 to 51 in HP1, and from 2 to 41 in HP7 (see Figure 2). The shortest and the longest sentences are given in examples (30) and (31), respectively. Even though the sentences in HP1 comprise fewer clauses (2.4 on average) than in HP7 (2.8 on average), they are longer in HP1 (16.8 words) than in HP7 (13.92 words). As illustrated by example (31)[20], one of the factors that contribute to the length of sentences is the coordination of main clauses, which occurred in 28% of sentences in HP1 and 24% of sentences in HP7.

30. *Harry sat.* (HP7)
31. *By the time he was allowed out of his cupboard again, the summer holidays had started* ***and*** *Dudley had already broken his new video camera, crashed his remote control airplane,* ***and****, first time out on his racing bike, knocked down old Mrs. Figg as she crossed Privet Drive on her crutches.* (HP1)

7.3 Dependent clauses

Table 11: Representation of types of dependent clauses (percentage of sentences which contain at least one clause of the particular type)

Syntactic function	Realization form	HP1	HP7
Noun phrase modifier	finite	16.0%	12.0%
	non-finite	4.0%	12.0%
Other phrasal modifiers	finite	8.0%	0.0%
	non-finite	4.0%	0.0%
Clause constituent	finite	36.0%	28.0%
	non-finite	12.0%	32.0%
Direct speech		16.0%	40.0%

The two books differ most in the representation of non-finite and verbless clauses which constitute clause elements. 32% of sentences in HP7 comprise at least one clause element realized by a non-finite or verbless clause; 8% of sentences in HP7 comprise more than one (example 32, illustrating an infinitival adverbial clause of purpose, and a participial adverbial clause). In HP1, clause elements realized by non-finite or verbless clauses occur merely in

[20] In example (31), the predicate verbs *had started, had broken, crashed, knocked down* were analyzed as belonging each to a separate clause, with the ellipsis of the subject in the third and fourth clause (cf. Dušková et al. (2012), chapter 16.1)

12% of sentences. Non-finite clauses are also more frequent in post-modifying function within noun phrases in HP7 (example 33, with post-modification expressed by a prepositional phrase with a complement realized by a gerund clause): 12% of sentences comprise a noun phrase with a non-finite post-modifier (as compared to 4% in HP1). It seems, therefore, that what contributes most to the increased syntactic complexity of HP7 is the use of non-finite and verbless clauses, whose interpretation is difficult for the reader due to the absence of a finite verb and – in most cases – the subject.

32. *One day in July, Aunt Petunia took Dudley to London* ***to buy his Smeltings uniform, leaving Harry at Mrs. Figg's.*** (HP7)
33. *Sixteen years* ***of being addressed thus*** *left Harry in no doubt whom his uncle was calling; nevertheless, he did not immediately respond.* (HP7)

Table 12: Dependent clauses according to the type in numbers

Syntactic function	Realization form	HP1	HP7
Noun phrase modifier	finite	4	3
	non-finite	1	3
Other phrasal modifiers	finite	2	0
	non-finite	1	0
Clause constituent	finite	11	10
	non-finite	5	11
Direct speech		5	11

The sentences in HP1 generally rely more on finite clauses, both in post-modification in noun phrases (a non-restrictive adjectival relative clause in example 34), and as realisation forms of clause constituents (example 35, with a finite adverbial clause and a nominal content clause). Direct speech was more frequent in the sentences in HP7. This also applies to the whole texts. While in HP1 the frequency of direct speech is 24,322.4 instances per million words, in HP7 the frequency is 25,368.8 per million words.

34. *Harry was glad school was over, but there was no escaping Dudley's gang,* ***who visited the house every single day****.* (HP1)
35. ***As he looked at Dudley in his new knickerbockers****, Uncle Vernon said gruffly* ***that it was the proudest moment of his life****.* (HP7)

8 Conclusions

Based on the comprehensive analysis presented in this thesis, the central research question—whether J.K. Rowling adjusts the linguistic complexity of her Harry Potter series to match the age of the intended readers—can be addressed with significant insights. The combination of macro- and micro-analyses provided a multifaceted understanding of this adjustment in linguistic complexity, particularly between "Harry Potter and the Philosopher's Stone" (HP1) and "Harry Potter and the Deathly Hallows" (HP7).

