A Corpus Linguistic Analysis of Characterisation in Chimamanda Ngozi Adichie's *Americanah*

Albert Omolegbé KOUKPOSSI[1] & Moustafa GUEZOHOUEZON[2]

[1] Faculté des Lettres Langue Arts et Communication, Université d'Abomey-Calavi, Bénin

[2] Université Nationale d'Agriculture (UNA), Bénin

Correspondence: Albert Omolegbé KOUKPOSSI, Faculté des Lettres Langue Arts et Communication, Université d'Abomey-Calavi, Bénin.

Abstract

Research in Corpus Linguistics has provided insights into the literary meanings of texts over the past few decades. Building on this foundation, this study aims to enhance the understanding of literary texts by employing corpus tools to analyze the methods of characterisation in Chimamanda Ngozi Adichie's novel, *Americanah* (2013). Specifically, it uses both quantitative and qualitative methods to examine the themes and characters in the novel, illuminating the author's ideas and style. Through a corpus-based approach, this research examines the keywords and clusters that define the novel's narrative structure, revealing the ideologies and literary techniques Adichie uses to craft characters and impart meaning about their experiences. The characterisation analysis uncovers recurring themes, rhetorical strategies, and linguistic patterns that enrich the portrayal of characters and narrative progression. The findings provide insights valuable for educational contexts, allowing students and readers to gain a thorough understanding of the text prior to delving into its more complex aspects.

Keywords: *Americanah*, characterisation, Corpus Linguistics, keyword analysis, n-gram

1. Introduction

A prevailing contentious issue exists between the domains of literary criticism and stylistics concerning the level of objectivity ascribed to their methodological approaches when analyzing elements such as literary style, characters, narrators, among others, in literary texts. Ayeomoni (2003) contends, "While literary criticism is traditionally based on subjective interpretations of texts, stylistics offers an analysis grounded in the linguistic structures inherent within the text." Consequently, as the quest for objectivity in language analysis methods intensifies, the incorporation of empirical data to substantiate claims within literary text analysis becomes imperative. In this vein, the application of Corpus Linguistics (CL), considered as "a scientific method of language analysis" by Brezina (2018), to the investigation of language in context presents a dual-faceted framework that yields both quantitative and qualitative insights, thereby underpinning the analytical and interpretive processes of literary texts.

Moreover, discussions on methods of characterisation in high esteemed literary works are not novel, as reflected in the contributions of Erwindriani (2012) and Akongo and Ngassaki (2022). The precedent of integrating computer-assisted analysis in literary studies also has been well-established, as evidenced by the works of Culpeper and Demmen (2015), Mahlberg (2007), and the collaborative research by Mahlberg, Smith, and Preston (2013). Many scholars have employed corpus tools, including collocations, colligations, Keywords in Context (KIWIC), and frequency analyses, to dissect literary texts. This study, however, sets itself apart by leveraging statistical data to elucidate and interpret theoretical concepts such as characterisation methods in Chimamanda Ngozi Adichie's acclaimed novel, *Americanah*. Corpus Linguistics thus emerges as a pivotal tool in deconstructing character development and portraying consistency through the analysis of keywords and n-grams. This study posits that a methodical approach, utilising the technical resources of CL, can unearth profound insights into Adichie's literary techniques and narrative strategies. It is not merely about delineating literary terms and devices or dissecting a work's tangible elements to recognize characters, themes, and plots. It is about enriching such discourse with empirical backing, specifically through statistical evidence extracted from the corpus, as advocated by Brezina (2018).

This study aims to investigate how Chimamanda Ngozi Adichie uses language and style to develop characters and narrative in her novel *Americanah*. It explores the linguistic and stylistic choices that shape the story and its thematic concerns, offering insights into how Adichie crafts a text that deeply engages the reader.

To crystallise the central aim and lay down a solid theoretical and methodological foundation for this study, the following key questions are posited:

- What keywords and key clusters (n-grams) have been used in shaping the narrative structure of *Americanah*?
- How do these linguistic constructs of keywords and key clusters serve to forge distinctive characterisations and craft a 'readerly text' within the novel?

Answering these questions will not only reveal the ideological underpinnings interwoven within the novel's narrative discourse but will also illuminate the means through which the author conveys her message to the readers, thereby aligning them with her ideological stance.

2. Theoretical Framework: Corpus Linguistics

This study anchors its theoretical framework in Corpus Linguistics to investigate the characterisation within Chimamanda Ngozi Adichie's *Americanah*. Understanding the term "corpus" is essential to grasp the objectives of Corpus Linguistics in language analysis. "Corpus," plural "corpora," refers to a systematically organised collection of text samples, chosen with explicit linguistic criteria in mind to serve as a representative sample of a language (Sinclair, 1994a, p. 2). Francis (1991, p. 17) similarly defines a corpus as a representative collection of texts or language subsets for linguistic scrutiny. The precise selection criteria and the move towards electronic corpora for multi-purpose analysis are central to these definitions.

Corpus Linguistics is more a methodological tool than a new linguistic theory. Brian Paltridge illuminates this by noting that corpus studies are based on electronically stored text collections analysed to track specific linguistic features' frequency and patterns within discourse (Paltridge, 2012, p. 144). This approach underlines the importance of Corpus Linguistics for analysing language systematically, particularly in its application to Stylistics and Discourse Analysis.

Applying this framework, our study focuses on the keywords and key clusters—lexical and structural elements that recurrently appear in the text of *Americanah*. These elements are instrumental for qualitative stylistic analysis, as they often reveal a text's underlying themes and stylistic nuances (Leech & Short, 2007). By examining these components, the study aims to dissect the methods Adichie employs for character development and to scrutinise how these linguistic patterns contribute to the narrative and thematic structure of the novel.

