Extraction of Key Motifs as a Preview from 2017 Nobel Prize Winning Novel, *'Never Let Me Go'*: An Interactive Word Cloud Study

Zafar Ullah[*]

Muhammad Uzair[**]

Arshad Mahmood[***]

Abstract

Word clouds manifest interactive visuals along with their statistical data. Thus knowledge discovery and aesthetic data visualization interlink to produce interactive word cloud which is an interesting, textual, statistical and visual data. This study aims to generate interactive word cloud—Cirrus—on the basis of statistical data to preview text of the novel for readers. So cirrus tool is selected from Voyant open access tools to produce interactive statistical word cloud. Then the generated word cloud and statistical data are analyzed with mixed method and its analysis draws insight from Rakesh Aggrawal's Knowledge Discovery Theory which seeks innovative and interesting knowledge patterns. This thematic word cloud verifies already known themes and discovers innovative interesting themes. Current study reveals that all mentioned key themes can be easily extracted from a voluminous novel with the help of Cirrus tool. Key motifs have been presented in the word cloud for the readers. On the other hand, unwritten themes can't be extracted through machine learning tools, rather it is the task of human cognition. Primarily, this novel based study reveals names of chief characters, for instance "Tommy (496)," "Ruth (455)" and "I (Kathy) (355)." Furthermore, motifs of nostalgic memories with word "remember (143)," "thought (126)" about "Hailsham (203)," "carer (74)," "sex (80)," sex "lectures (8)" have been discovered as a preview. Previewing technique prepares reader's mind and gives an epigrammatic digital view of the text. The visual themes as knowledgeable word cloud leave an indelible mark on the slate of memory.

Keywords*:* Word cloud/Cirrus, Voyant, text visualization, topic modelling, preview

Introduction

Technology facilitates human learning to extract knowledge patterns instantly. Technological gadgets save humans from laborious tasks and they produce accurate results even before actual reading which is called previewing. Previewing is a skimming technique to get a brief idea about key themes and major characters of a book or text without its

---

[*] Lecturer in English, National University of Modern Languages, Islamabad.
   Email: zafarullah@numl.edu.pk
[**] Director, Academics, National University of Modern Languages, Islamabad.
[***] Professor/Head, Department of English, National University of Modern Languages, Islamabad.

through reading. Current study aims to extract the word cloud as a preview of the novel, *'Never Let Me Go'* which won Nobel Prize in literature in 2017.

Cirrus or word cloud exhibits visual and statistical representations of central themes of the uploaded texts. The word cloud is created on the basis of statistical weight, hence, the most occurring words occupy more prominent place and bigger font size in the word cloud frame. Word clouds statistically quantify qualitative and linguistic data and change them into condensed and refined information for the viewers.

Technically, word clouds are generated with Java, a computer programming language, to extract central themes. Wordle tool accepts Unicode text to create word cloud. At that juncture, stop words (prepositions, articles, conjunctions etc) are cleansed from data. Afterwards weight, statistical occurrence of any word, is assigned to words. Subsequently, a rectangular layout is usually designed, however, Wordle tool also gives other layout options. At that point, playing field or total covered area is determined and the selected words are changed into Java 2D shape. Lastly, statistical weight determines positioning and size of words (Steele & Iliinsky, 2010).

As a pioneer of word clouds methodology, Milgram used word clouds to visualize word frequency in 1970s (Milgram, 1976). Similarly, word clouds were also used in 2004 when Millen and his fellows worked on book marking project named "dogear". In this project, tag content tool was used to produce tag clouds in order to display key themes of the selected documents (Millen, Feinberg & Kerr, 2006). The use of Wordle tool was followed by other tools to design word clouds.

Afterwards, Matt Jones, a blogger, started to post word clouds in different calligraphic forms, popularizing them in the world. The word clouds produce statistical information along with having a pleasing effect. From this premise, word clouds are employed for playful learning and teaching purposes. Now a days, word clouds are frequently shown in workshops and academic materials to visualize key themes. Extending this concept, generation of word cloud/Cirrus is a part of text mining/analytics which is an emerging field in Digital Humanities.

As far as scope of word clouds is concerned, data visuals have stronger and lasting effect on cognition, memory and aesthetic sense than the mere delineation of random list of words. Moreover, implicit logic and visual learning techniques perform their role in the word cloud because these visuals are embedded in statistical data. Novelty of this study is that Cirrus tool can easily extract most of the key themes which can be used as a preview of any linear or nonlinear text.