The macro analysis revealed a substantial increase in lexical diversity in HP7, with a marked rise in the number of tokens (2.5 times) and types (1.9 times), compared to HP1. The standardized TTR, with HP1 at 0.3665 and HP7 at 0.3706, further indicates a more varied vocabulary in HP7, despite its longer length. Additionally, the more than double number of chapters and sentences in HP7 points to a more elaborate narrative structure. These findings suggest an escalation in complexity.

The analysis of word usage, such as the frequent employment of *very* and *wand* in varying contexts, reflects a shift in style and thematic focus. In HP1, *very* is more frequently used, mostly to modify adjectives and adverbs like *good, well, important, last, difficult, fast, hard, long, angry, cold, first, old, and strange* in the narrative parts of the text. In HP7, *very* is frequently paired with similar words such as *well, much, good, long, clever, close, end, least, little, nice, odd, same, sorry, valuable*. However, in HP7, these instances are more typically found in direct speech. Similarly, the more frequent use of the word *wand* in HP7 highlights the increased magical action and conflict. In this book, *wand* often appears alongside action-oriented verbs like *draw, thrust, withdraw, pull out, point, give, raise, seize* and *wave*. Additionally, in HP7, wand is seldom modified by adjectives. In contrast, HP1 features fewer dynamic wand-related actions (mainly towards the end) and *wand* is more frequently described with adjectives such as *magic, mahogany, single, best, powerful* and *battered-looking*. The word *death* appears more in HP7, contributing to its darker tone. Notably, *gotten* is unique to HP1, perhaps initially used to facilitate reading for American readers. The keyword lists for both volumes, categorized by word forms, do not indicate plausible signs of complexity. These words are uniquely associated with specific events, objects, or characters distinctive to each book. As for HP1, the words are, for instance, *Nimbus, Piers,*

Scabbers, mail, turban, Halloween, chessmen etc. As far as HP7 is concerned, only 6 out of 20 keywords are related to the gloomy atmosphere, antagonists and dark magic – namely *Bellatrix, Horcrux, locket, Greyback, Scrimgeour and Yaxley*.

HP7's sentence structure, characterized by longer sentences and higher use of commas, prepositions, subordinating conjunctions, and possessive *wh*-pronouns (typically post-modifiers contributing to the complexity of noun phrases), suggests increased processing effort and thus lowered readability. Conversely, HP1 is marked by more cardinal numbers and proper nouns, indicative of children's literature.

The micro-analysis, however, presents a more nuanced picture and reveals distinct types of complexity within each book. While HP1 has fewer clauses per sentence on average (2.4) compared to HP7 (2.8), its sentences are longer (16.8 words in HP1 vs. 13.92 words in HP7). The difference in sentence length, however, appears to be due to the selection of the sample since the macro-analysis has shown that sentences are generally longer in HP7. A significant aspect of complexity in HP7 is the frequent use of non-finite and verbless clauses as clause elements. These non-finite clauses occur in 32% of sentences, with 8% containing more than one such clause. This contrasts sharply with HP1, where only 12% of sentences include non-finite or verbless clause elements. Furthermore, non-finite clauses are more commonly found within post-modifying functions in noun phrases in HP7. The absence of a finite verb in these clauses presents a higher level of interpretative challenge, thereby contributing to the syntactic complexity of HP7. Direct speech, another factor contributing to complexity, appears more frequently in HP7. While the occurrence of direct speech in HP1 stands at 24,322.4 instances per million words, this rate rises to 25,368.8 instances per million words in HP7. This increased use of direct speech potentially adds to the narrative's depth and complexity, offering varied perspectives and enhancing reader engagement.

In summary, while the macro analysis clearly indicates that HP7 is linguistically more complex than HP1, reflecting Rowling's adaptation to her audience's maturing abilities, the micro-analysis provides a more balanced view. It reveals that the increased complexity in HP7 is not uniformly distributed across all aspects of language. HP1, with its longer but structurally simpler sentences, contrasts with HP7's shorter sentences, which are syntactically more intricate. The findings illustrate Rowling's approach to narrative

complexity, adapting her writing style in a way that gradually escalates the linguistic challenge to match her audience's developing reading skills.

References, Sources and Tools

References

Biber, D., et al. 2011 “Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?” *TESOL Quarterly*, vol. 45, no. 1, pp. 5–35.