3. Literature Review

A wide range of approaches exists when describing the language of literary texts (White, 1998). From a stylistic point of view, it is possible to unveil a literary text's communicative and thematic functionality by attempting to uncover salient lexical and grammatical features of a given text. So far, traditional approaches to stylistics have been used to carry out such an exercise of text explication. However, some prominent figures in the field see these approaches as limited, particularly in their restriction to short texts and text extracts (Fischer-Starche, 2010, quoted in Moustafa, 2022, p. 47). Another criticism of traditional stylistics, as opposed to modern stylistics based on quantitative methods, is that it is mostly intuition-based. This is where Corpus Linguistics makes an innovative contribution to literary studies. Moustafa (2022) argues that it provides "a rather objective means of analysing a fictional work" (p. 47).

Considering all the added value Corpus Linguistics brings to the systematic description of the language of literary texts, this study adopts corpus stylistics as the theoretical foundation to identify recurrent linguistic patterns focusing on a quantitative approach to meaning detection and methods of characterisation in *Americanah*, and then proceeds to interpret the quantitative findings (Egbert, Larson, & Biber, 2020).

The Corpus Linguistics analytical tools used in this study include keyword analysis, concordance lines, and n-gram analysis, etc. In Corpus Linguistics, keyword analysis is an essential tool for systematically investigating the subject matter and thematic concerns of a target corpus. Jonathan Culpeper and Jane Demmen (2015) observe that this analytical tool is crucial for determining the "keyness" (aboutness) of a text "because they capture the essence of particular social, cultural, or political themes, thoughts, or discourses" (p. 90). Technically, "keyness" is determined by comparing elements of the reference corpus (often more extensive than the target corpus) but following the discourse type or register. In this study, the American Corpus 06 (AmE 06), especially the subcorpus comprising texts including fiction, science fiction, etc is considered. Moreover, the Log-likelihood and its relative effect are the statistical measures to be considered in this keyword analysis exercise (Anthony,

2019). AntCorc, which is one of the most used concordance programmes, automatically generates this measure. It is based on the calculation/identification of the frequency of a given word to the total number of words in the target corpus (*Americanah*) and compares it to those of the same word found in the reference corpus. From this perspective, some scholars (Moustafa, 2022; Michaela Mahlberg & Dan McIntyre, 2011), believing in the accuracy and scientific value of such a corpus analytical tool, use it to explore literary meanings in works of fiction.

More specifically, Moustafa (2022), in a comparative analysis based on the Corpus Stylistics approach to literary studies, uses the keyword analytical tool to pinpoint the concepts of thematization and characterisation in two different novels crafted by two Nobel Prize winners from South Africa, Nadine Gordimer in *My Son's Story*, and John Maxwell Coetzee in *Disgrace*. Each of these novels has been treated as two distinct corpora using Sketch Engine (Kilgarriff, 2009). They have also been programmed as the reference corpus for each other to sort out a list of the most recurrent words in each corpus. The findings that Moustafa has come to in such an exercise in literary text analysis are of paramount interest for this study. In well-organized tables, he displayed the most recurrent keywords of each novel "with their rank, frequencies, and keyness score" (p. 53). Placing each keyword in its concordance line generated by Sketch Engine (a concordancer), he has been able to systematically determine the thematic concerns of each literary work. For example, the keywords "cottage," "cinema" and "kid" found in *My Son's Story*, and placed in their respective concordance lines (context), reveal that it is all about love affairs. Moreover, the word "kid," collocating with many other keywords such as "detention," "liberation," "revolutionary," etc, supports the argument that Nadine Gordimer's *My Son's Story* is about "racial discrimination and the struggle against political injustice."

In the same way, Collins' (2015) contribution to *Linguistik Online* informs about the possibility of using keyword analytical tools to pinpoint the 'aboutness' of a corpus and annotate data for their syntactic and semantic elements. In his article titled "How can semantic annotation help us to analyse the discourse of climate change in online user comments," Collins, using the keyword analytical tool offered by Wmatrix concordancer, has been able to generate the keyword list about the user comments of some newspaper articles published by "The Guardian". The analysis and interpretation of his findings help to identify the recurrence of a set of ten words that offer a rich and fascinating perspective in characterising discussions about climate change among Guardian readers.

Another exciting and impressive work related to the characterisation analysis of fictional work, which informs the theoretical and methodological orientation adopted in this study, is an online video (https://youtu.be/8BWXYbR_lo4) accounting for Professor Monika Bednarek's (University of Sydney) contribution to the 2021 annual Sinclair Lecture (2022). The title of this talk is 'Language and Characterization in Television Series'. Although the corpus of analysis is not directly linked to a literary corpus, it encapsulates innovative and interesting technical analytical tools applicable to literary fiction for optimal cogent results and valuable conclusions. Bednarek (2022) provides an analysis typically based on the corpus analytical tool, mainly on keyword analysis and N-gram, through a case study of "Gilmore Girls," "The Big Bang Theory," and Ngara Corpus, three famous television series, to perform what she terms "television characterisation." To achieve this objective, she designs a framework organised into four axes: character differentiation, character expressivity, character stability, character stereotypicality, character self and other representation. Specifically, character differentiation is about whether linguistic resources differentiate characters, and character expressivity is related to the role of emotion and other expressions in characters' language. Regarding character stability and character stereotypicality, they are related to characters' language changes across the narrative and the role of stereotypes in building a character. Finally, character self and other representation are about how characters represent themselves compared to others. In essence, to achieve the objectives presented in her study's framework, Bednarek (2022) relies on crucial and fundamental analytical tools offered by concordancing programmes which are keyword and N-gram analytical tool (clusters).