Study in hand concentrates on the issue that most of the students remain unaware of uniformed knowledge patterns, central ideas and major characters of the text. Basically previewing technique enables the viewers to grasp brief ideas even prior to complete reading of the text. So to address these academic needs, Rockwell and Sinclair (2015)

introduced Cirrus tool as part of Voyant tools. This tool produces word clouds and their statistical weight interactively. Students require to know briefly about basic concepts and characters prior to full text reading. So word clouds as a previewing technique are needed for learners and teachers. Word clouds being a previewing tool empower "digital natives" (Prensky, 2001) to unveil key themes interestingly and aesthetically. There is a big problem of dearth of time and numerous information sources (Godwin-Jones, 2006). In order to cope with this situation, Cirrus/word clouds are apt to condense the most occurring information in the form of word clouds and statistical data. Thus, qualitative and quantitative data strengthen results of each other.

This study aims to preview extracted prominent motifs in the form of word clouds which fall in the domains of Computational Linguistics and Natural Language Processing (NLP). To accomplish this aim, this research raises the following research question:

How does Cirrus tool reveals major themes and characters from the novel, *'Never Let Me Go'* during preview phase?

Novel *'Never Let Me Go'* is a dystopian novel focusing on lives of clones who are nurtured to donate their organs one by one to the donors who used to bear their expenses in Hailsham and Cottages. After two or three donations, lives of clones end miserably. Kathy tells her story along with her fellows, Ruth and Tommy. In their early life, they stay at Hailsham as students, afterwards, they move to Cottages and lastly, in the third phase, they become carer. Having been grown up, they understand drastic future, transient life and exploiting purpose of their lives.

Literature Review

Word clouds/Cirrus were used as a previewing technique for voluminous textual data. After getting the basic idea about key motifs and characters through word clouds, deeper level of textual and thematic analysis was done. Digital tools such as Wordle, Tagxedo, and Word Cloud Explorer also made word clouds which functioned on the basis of certain algorithms. Several literary studies were conducted with digital tools to exhibit key themes with word clouds. This segment reviewed some previous relevant studies.

Novel 'Romance of Flamenca' and its English translation were studied to explore major themes in the form of word clouds with the help of toolkit of Interactive Text Mining Suite. In this paper, corpus and digital humanities were combined and data regarding distant and close reading were visualized (Scrivener, & Davis, 2017). There was a similarity of genre of novel and word cloud with current research but analysing tool was different. Voyant tools of current study also amalgamated corpus and digital humanities to explore new word cloud patterns in the knowledge discovery process.

Some other circular data presenting digital tools also displayed significant themes. Lohmann, Heimerl, Bopp, Burch and Ertl (2015) claimed that ConcentriCloud was akin to

RadCloud approach. ConcentriCloud was better than other word cloud tools because it laid three layers of a circle where the inward circle showed the most common words of various selected documents. Suppose if there was no common word, the central circle would be empty. Empirically, 7 novels of Harry Potter were analysed with CorecentriCloud. These words clouds could also be customized and coloured according to needs and choices. This study had a deficiency of interactivity which was filled by Voyant Cirrus tool for showing interactive word clouds in a rectangular shape with its occurrence number to merge word cloud and statistical data. Some other tools, for instance Parallel Tag Clouds, ManyEyes and DocuBurst, showed key concepts through digital extraction process. So every document was shown as a bag of words in digital data.

A similar study was carried out by Jannidis (Burrows, 2002, p. 267) in which Voyant tools were applied on 350 Canonical German novels. In the study, unique and interesting knowledge patterns were derived in the form of data visualization, statistics etc. One of the limitations of this study, due to its magnitude, was its inability to specify each theme pertaining to each novel for its readers. Another study was conducted on Jane Austen's all novels with Voyant tools (Sinclair & Rockwell, 2015). It also dealt with big textual data. Likewise, word clouds were generated from 1500 apocalyptic fictions by Yeates (2013). Owing to usefulness and result-oriented nature of Voyant tools, a number of studies were taken up: for instance, Shakespeare's 37 plays, Jane Austen's complete work, George Eliot's novel 'Mill on the Floss,' Harry Potter series and novels of 'Sherlock Holmes' were mined with Voyant tools (Sinclair, & Rockwell, 2015).