Biber, D., et al. 2023. “The linguistic organization of grammatical text complexity: Comparing the empirical adequacy of theory-based models.” *Corpus Linguistics and Linguistic Theory*, vol. 0, no. 0, 2023

Bulté, B. and Housen, A. 2012. “Defining and operationalising L2 complexity.” In Housen, A. Kuiken, F. and Vedder, I. (eds) *Dimensions of L2 Performance and Proficiency*. Amsterdam: John Benjamins. pp. 21-46.

Durão, de A.B., et al. 2018. “Children’s Literature Parallel Corpora: A Hybrid Experimental Model to Evaluate Transfers of Language Complexity via Linguistic Transcoding.“ *Ilha Do Desterro: A Journal of English Language, Literatures in English and Cultural Studies*, vol. 71, no. 1, Jan. 2018, pp. 27–52.

Dušková, L., et al. 2012. *Mluvnice současné angličtiny na pozadí češtiny.* Praha: Academia.

Haueng, M. 2019. *Development of Linguistic Complexity in Harry Potter*. Norwegian University of Science and Technology.

Montoro, R. 2018. “Investigating Simplicity in Popular Fiction. A corpus Stylistic Approach.” In Page, R., Busse, B. and Nørgaard, N. (eds) *Rethinking Language, Text and Context: Interdisciplinary Research in Stylistics in Honour of Michael Toolan.* London and New York: Routledge. pp. 63-78.

Montoro, R. and McIntyre, D. 2019. “Subordination as a potential marker of complexity in serious and popular fiction: a corpus stylistic approach to the testing of literary critical claims”. Corpora, 14(3), pp. 275-299.

Puurtinen, T. 1998. “Syntax, Readability and Ideology in Children's Literature.” Meta, 43(4), pp. 524-533.

Sources

Rowling, J. K. 1998. *Harry Potter and the Sorcerer's Stone.* Scholastic Press

Rowling, J. K. 2007. *Harry Potter and the Deathly Hallows*. Scholastic Press

Tools

Anthony, L., 2022. AntFileConverter (Version 2.0.2) [Computer Software]. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software Accessed 20 Nov. 2023.

Hardie, A, 2012. CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17 (3): 380–409. https://cqpweb.lancs.ac.uk Accessed 20 Nov. 2023.

Appendices

Appendix 1

HP1 – Chapter 3, The Letters from No One (initial 25 sentences)

The escape of the Brazilian boa constrictor earned Harry his longest-ever punishment. By the time he was allowed out of his cupboard again, the summer holidays had started and Dudley had already broken his new video camera, crashed his remote control airplane, and, first time out on his racing bike, knocked down old Mrs. Figg as she crossed Privet Drive on her crutches. Harry was glad school was over, but there was no escaping Dudley's gang, who visited the house every single day. Piers, Dennis, Malcolm, and Gordon were all big and stupid, but as Dudley was the biggest and stupidest of the lot, he was the leader. The rest of them were all quite happy to join in Dudley's favorite sport: Harry Hunting. This was why Harry spent as much time as possible out of the house, wandering around and thinking about the end of the holidays, where he could see a tiny ray of hope. When September came he would be going off to secondary school and, for the first time in his life, he wouldn't be with Dudley. Dudley had been accepted at Uncle Vernon's old private school, Smeltings. Piers Polkiss was going there too. Harry, on the other hand, was going to Stonewall High, the local public school. Dudley thought this was very funny. "They stuff people's heads down the toilet the first day at Stonewall," he told Harry. "Want to come upstairs and practice?" "No, thanks," said Harry. "The poor toilet's never had anything as horrible as your head down it — it might be sick." Then he ran, before Dudley could work out what he'd said. One day in July, Aunt Petunia took Dudley to London to buy his Smeltings uniform, leaving Harry at Mrs. Figg's. Mrs. Figg wasn't as bad as usual. It turned out she'd broken her leg tripping over one of her cats, and she didn't seem quite as fond of them as before. She let Harry watch television and gave him a bit of chocolate cake that tasted as though she'd had it for several years. That evening, Dudley paraded around the living room for the family in his brand-new uniform. Smeltings boys wore maroon tailcoats, orange knickerbockers, and flat straw hats called boaters. They also carried knobbly sticks, used for hitting each other while the teachers weren't looking. This was supposed to be good training for later life. As he looked at Dudley in his new knickerbockers, Uncle Vernon said gruffly that it was the proudest moment of his life.