4. Research Methodology

4.1 Corpus

The corpus under examination is Chimamanda Adichie's novel *Americanah*. To explore the author's methods of characterisation to create a text that engages the reader, the entire novel has been considered. The PDF version, available for sale on the internet, has been processed and converted into a plain text file (txt extension with UTF-8 encoding) and then uploaded into Antconc (version 4.0). The analysis reveals that the target corpus comprises 171,016 tokens (the total number of words) and 11,582 word types (the number of distinct words). For instance, the word 'the' is a word type that occurs 6,260 times in the corpus.

4.2 Procedure

4.2.1 Keyword Analysis

Bednarek (2022) specifies that keyword analysis in Corpus Linguistics is "a technique that identifies words (or other items) that are salient in one corpus compared to another. Keywords are, by definition, statistically salient in one corpus when contrasted with a reference corpus. The analysis prioritizes the frequencies and distribution of these keywords within the target corpus. An important question is how keywords and n-grams relate to characterisation. Culpeper (2001), as cited by Bednarek (2022), observes that keyword analysis "offers a unique opportunity to reveal style markers in characters' speech." To analyze characterisation methods through keywords, one should examine the frequency and range of keywords throughout the corpus to address the differentiation, expressiveness, and representation of the characters.

Mahlberg and McIntyre (2011) suggest that the classification of keywords for analysis should follow a categorisation of the most salient words (keywords) according to their semantic domain. Thus, two main categories are identified: keywords related to 'fictional worlds' and those related to 'thematic signals'. 'Thematic signal' keywords, which are less concrete than 'fictional world' keywords, tend to allow for broader interpretations. The analysis of fictional world keywords, which are linked to concrete characters and objects, mainly involves identifying relevant connections to characters, objects, and places. The thematic category predominantly comprises abstract and metaphorical terms that necessitate more nuanced interpretation. Identifying 'fictional world' and 'thematic signal' keywords requires analysing concordance lines to discern the meanings of the Keywords in Context.

4.2.2 Cluster Analysis

In Corpus Linguistics, clusters, also known as N-grams, are sequences of contiguous words that occur repeatedly within a corpus. These sequences can consist of a single word, referred to as a 1-gram or unigram. Sequences of two words are known as 2-grams or bigrams, sequences of three words as 3-grams or trigrams, and so on, with sequences of N words termed N-grams. N-grams are instrumental in word prediction and language modeling. The significance of cluster analysis in this study lies in how a consistent sequence of words and their frequencies can illuminate local textual functions as stylistic features in fiction (Mahlberg, 2007). Table 1 presents the top twenty examples of two-, three-, and four-grams found in *Americanah*.

Table 1. Top twenty-two-, three- and four-word clusters in *Americanah*

Two-word clusters	F	Three-word clusters	F	Four-word clusters	F
in the	539	i don t	134	i don t know	33
she had	445	she did not	126	did not want to	27
she was	420	you don t	64	i m going to	22
of the	396	he did not	57	she did not want	22
it was	387	it s not	54	from time to time	21
did not	375	she wanted to	54	i don t want	19
on the	360	don t know	47	that she did not	19
don t	356	she could not	47	for a long time	18
he was	339	aunty uju s	45	she did not know	17
it s	311	aunty uju said	45	i don t have	15
aunty uju	308	i can t	45	but she did not	14
she said	307	that she was	44	don t want to	14
he had	301	did not want	43	in front of the	14
and she	288	he told her	43	what do you mean	14
in a	287	she had not	41	as though she had	13
at the	228	as though she	40	the first time she	13
to be	227	did not know	40	in the living room	12
to the	257	that she had	40	to go to the	12
i m	251	there was a	40	you don t have	12
as though	230	i m not	39	he did not want	11

Note. F = Frequency.

AntConc can generate N-grams that extend beyond four words, as Table 1 illustrates. For example, 5-word clusters in *Americanah* include "she did not want to" and "America for the non-American black." Six-word

clusters feature phrases such as "understanding America for the non-American." The software is capable of producing 7-, 8-, or more word clusters tailored to the specific objectives of a study. According to Mahlberg (2007), while shorter clusters are often more frequent and flexible, making them challenging to characterize, longer clusters tend to be more distinctive, closely associated with a particular text. In this study, we have focused on four-word clusters for analysis.

Table 1 reveals that the N-grams manifest in diverse syntactic structures: clauses with a pronoun subject followed by a verb (e.g., "I don't know," "I'm going to"), noun phrases (e.g., "Aunty Uju"), and prepositional phrases (e.g., "from time to time"). What distinguishes Mahlberg's (2007) analysis—and what is particularly relevant to this study—is not solely the grammatical composition of the clusters but the functional roles they play within the corpus. Mahlberg categorizes these functions into five distinct groups: labels, speech clusters, 'as if' clusters, body part clusters, and time and place clusters.

Labels

Clusters in this category either include names of characters or places or are directly associated with them, as noted by Mahlberg (2007). Examples such as "Aunty Uju said" and "the house in Dolphin Estate" clearly illustrate this. However, clusters like "the kind of man who" require further context to understand their connection to specific characters or concepts. Mahlberg provides additional insight: "Clusters in the Labels group tend to be linked to particular characters and concepts, which is evident when we examine both quantitative and qualitative aspects of the data" (p. 14).