Another toolkit, namely Wordstorms tool, generated word clouds to compare themes of data. Just one cursory glance at word cloud generated through this tool enabled viewers to discriminate themes of different documents. High dimensional data were placed in a high dimensional space to visualize corpus of a document. Principally, a good word cloud must be arranged coherently until no significant element is eliminated and no function word (preposition, articles, conjunction) is incorporated (Castella & Sutton, 2014). Voyant stopwords tool automatically eliminates function words to refine the word cloud.

The idea of word clouds was further spotlighted during the general elections in the US when, in 2008, Obama and McCain's presidential speeches were text mined in order to generate and examine the key themes used by both contestants (GitHub, n.d.). These speech texts were small without statistical data and their political themes were self-evident to some extent but this research paper analysed a long Nobel Prize winning novel, *'Never Let Me Go'* and generated thematic as well as statistical word clouds as a previewing technique.

Some other studies extended the focus of word cloud studies beyond theme extraction. Another research task was mining emotion words from Shakespeare's plays 'Hamlet' and 'As you Like It,' Shelley's 'Frankenstein' and other fairytales. Subsequently,

timeline of their emotion words were mutually compared (Muhammad, 2012). Emotions were vital themes of Shakespearean tragedies.

Topic modelling, an unsupervised learning method in machine learning, had worked successfully in the domain of text mining to extract significant themes. It is an unbiased academic method to bring key motifs to limelight. It was based on algorithm of Latent Dirichlet Allocation (LDA). So interesting themes were extracted from French dramas and those themes were associated with epistemological domain of philosophy (Schoch & Schoech, 2012).

Primarily word clouds were meant for topic modelling from any corpus. Therefore, topic modelling technique was employed to extract key themes from the corpus of 3346 British, American and Irish fictional works. Function words were stopped manually to discover major themes. Key themes pertaining to word "enemy" were derived as word clouds from Jocker's work (Jockers & Mimno, 2013). Deficiency of this tool was made up in Cirrus tool that automatically stops function words. Moreover, stop words can be disabled or edited according to the research needs.

The higher the frequency of certain words in a corpus, the bigger size of words would be in the word cloud. Frequency and significant words were not in harmony with a word cloud. One more deficiency of word cloud was that they didn't provide context of word and same forms of a stem word didn't group together in the word cloud. To minimize ambiguity, context must be read along with the analysis of word clouds (Ramsden & Bate, 2008). Mere word clouds didn't create deep sense, so its context was integral to avoid obscurity. Contexts tool provided context for clarity of visions. Collaboration of contexts tool with word cloud made previewing an error free knowledge extraction activity. This option was not incorporated in current study.

Support of Context tool along with word clouds refined true semantic comprehension. In another study, Google Books Corpus N grams data were presented in word clouds with other visuals (Muhammad, 2012). Mere word clouds could be thematically misleading, for example, word "like" was used as simile and adoration. So to avoid such thematic and semantic opacities, Voyant tool might use Contexts tool to disambiguate word sense.

With flow of time, innovativeness and new technology replaced old tools. Thereupon, previous static word clouds were transformed into interactive word clouds. Wordle was a static word cloud generating tool while following several interactive word clouds had been introduced, for instance ManiWordle (Koh et al., 2010 as cited in Seyfert & Viola, 2017) showed rotating and moving interactive word clouds. Later, WordlePlus (Jo et al., 2015 as cited in Seyfert & Viola, 2017) added words pop and grouping in interactive word clouds. SparkClouds (Lee et al., 2010 as cited in Seyfert & Viola, 2017) showed sparkline below every word. Dynamic word clouds were used to observe time

varying data and word frequency (Seyfert & Viola, 2017). Cirrus tool was also in harmony with above-mentioned tools in terms of their interactivity.

After embedding interactivity in word clouds, another progression was induction of semantic word clouds with ReCloud which were rooted in its context. Voyant used separate context tool to clarify the ambiguity but ReCloud integrated context within word clouds. Customers' reviews about Amazon and Yelp were analysed quantitatively and qualitatively through semantic word clouds to find satisfaction level of customers (Wang, Zhao, Guo, North, & Ramakrishnan, 2014). Semantic word clouds had more significance than mere word clouds. In this research, LDA algorithm model was applied to comprehend contextual meanings. These word clouds were also being used in business intelligence, media studies and other fields of life. Showbiz contents were studied to extract movie themes in the form of word clouds through TopicVec tool (Nguyen, Chang, Hui, 2011).