Appendix 2

HP7 – Chapter 3, The Dursleys Departing (initial 25 sentences)

The sound of the front door slamming echoed up the stairs and a voice yelled, “Oi! You!” Sixteen years of being addressed thus left Harry in no doubt whom his uncle was calling; nevertheless, he did not immediately respond. He was still gazing at the mirror fragment in which, for a split second, he had thought he saw Dumbledore’s eye. It was not until his uncle bellowed, “BOY!” that Harry got slowly to his feet and headed for the bedroom door, pausing to add the piece of broken mirror to the rucksack filled with things he would be taking with him. “You took your time!” roared Vernon Dursley when Harry appeared at the top of the stairs. “Get down here, I want a word!” Harry strolled downstairs, his hands deep in his jeans pockets. When he reached the living room he found all three Dursleys. They were dressed for traveling: Uncle Vernon in a fawn zip-up jacket, Aunt Petunia in a neat salmon-colored coat, and Dudley, Harry’s large, blond, muscular cousin, in his leather jacket. “Yes?” asked Harry. “Sit down!” said Uncle Vernon. Harry raised his eyebrows. “Please!” added Uncle Vernon, wincing slightly as though the word was sharp in his throat. Harry sat. He thought he knew what was coming. His uncle began to pace up and down, Aunt Petunia and Dudley following his movements with anxious expressions. Finally, his large purple face crumpled with concentration, Uncle Vernon stopped in front of Harry and spoke. “I’ve changed my mind,” he said. “What a surprise,” said Harry. “Don’t you take that tone —” began Aunt Petunia in a shrill voice, but Vernon Dursley waved her down. “It’s all a lot of claptrap,” said Uncle Vernon, glaring at Harry with piggy little eyes. “I’ve decided I don’t believe a word of it. We’re staying put, we’re not going anywhere.” Harry looked up at his uncle and felt a mixture of exasperation and amusement. Vernon Dursley had been changing his mind every twenty-four hours for the past four weeks, packing and unpacking and repacking the car with every change of heart.

Appendix 3

HP1 Table for micro-analysis

number	book	sentence	number of words	number of clauses	coordination main clauses	NP modifiers - finite	NP modifiers - nonfinite	other phrase - finite	other phrase nonfinite	clause constituent - finite	clause constituent - nonfinite	direct speech
1	HP1	The escape of the Brazilian boa constrictor earned Harry his longest-ever punishment.	12	1								
2	HP1	By the time he was allowed out of his cupboard again, the summer holidays had started and Dudley had already broken his new video camera, crashed his remote control airplane, and, first time out on his racing bike, knocked down old Mrs. Figg as she crossed Privet Drive on her crutches.	51	6	4	1				1		
3	HP1	Harry was glad school was over, but there was no escaping Dudley's gang, who visited the house every single day.	20	4	2	1		1				
4	HP1	Piers, Dennis, Malcolm, and Gordon were all big and stupid, but as Dudley was the biggest and stupidest of the lot, he was the leader.	25	3	2					1		
5	HP1	The rest of them were all quite happy to join in Dudley's favorite sport: Harry Hunting.	16	2					1			
6	HP1	This was why Harry spent as much time as possible out of the house, wandering around and thinking about the end of the holidays, where he could see a tiny ray of hope.	33	5		1				1	2	
7	HP1	When September came he would be going off to secondary school and, for the first time in his life, he wouldn't be with Dudley.	24	3	2			1				
8	HP1	Dudley had been accepted at Uncle Vernon's old private school, Smeltings.	11	1								
9	HP1	Piers Polkiss was going there too.	6	1								
10	HP1	Harry, on the other hand, was going to Stonewall High, the local public school.	14	1								
11	HP1	Dudley thought this was very funny.	6	2						1		
12	HP1	"They stuff people's heads down the toilet the first day at Stonewall," he told Harry.	15	2								1
13	HP1	"Want to come upstairs and practice?"	6	1								1
14	HP1	"No, thanks," said Harry.	4	2								1
15	HP1	"The poor toilet's never had anything as horrible as your head down it — it might be sick."	16	2	2							2
16	HP1	Then he ran, before Dudley could work out what he'd said.	11	3						2		
17	HP1	One day in July, Aunt Petunia took Dudley to London to buy his Smeltings uniform, leaving Harry at Mrs. Figg's.	20	3							2	
18	HP1	Mrs. Figg wasn't as bad as usual.	7	1								
19	HP1	It turned out she'd broken her leg tripping over one of her cats, and she didn't seem quite as fond of them as before.	24	4	2					1	1	
20	HP1	She let Harry watch television and gave him a bit of chocolate cake that tasted as though she'd had it for several years.	23	4	2	1				1		
21	HP1	That evening, Dudley paraded around the living room for the family in his brand-new uniform.	15	1								
22	HP1	Smeltings boys wore maroon tailcoats, orange knickerbockers, and flat straw hats called boaters.	13	1								
23	HP1	They also carried knobbly sticks, used for hitting each other while the teachers weren't looking.	15	3			1			1		
24	HP1	This was supposed to be good training for later life.	10	1								
25	HP1	As he looked at Dudley in his new knickerbockers, Uncle Vernon said gruffly that it was the proudest moment of his life.	22	3						2		