Speech clusters

This category encompasses clusters that include pronouns such as "I" or "you," or possessive forms, suggesting a dialogic or interactive element. It is important to note that when clusters could fall into both this category and Labels, the Labels category is prioritised for classification purposes. Representative examples include "I don't know" and "what do you mean."

As if clusters

Characterized by the initiating phrase "as if," these clusters often set up similes or hypothetical scenarios. In *Americanah*, for instance, we find "as if she and Ifemelu" and "as if a novel had," which add a layer of speculation or comparison to the narrative.

Body part clusters

Clusters containing terminology related to human anatomy are classified here. Phrases like "to shake Obinze's hand" and "her head like a halo" are prime examples that evoke imagery centred around physical features.

Time and place clusters

This group comprises clusters that explicitly mention temporal or locational elements, such as "from time to time" and "in the living room." These clusters help anchor the narrative in a specific setting or time frame.

5. Analysis of Characterisation in Adichie's *Americanah*

5.1 Keyword Analysis

The emphasis on keywords in the corpus analysis of a literary text is well-founded, given that they act as "signals for the building of fictional worlds as well as triggers for the thematic concerns of the novel" (Mahlberg & McIntyre, 2011). Consequently, keyword analysis proves to be a powerful tool in examining thematic elements and characterisation within a literary work, as it involves the systematic examination of word frequencies. To effectively utilize this method, Mahlberg and McIntyre recommend comparing the identified keywords against established key semantic domains. In their approach, the initial broad categorisation of these domains includes three primary categories: "names," "words that indicate the aboutness of the text," and "grammatical words or words that relate to the style of the text." Adhering to Mahlberg and McIntyre's (2011) analytical framework, this study concentrates on the first 150 keywords with a Likelihood Ratio (LL) value of 31.31 or higher. This threshold is chosen to ensure the relevance and significance of the keywords in the context of the novel.
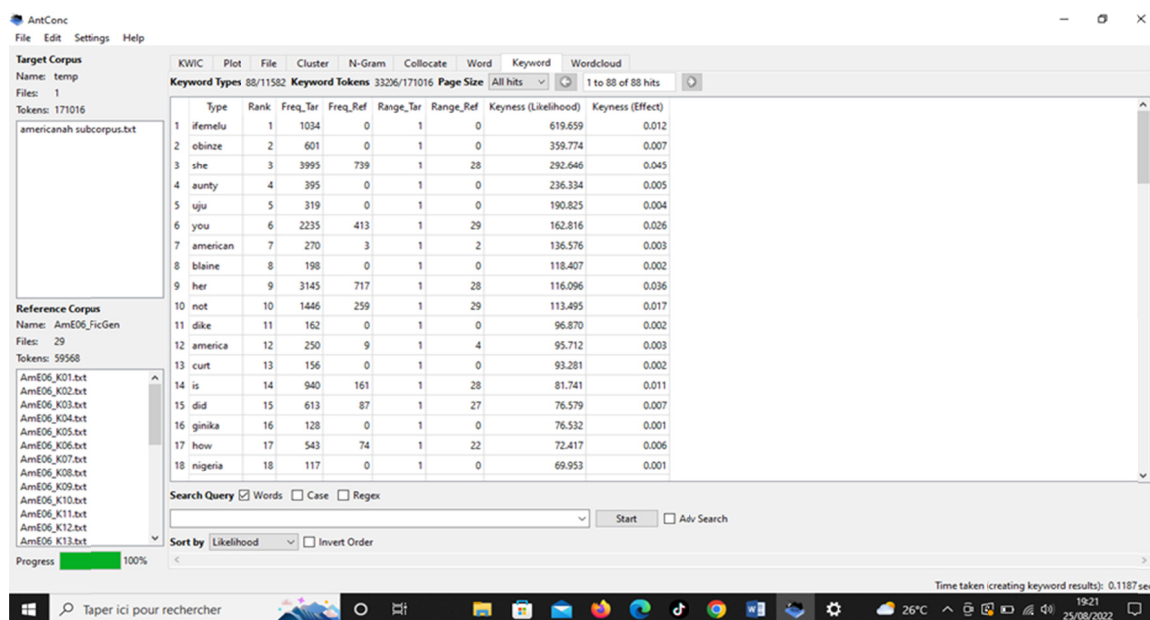
Figure 1. Keywords: The first 18 statistically significant keywords in *Americanah*

Figure 1 illustrates the first 18 keywords of the target corpus, *Americanah*, along with their frequency in the American fictional corpus (AmE 06) and their log-likelihood (LL) values. An in-depth analysis of the top 150 keywords reveals several categories, such as names, including Ifemelu, Obinze, Uju, Blaine, etc.; words that signify the novel's central themes, like "American," "black," "race" and "blog"; and grammatical words such as "she," "you," and "her." Through these keywords generated by AntConc and categorised into "keywords related to the fictional world" and "keywords related to thematic signals," we can quantitatively discern the novel's thematic and character-driven aspects. This categorisation is based on an examination of concordance lines, which helps determine if the keywords are concrete and plot-centric or more abstract and emotionally charged. Similar keywords are grouped together, with the findings presented in Table 2.