There were multiple uses of word clouds in the different domains of academia. One use of word clouds was to summarize the entire corpus data. Therefore, word clouds were positively advantageous for summarizing various interviews (McNaught & Lam, 2010). Same approach could be implemented on any linear or nonlinear texts for summarizing purposes.

Word cloud was a learning tool which could be manipulated for numerous purposes. Word clouds facilitated teachers and learners to learn and teach all language skills (Hayes, 2008). Kolb's four learning styles: accommodation, assimilation, convergence, and divergence were studied with the help of word clouds which enhanced interest, motivation and knowledge of learners. Consequently, word clouds developed learning style, teaching pedagogy, learning techniques and workplace skills for learners (Miley & Read, 2011). Word clouds displayed intellectual information and aesthetic pleasure simultaneously. They were appropriate in the previewing process, learning vocabulary, brain storming and theme learning. It was an advanced step towards previewing, self-paced learning, maintaining interest and enhancing motivation level of ESL learners.

Clement, Plaisant and Vuillemot (2008) reported that PosViz produced visuals to understand use of words. Assigning different colours to different parts of speech was done to understand them. It was deficient because it was an old static word cloud without statistical data while the target research took the advanced form of word clouds and changed static data into an interactive word cloud. Moreover, statistical word cloud was displayed with Cirrus tool which showed 25-500 themes.

Methods

Research onion comprised philosophical dimensions, approach, strategies, time horizons, data collection and its evaluation in the light of theoretical framework (Saunders, Lewis, & Thornhill, 2012).

Philosophical stance leads to research prototype, hence, epistemology and positivism were applied on this study. Epistemology and positivism led to experimental research design with machine learning tools, deep learning and artificial intelligence (The Writepass Journal, 2012, June 5). To accomplish these ends, mixed method approach was opted. Mixed method means the "combination of qualitative and quantitative approaches" (Durant, 2004, p.10). In this study design, corpus of the novel, statistical data and visuals were merged. Therefore, it intersected domains of literature, computer science and data visualization. In other words, the mixed method amalgamated hard sciences and social sciences. Its reliability and validity produced same results in the form of statistics, visual data as word cloud and corpus. Consequently, mixed method crosschecked the study qualitatively and quantitatively.

This research followed Rakesh Aggrawal's Knowledge Discovery Theory in Data Mining (KDD) to extract unfamiliar, interesting and beneficial knowledge patterns (Cabena, Hadjinian, Stadler, Verhees, Zanasi, 1998). It encompassed several disciplines, for instance, hermeneutics, deep learning, artificial intelligence and machine learning (Fayyad, Shapiro, Smyth, & Uthurusamy, 1996). Integration of technology, pedagogy and content was named as TPACK model (Koehler & Mishra, 2009) and this model rationalized the use of technology for learning purposes.

Voyant tools were applied in this research and they were created by Geoffrey Rockwell and Stefan Sinclair in 2003, later improved by them with better and precise results. The Voyant has 51 tools under broader categories of corpus, grid, visualization, document and others. Thereupon, they were further extended into various data visualization forms. Cirrus is another name for word clouds in the broader category of visualization. This tool showed interactive word cloud to extract 25 to 500 key themes. Editable stopwords tool stopped function words (prepositions, articles, conjunctions, and numbers) to refine the results. As far as academic worth of Cirrus tool is concerned, Graham, Milligan & Weingart (2013) declared it an "ideal" tool for classroom and academic purposes.

Technically, text data and its classification was termed as "a bag of words" (Aggarwal & Zhai, 2012, p.3). So text data or research document for this study was a 2017 Nobel Prize winning novel, *'Never Let Me Go'* (2005) written by Kazuo Ishiguro, a Japanese origin British novelist.

The research strategy was that pdf file of the novel *'Never Let me Go'* was uploaded on Voyant tool and it revealed its data in the five panels. As current study focused on Cirrus, so 105 words Cirrus/word cloud was generated for topic modelling. The word cloud was exported as a PNG image because it gave clearer view. Later, it was copied and its visual and statistical data were evaluated.

There was a limitation that statistical data were only visible by placing cursor on the digital screen. When Cirrus as a static image was exported on the word file, it didn't

show any statistical value interactively. To overcome this deficiency, Summary tool in Voyant was used to take statistical data of first 59 the most frequent words and it was pasted below Cirrus. We claimed to present interactive word cloud on Cirrus panel while PNG image on the paper hampered interactivity of the word cloud. Therefore, word cloud and complete corpus can be accessed online just by CTRL and click on any yellow coloured theme in the word file.