Appendix 4

HP7 Table for micro-analysis

number	book	sentence	number of words	number of clauses	coordination main clauses	NP modifiers - finite	NP modifiers - nonfinite	other phrase - finite	other phrase nonfinite	clause constituent - finite	clause constituent - nonfinite /verbless	direct speech
1	HP7	The sound of the front door slamming echoed up the stairs and a voice yelled, "Oi! You!"	17	4	2		1					1
2	HP7	Sixteen years of being addressed thus left Harry in no doubt whom his uncle was calling; nevertheless, he did not immediately respond.	22	4	2	1	1					
3	HP7	He was still gazing at the mirror fragment in which, for a split second, he had thought he saw Dumbledore's eye.	21	3		1				1		
4	HP7	It was not until his uncle bellowed, "BOY!" that Harry got slowly to his feet and headed for the bedroom door, pausing to add the piece of broken mirror to the rucksack filled with things he would be taking with him.	41	8		1	1			3	2	
5	HP7	"You took your time!" roared Vernon Dursley when Harry appeared at the top of the stairs.	16	3						1		1
6	HP7	"Get down here, I want a word!"	7	2								2
7	HP7	Harry strolled downstairs, his hands deep in his jeans pockets.	10	2							1	
8	HP7	When he reached the living room he found all three Dursleys.	11	2						1		
9	HP7	They were dressed for traveling: Uncle Vernon in a fawn zip-up jacket, Aunt Petunia in a neat salmon-colored coat, and Dudley, Harry's large, blond, muscular cousin, in his leather jacket.	30	4							3	
10	HP7	"Yes?" asked Harry.	3	2								1
11	HP7	"Sit down!" said Uncle Vernon.	5	2								1
12	HP7	Harry raised his eyebrows.	4	1								
13	HP7	"Please!" added Uncle Vernon, wincing slightly as though the word was sharp in his throat.	15	4						1	1	1
14	HP7	Harry sat.	2	1								
15	HP7	He thought he knew what was coming.	7	3						2		
16	HP7	His uncle began to pace up and down, Aunt Petunia and Dudley following his movements with anxious expressions.	18	2							1	
17	HP7	Finally, his large purple face crumpled with concentration, Uncle Vernon stopped in front of Harry and spoke.	17	3	2						1	
18	HP7	"I've changed my mind," he said.	6	2								1
19	HP7	"What a surprise," said Harry.	5	2								1
20	HP7	"Don't you take that tone —" began Aunt Petunia in a shrill voice, but Vernon Dursley waved her down.	18	3	2							1
21	HP7	"It's all a lot of claptrap," said Uncle Vernon, glaring at Harry with piggy little eyes.	16	3							1	1
22	HP7	"I've decided I don't believe a word of it.	9	2						1		
23	HP7	We're staying put, we're not going anywhere."	7	2	2							
24	HP7	Harry looked up at his uncle and felt a mixture of exasperation and amusement.	14	2	2							
25	HP7	Vernon Dursley had been changing his mind every twenty-four hours for the past four weeks, packing and unpacking and repacking the car with every change of heart.	27	2							1	