Table 2. Keyword classification of fictional words and thematic signals in *Americanah*

Categories	Examples
Fictional world	
Characters	Ifemelu, obinze, Uju, Blaine, Dike, Curt, Ginika, Emenike, Kosi, Obama, Nigel, Araminata,
Clothes and accessories	Craftan, afro
places	America, Nigeria, Lagos, Nsukka, Ibadan, Ikoyi, Abuja
Professions	Housegirl, braiding, braider, babysit, waiter, gateman
Authorities	Oga, Chief
Literature and Music	Yori
Food and drinks	Jollof rice
Tools	Relaxer, e-mails
Thematic signals	Nigerpolitan, halfcaste, race, black,

The analysis involves correlating keywords associated with the fictional world to specific characters and tangible elements within the text, thus elucidating their significance in both characterisation and thematic development. Excluding character names—which McIntyre and Walker (2019) assert will invariably emerge as keywords when contrasted with a reference corpus (as cited in Moustafa, 2022)—the subsequent table outlines these findings:

Table 3. Keywords with characters' names excluded

Rank	Keywords	Frequency	Keyness (Likelihood)
3	she	3995	292.646
6	you	2235	162.816
7	American	270	136.576
9	her	3145	116.096
10	not	1446	113.495
12	America	250	95.712
14	is	940	81.741
15	did	613	76.579
17	how	543	72.417
18	Nigeria	117	69.953
19	told	358	68.587
20	about	746	68.030
22	black	322	66.455
23	race	108	64.571
24	said	1646	64.557
25	Lagos	105	62.777
29	are	524	53.624
30	blog	89	53.209
31	would	921	52.737
35	Nigerian	84	50.219
36	asked	395	46.515
37	because	467	46.480
38	African	72	43.043

The most prevalent keywords in the corpus are the pronouns "she" and "you," with log-likelihood values of 292.646 and 162.816, respectively. Despite being grammatical words, their high frequency underscores a dialogic narrative style within the novel. When examining the concordance lines for "she," we predominantly encounter references to Ifemelu, the central character whose experiences in America and Nigeria frame the narrative. This narrative perspective is reinforced by the recurrent use of "her," the fourth-ranked keyword, hinting at the novel's emphasis on female perspectives.

While "America" and its variants appear frequently, highlighting the setting's centrality, they also signify more than mere geography. They encapsulate Ifemelu's confrontations with cultural identity, race, and gender issues, underpinning the novel's thematic 'aboutness.'

The term "race" emerges as a philosophical cornerstone in *Americanah*, illustrated by its log likelihood of 64.571 and 108 occurrences within the text. Analysis of this keyword's concordance lines reveals the narrator's profound engagement with issues of race, providing critical insights into the underlying ideologies at play.

Because his mother was white. But	race	is not biology; race is sociology. Race is not genotype; race
is sociology.	Race	is not genotype; race is phenotype. Race matters because of racism.
So What's the Deal? They tell us	race	is an invention, that there is more genetic variation between two
Race is not genotype;	race	is phenotype. Race matters because of racism. And racism is absurd
But race is not biology;	race	is sociology. Race is not genotype; race is phenotype. Race matters
"We're tired of talking about race" or "The only	race	is the human race." American Blacks, too, are tired of talking
"The first step to honest communication about	race	is to realise that you cannot equate all racisms," she said,
he was a social warrior and might make a good guest blogger.	Race	is totally overhyped these days, black people need to get over
Of all their tribalisms, Americans are most uncomfortable with	race.	If you are having a conversation with an American, and you
of the bad morals or the poverty or the illiteracy of your	race?	If you do well in a situation, do you expect to

Figure 2. Concordance lines of the keyword "race" in *Americanah*

Figure 2 displays some examples of the 108 occurrences of the word "race" in *Americanah*, from which we can cite clause fragments such as "race is not biology," "race is not genotype," "race is an invention," "race is sociology," "the only race is the human race," and "race is totally overhyped these days." Most of these instances appear in local contexts where Ifemelu and other characters engage in self-reflections and make statements about the significant issue of race and racism in America, particularly its implications for relationships between black and white individuals.

The occurrence of the characters' is also significant to novel's "aboutness". Characters' names, which are integral to the fictional world, serve as narrative elements around which the plot is constructed, and they can also indicate the overall themes and characterisation within the novel. The name Ifemelu appears 1,034 times with a frequency of 619.659, followed by Obinze, her boyfriend, whose name occurs 601 times. There is a strong bond between both characters, and their presence can be traced from their early school days to Ifemelu's return to Nigeria. When the name Ifemelu is mentioned in specific contexts, it is often in association with kinship terms such as mother, father, and aunt, revealing different facets of her personality.

5.2 Cluster Analysis

As previously mentioned, identifying the five-cluster groups (Mahlberg, 2007) is essential for analyzing character development in Americanah. These N-grams are instrumental in pinpointing stylistic techniques related to:

- Characterisation of people, places, and things within the narrative (label clusters)
- Depiction of interactions between characters (speech clusters)
- Descriptions of physical appearance and movement (body part clusters)
- Creation of textual worlds via comparison and contrast (as if clusters)
- Setting of actions in time and place within the story (time and place clusters)

The forthcoming analysis will classify the top fifty four-word clusters identified in *Americanah*, using the following key:

L = Label Cluster S = Speech Cluster TP = Time and Place Cluster AI = As If Cluster BP = Body Part Cluster O = Other.