Generated word cloud/Cirrus and its statistical data were valid and reliable because every time, the same Cirrus and word frequency were generated with Cirrus tool. Furthermore, word cloud is considered as a valid and appropriate tool for conducting research (McNaught & Lam, 2010).

Results and Discussion

The problem of previewing themes from a voluminous data is a frequent one. This study visualizes Kazuo Ishiguro's 2017 Nobel Prize winning novel *'Never Let Me Go'* in a word cloud to present its major themes as a previewing technique. It generates word cloud with Cirrus tool. It produces interactive word cloud and a number of themes which can be searched from 25 to 500. Cirrus tool in Voyant performs better than Wordle tool because Cirrus tool shows statistical number of each theme and character just by placing the cursor on it while Wordle tool is unable to show statistical data. Themes are extracted on the basis of statistical weight, so the following word cloud shows topic modelling, character and theme extraction.

Cirrus and Statistical Data of the Novel *'Never Let Me Go'*

said (584); just (509); like (500); tommy (496); ruth (455); i'd (355); time (342); way (302); didn't (287); it's (286); got (262); really (249); we'd (230); say (226); know (224); think (217); wasn't (215); miss (210); don't (205); hailsham (203); came (196); that's (192); things (189); going (181); went (181); i'm (178); maybe (177); right (172); she'd (172);

come (157); long (143); remember (143); little (138); kept (129); looking (126); people (126); thought (126); day (125); away (123); look (123); looked (123); room (120); sure (1 18); he'd (116); talking(114); talk (111); told (109); lot (103); bit (102); doing (101); wan ted (101); knew (99); suppose (98); years (98); good(97); i've (97); want (96); saw (93); tell (93)

Dialogic Theme: The most occurring statistical theme of the novel is "said (584)" because the novel characters directly talk to each other. With dialogues, plot of the novel progresses and all characters evolve and understand situations minutely. Some other relevant themes are "told (109)," "talking (114)" and "talk (111)." They reveal that most of the time characters directly talk and express their intimate and heartfelt emotions particularly regarding their childhood, boyhood or girlhood at "Hailsham (203)" and "Cottages (68)" which are geographical settings and rural settlements in England.

First Person Narrative Style: The novel is written in the first person narrative as word cloud and statistical data express, for instance, "I'd (355)," "I'm (178) and "we'd (230)." In totality, first person pronouns are used 763 times in first person narrative. This statistical and visual data suggest that Kathy, the narrator, dominates all characters. Heartfelt feelings are more precisely expressed in the first person narrative technique.

Joined Words as Distinct Style: Firstly, the researcher considered joined words as a spelling error but later found the joined words to exhibit the distinct style, for example "some response," "fromtrash," "feltI," "toremeber," "fromtrash." Probably, this style is adopted to draw attention of the readers to express distinctness.

Characters: Major characters of the novel are "Tommy (496)," "Ruth (455)" and first person pronouns "I'd (355)" and "I'm (178)" refer to Kathy or "Kath (82)." Ruth, Tommy and Kathy spend their childhood and adolescent period in Hailsham and Cottages under the guardianship of Emily, "Madame (78)" and Lucy who teaches them how to lead their lives and how to cope with the challenges faced by a "special human being." Another couple in the novel is "Chrissie (84)" and "Rodney (80)" who live in the cottages. Ruth, Kathy, Tommy go to a trip to Norfolk with Chrissie and Rodney to test theory of "gallery." The aforementioned characters are extracted due to their excessive occurrence. When statistical weight of any character's name increases, its dominance and size increases in the Cirrus. In this corpus, "I" as a character occurs frequently.

Special Boys and Girls: Children of Hailsham are called "special." They know about themselves and they say, "We were different from our guardians and also from the normal people outside" (Ishiguro, 2005, p.34). They admit, "We all knew … none of us could have babies" (Ishiguro, 2005, p.36). These children realise their shortcoming of derogatory comments, for instance "You were less than human" (Ishiguro, 2005, p.125); "You poor

creatures" (Ishiguro, 2005, p.129). A sense of deprivation and a shadow of imminent death loom on the lives of these clones.