Table 4. Top fifty four-word clusters in *Americanah*

Rank	clusters	Fonction group
1	i don t know	S
2	did not want to	O
3	i m going to	S
4	she did not want	L
5	from time to time	TP
6	i don t want	S
7	that she did not	L
8	for a long time	TP
9	she did not know	L
10	i don t have	S
11	but she did not	L
12	don t want to	O
13	in front of the	TP
14	what do you mean	S
15	as though she had	AI
16	the first time she	TP
17	in the living room	TP
18	to go to the	O
19	you don t have	S
20	he did not want	L
21	said i don t	S
22	as though he was	AI
23	at the same time	TP
24	but i don t	S
25	i don t think	S
26	in the middle of	TP
27	as though he had	AI
28	as though it was	AI
29	don t know what	O
30	she told him that	L
31	you don t want	S
32	did not know what	O
33	he was looking at	L
34	i didn t know	S
35	i don t even	S
36	i don t understand	S
37	i just want to	S
38	it had to be	O
39	it was as if	AI
40	was looking at her	L
41	we re going to	L
42	you don t know	S
43	you know what i	S
44	and for this she	L
45	and she did not	L
46	as though they had	AI
47	as though to say	AI
48	at the dining table	TP
49	but he did not	L
50	do you worry that	S

As can be observed from the table of the top fifty four-word clusters in *Americanah*, there are five different categories into which the four-word clusters are classified: the Label group, the Speech group, the As if group, the Time and Place group, and the last group, labelled "O," which does not fit any of the aforementioned categories. The Label group and the Speech group are the most prevalent and typically determine the methods of characterisation and the building of relationships between characters. While no instances of body part clusters are present in the top fifty four-word clusters, they become identifiable upon further examination of the cluster occurrence rankings.

5.2.1 Label Analysis

The first cluster identified to belong to this group is "she did not want," a clause fragment where "she" typically refers to characters in the corpus. Concordance line analysis reveals predominantly relates to Ifemelu, as illustrated by examples such as:

> Ifemelu was already looking for a studio apartment, … and *she did not want* to be further enmeshed in the lives of the Turners, …

> They kissed often, but always with her lips firmly pursed; *she did not want* his tongue in her mouth.

This cluster consisting of "she" (the referent to Ifemelu across the novel) and the mental desiderative process "want" in the negative form, significantly characterises Ifemelu's assertion of independence and her resistance to external control other her choices.

The next cluster, "that she did not" appears 19 times in the novel. The pronoun "she" predominantly refers to Ginika, the Russian girl, Kosi, Ifemelu, and occasionally other female characters. Most occurrences are associated with Ifemelu. An in-depth analysis of the concordance lines reveals that it is usually paired with verbs denoting mental processes like "like" "understand," "want," "care," "know," "think", and others forming phrases such as "that she did not like," "that she did not understand," and "that she did not care." These instances further illustrate Ifemelu's portrayal an assertive and independent character, consistently described as someone who asserts her will. Overall, these clusters contribute to the characterisation of Ifemelu as a decisive and strong-willed woman.

5.2.2 Speech Clusters

Speech clusters are notably significant in character description within the novel. Mahlberg (2007) highlights that analysing such N-grams provides insights into character interactions and emphasises their dialogues. The most frequently identified four-word clusters in the novel include "I don't know," "I did not want to," "I'm going to," "I don't want," "I don't have," and "what do you mean." The cluster "I don't like" reveals much about Ifemelu's character, shedding light on her cultural values, identity, and philosophical outlook. For instance, phrases such as "I don't like that" and "I don't like America" express a negative sentiment and a natural aversion that Ifemelu harbors towards American culture and environment. Despite her achievements in America, this recurring cluster underscores Ifemelu's feelings of alienation and her struggle with cultural belonging.

5.2.3 "As if" Cluster

The significance of this cluster lies in its role in creating textual worlds in fiction. Mahlberg (2007, p. 21) states that the "As-If" group concerns clusters that compare or comment on actions and situations in a story. In other words, it helps the narrator comment on a given situation through comparison and contrasting linguistic elements. In *Americanah*, clusters made up of the conjunctive particle "as if" occur 69 times and contribute to the characterisation of some characters' physical, mental and psychological states. The clusters in this category include "as if he could," "as if he had," "as if he was," "as if I'm," "as if she had," "as if they were," "as if we are," etc. Other structures can also be classified in this category, such as "as though," which appears 148 times in the novel. Instances of this type include "as though she had," "as though he was," "as though he had," "as though it was," "as though they had," "as though to say," etc. It is essential to notice at this level that "as if" or "as though" is followed by the first, second, and third-person pronouns, which refer to characters. Therefore, they can be argued to be used to portray characters and create textual worlds. For example, when Blaine, a black American, met Ifemelu for the first time, they exchanged smiles and jokes, and the impression that Blaine had in this situational event is presented as follows:

> "He asked, still smiling. There was something both ironic and gentle about him. It was *as if he* believed that they shared a series of intrinsic jokes that did not need to be verbalized" (Chapter 17).

"As if he" here prompts the description of Blaine as a humorous character, an attribute that he and Ifemelu share. This is a foreshadowing of the love relationship between both characters.

5.2.4 "As If" Clusters

According to Mahlberg (2007, p. 21), the "As If" cluster plays a pivotal role in constructing textual worlds by allowing for commentary on actions and situations through comparison. In *Americanah*, the use of "as if" appears 69 times, contributing to the portrayal of characters' states. This category encompasses structures like "as if he could" and "as if she had," as well as the similar "as though," which is found 148 times in the text. These phrases often precede pronouns that refer to characters, thus serving to depict characters' mental and emotional states. For instance, when Blaine first meets Ifemelu, his impression is captured as: "He asked, still

smiling. There was something both ironic and gentle about him. It was as if he believed that they shared a series of intrinsic jokes that did not need to be verbalized" (Chapter 17). Here, "as if he" introduces a description of Blaine that aligns him with Ifemelu's sense of humor, subtly foreshadowing their ensuing romantic connection.

6. Discussion

A linguistic approach to literary analysis may include a discussion of the various methods of characterisation to explore strategies of text creation for the reader. Characters play a pivotal role in creative writing, and a careful analysis can reveal that they serve more than just social roles assigned by the narrator (implied author). Characters often embody the ideological positions of the authors, who imbue them with qualities that reflect their own worldviews, their stance on the plot, and the thematic essence of the narrative. The American author William Faulkner, as cited by Houessou-Adin (2011, p. 63), highlights the essential function of characters in imparting meaning within a narrative. He asserts, "It begins with a character, usually, and once he stands up on his feet and begins to move, all I do is trot along behind him, recording what he says and does."

The comprehensive analysis of the novel through keyword and cluster analysis has revealed significant insights into the methods of characterisation. The "aboutness" of *Americanah* is initially investigated by systematically identifying the most salient keywords and analysing them through their concordance lines to ascertain their roles in developing the novel's thematic structure. Characters' names, such as Ifemelu, Obinze, Uju, Blaine, Dike, among others, are instrumental in crafting a fictional realm and underpinning the narrative. Additionally, place names like America, Nigeria, Lagos, Nsukka, Ibadan, Ikoyi, Abuja, and others, guide the reader's understanding of *Americanah* as a tale of immigration and its accompanying challenges. Profession terms such as "housegirl," "braiding," "braider," "babysitting," "waiter," "gateman," and more, also feature prominently and contribute to the portrayal of Ifemelu and other characters' sense of self and their interactions with others. For instance, Kosi, Obinze's wife, opts for a housegirl from the Benin Republic over Nigerian options, reflecting her belief, as stated in the novel, that Nigerian housegirls are unsuitable.

Regarding thematic signals, the words "Nigerpolitans," "halfcaste," "race," and "black" are particularly prominent in *Americanah*. As Mahlberg and McIntyre (2011) observe, these terms are "more reader-centered and relate to the effects that the text creates on the reader at a higher level of abstraction" (p. 211). They are crucial for they structure the novel's thematic organization and ideological positioning.

The term "Nigerpolitan" appears seven times within the novel. It denotes a group of young Nigerian returnees from abroad, symbolising the new identities these individuals assume upon their return to Nigeria for various reasons. Rather than adapting to their renewed life in Nigeria, they persistently lament their challenges and the process of acclimatisation. While Ifemelu is among this group, she adopts a critical view of the mentality of these Nigerians, as illustrated in her blog, *The Small Redemptions of Lagos*:

> "Lagos has always been undeniably itself, but this is not apparent at the Nigerpolitans Club meetings, where young returnees weekly bemoan how Lagos fails to resemble New York, as though Lagos had ever aspired to mimic New York. Full disclosure: I am one of them. Many of us have returned to capitalize on Nigeria's economy, to launch businesses, or to secure government contracts. Others have arrived with aspirations to transform the nation, yet we invariably find ourselves critiquing Nigeria, and even though our grievances are justified, I sometimes envision myself as an outsider exhorting: Go back where you came from!" (Chapter 12)

In this passage, Ifemelu's character is portrayed through her discerning perspective on Nigerian society and her critique of the 'Americanahs' who idealise life abroad.

Americanah also delves into the themes of race and black identity, underscored by the frequent references to "race" and "black." The term "race" is used 108 times, first appearing in chapter 37. Whether in Nigeria, England, or America, issues of race and black identity are pivotal to Adichie's narrative. Through her journey to America, Ifemelu discovers the stark realities of being "black" in a society dominated by white culture and inherent racism.

Additionally, Ifemelu's blog becomes a platform for articulating her insights on race, challenging the condescending attitudes of a white-dominated America. The complexity of race in America is such that Adichie, through Shan's character, portrays it as a delicate issue that must be approached with subtlety. As Shan explains to her editor, who desires her book to go beyond the topic of race: "…like race is a brew best served mild, mixed with other elements, lest it becomes too potent for white folk to ingest" (p. 415).

The analysis of clusters further elucidates the local functions of specific linguistic patterns, particularly in terms of character descriptions, interpersonal dynamics, and the construction of narrative realms. For instance, the

repeated use of the cluster "she did not want" is instrumental in depicting Ifemelu's assertive nature, her forthrightness, her strength, and her autonomy. The speech clusters also serve to distinguish characters, as exemplified by "I don't like," which portrays Ifemelu as a character grappling with feelings of alienation.

The "As if" clusters are notably effective in crafting the narrative's texture and offering insightful interpretations of characters' appearances and behaviours. For example, the use of "as if he" in phrases such as "It was as if he believed they shared a series of unspoken jokes…" and "…as if he could effortlessly float…" paints two distinct images: Blaine as a character imbued with humour and Obinze as one characterised by his elusive nature.

In summary, the linguistic analysis of *Americanah* through the lens of keywords and clusters yields significant revelations about the narrative's character development strategies. This examination not only uncovers the thematic and ideological fabric of the novel but also enhances our understanding of how Adichie uses language to craft her characters and their worlds.

7. Conclusion

This research paper aims to examine the methods of characterisation employed in Chimamanda Ngozi Adichie's novel *Americanah* through the analysis of keywords and n-grams. By applying a corpus linguistic approach, the objective is to gain valuable insights into the portrayal of characters and the creation of textual worlds within the narrative.

Through the analysis of keywords, we identified prominent terms that not only shape the thematic organisation of the novel but also provide an ideological positioning. Character names such as Ifemelu, Obinze, Uju, Blaine, and Dike emerged as major figures, underscoring their significance in the storyline. Additionally, place names like America, Nigeria, Lagos, and others revealed the novel's exploration of immigration and related issues. The inclusion of profession-related terms further contributed to the characterisation of Ifemelu and other female characters, offering glimpses into their self-representation and how they are perceived by others. Thematic signaling words such as "Nigerpolitans," "halfcaste," "race," and "black" play a pivotal role in conveying the central themes of the novel. The recurrence of these keywords not only reflect the narrative's focus on race and black identity but also showcase the experiences of characters in Nigeria, England, and America. Furthermore, Ifemelu's blog provides a platform for her to express her ideology and understanding of race, challenging mainstream perceptions and highlighting the complexity of racial dynamics. The analysis of clusters, particularly the "she did not want" and "I don't like" patterns, has revealed important character traits and the individualization of characters. Ifemelu's assertiveness, bluntness, and refusal to be dictated to are evident through these linguistic patterns, contributing to her characterisation as an independent and self-assured protagonist. Additionally, the "as if" cluster patterns created textual worlds, presenting characters like Blaine as humorous and Obinze as evasive, adding depth and complexity to their portrayals.

By conducting a corpus linguistic analysis of *Americanah*, we have deepened our understanding of the methods employed by Adichie in crafting and shaping her characters. The examination of keywords and n-grams has shed light on the thematic concerns, ideological positioning, and strategies of readerly text construction in the narrative fiction.

Overall, this research paper underscores the importance of linguistic analysis in literary studies. Employing corpus linguistics techniques has uncovered significant patterns and insights into the methods of characterisation in a literary text like *Americanah*. This study contributes to the broader understanding of Adichie's narrative techniques and provides a framework for future research exploring characterisation in literary works through corpus linguistic analysis.

References

Akongo, M. M. L., & Ngassaki, B. M. (2022). Characters in Chimamanda Ngozi Adichie's Americanah. *South Asian Research Journal of Arts, Language and Literature*, *4*(1), 1–9. https://doi.org/10.36346/sarjall.2022.v04i01.001

Anthony, L. (2022). *AntConc* (Version 4.2.0). Computer Software. Tokyo, Japan: Waseda University. Retrieved from https://www.laurenceanthony.net/software

Ayeomoni, N. (2003). The role of stylistics in literary studies. In L. Oyeleye & M. Olateju (Eds.), *Readings in Language and Literature*. Ile-Ife: Obafemi Awolowo University Press.

Brezina, V. (2018). *Statistics in Corpus Linguistics*. Cambridge. Cambridge University Press. https://doi.org/10.1017/9781316410899

Collins, L. (2015). How can semantic annotation help us to analyse the discourse of climate change in online

user comments? *Linguistik Online*, *70*(1), 43. https://doi.org/10.13092/lo.70.1743

Culpeper, J., & Demmen, J. (2015). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.

Egbert, J., Larson, T., & Biber, D. (2020). *Doing Linguistics with a Corpus: Methodological Considerations for the Everyday User*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108888790

Erwindriani, T. (2012). Scrooge's character development in Charles Dickens' A Christmas Carol. *Celt*, *12*(1), 1–100.

Francis, W. N. (1991). *Language corpora BC. Directions in Corpus Linguistics* (pp. 17–31). Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991.

Houessou-Adin, T. (2011). *Prolegomena to Critical Literary Analysis and Text Explication*. Cotonou: Hotepbook.

Larsson, T., Egbert, J., & Biber, D. (2022). On the status of statistical reporting versus linguistic description in corpus linguistics: A ten-year perspective. *Corpora*, *17*(1). https://doi.org/10.3366/cor.2022.0238

Leech, G., & Short, M. (2007). *Style in Fiction: A Linguistic Introduction to English Fictional Prose* (2nd ed.). London: Pearson Education Limited.

Mahlberg, M. (2007). Clusters, key clusters and local textual functions in Dickens. *Corpora*, *2*(1), 1–31. https://doi.org/10.3366/cor.2007.2.1.1

Mahlberg, M., & McIntyre, D. (2011). A case for corpus stylistics: Ian Fleming's Casino Royale. *English Text Construction*, *4*(2), 204–227. https://doi.org/10.1075/etc.4.2.03mah

Mahlberg, M., Smith, C., & Preston, S. (2013). Phrases in literary contexts Patterns and distributions of suspensions in Dickens's novels. *International Journal of Corpus Linguistics*, *18*(1), 35C56. https://doi.org/10.1075/ijcl.18.1.05mah

Moustafa, B. (2022). A comparative corpus stylistic analysis of thematisation and characterisation in Gordimer's My Son's Story and Coetzee's Disgrace. *Open Linguistics*, *8*(1), 46–64. https://doi.org/10.1515/opli-2020-0183

Paltridge, B. (2012). *Discourse Analysis: An Introduction Continuum* (2nd ed.). London & New York: Bloomsbury. https://doi.org/10.5040/9781350934290

Pearson, J. (1998). *Terms in Context*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.1

Sinclair, J. (1995). Corpus Typology: A Framework for Classification. In G. Melchers & B. Warren (Eds.), *Studies in Anglistics* (pp. 17–34). Stockhom: Almquist and Wiksell International.

White, P. R. R. (1998). *Telling media tales: The news story as rhetoric*. Unpublished PhD Thesis Department of Linguistics, University of Sydney. Retrieved April 12, 2022, from https://www.prrwhite.info/prrwhite,%201998,%20Telling%20Media%20Tales%20(unpublished%20PhD).pdf

Copyrights