Carer as a Career: Theme of "carer" is found 74 times and most of the students from Hailsham have to be carer in their future life as it is mentioned in the novel, "More and more students were going off to be carers" (Ishiguro, 2005, p.94). So Ruth and Kathy also serve as carer for Tommy; and Laura also serves as a carer for other donors. These carers have to sacrifice their organs gradually. This act is rather extreme human exploitation and humiliation. As the novel raises voice for the carers, it wins the Nobel Prize for Kazuo Ishiguro.

Term "Miss" for Hailsham Guardians: Theme of "Miss (210)" is very prominent in the novel and every time it refers to the names of "guardians (63)" or teachers, for instance, Miss "Emily (70)", head of Hailsham, Miss Geraldine and Miss Lucy who also serve in Hailsham. Miss "Lucy (80)" is very considerate and idealistic to show students the full picture of life including rosy and ugly sides. She thinks it is wrong to show a deceitful rosy life to students. Later on, she leaves Hailsham due to her conflicting idealistic thoughts while Miss Emily wants to nurture some myths of "gallery (50)" and deferral (13)" in the minds of students. Thus, it was a conflict between illusion and bitter reality.

Moreover, the theme of "miss" also reveals that Hailsham students don't call their teachers by their first names contrary to English norm. It shows the ideology of Hailsham that creates sophisticated and respectful children. The world considers carers "less than human" but ethically they are more or equal to normal humans in their mannerism. Simultaneously, Miss Emily and Madame "fought that battle for many years" for equal rights of deprived clone children.

Nostalgia: One of the major themes of the novel is nostalgia and its statistical evidence is "remember (143)" that characters think about innocent and glorious past in Hailsham. The wave of nostalgia prevails throughout the novel as Ruth, Tommy and Kathy miss one another and they think about their guardians especially Miss Lucy.

In fact, love triangle of Ruth, Tommy and Kathy is the strongest bond. Once Ruth and Tommy tie the knot and later Kathy joins Tommy but in the end, this bond shatters. Even after this segregation, Tommy is visible to Kathy in a fantasy and this is told in these words, "a tiny figure would appear on the horizon across the field and gradually got larger until I'd see, it was Tommy, and he'd wave, may be even call" (Ishiguro, 2005, p.137). In addition, theme of "past" is mentioned 63 times and this indicates that characters of the novel mostly talk about their past events and cherish those pleasant memories. One rationale can be that every coming moment is getting worse than the past, so human beings seek pleasure and refuge in the old enjoyable memories. Above all, childhood is a carefree

span of life while adulthood brings drastic realities of life. Being a weak creature, human beings intend to live in the murky world of illusions rather than in bony arms of reality.

Theme of "memory" and "memories" are found 16 and 12 times respectively in the corpus. A donor said, "… memories even your most precious ones, fade surprisingly quickly." Kath says, "I lost Ruth, then I lost Tommy, but I won't lose my memories of them. I lost Hailsham too" (Ishiguro, 2005, p. 136). Memories about past are never washed-out from mind slates of all characters. Here nostalgic feelings are so powerful and sometimes it seems that whole novel is weaved with the fabric of past memories and verification of previous theories and ideas because the adulthood leads to find proofs of childhood assumptions.

Thought: Theme of "thought (126)" is also prevailing in the novel since all characters establish theories and later they strive to testify their theories. Ruth, Tommy and Kathy think about one another and all of them ponder about Hailsham and Cottages too. They think about "donors (73)," "donations (52)" and their career as "carer (74)" for which they are nurtured.

Theories and Testing of Theories: Theme of "theory" is found 28 times in the corpus of novel. Students hypothesize different theories and try to test them in their future lives. One theory of Hailsham students is that "Madame is afraid of us." Later, they test the theory and find that she is fearful of spiders and students alike. Another theory is "Tommy's big gallery theory" which is tested by visiting to Madame's house. It reveals that students' art material is collected just to show to authorities, donors and for introspection of students. Madame said, "… my gallery help in telling which of you were really in love?" "Your art will reveal your inner selves" (Ishiguro, 2005, p.120). "We took away your art because we thought it would reveal your souls." (Ishiguro, 2005, p. 123). Consequently, their teachers use art for the analysis of students.

Sex Teaching and Sex: Theme of "lecture (8)" about safe sex and its effects occur 8 times in this corpus and two phrases express sex lectures thus: "proper lectures about sex" and "a lot of sex lectures." First time the students are taught about sex by Miss Emily who teaches them sexual relations with the help of a skeleton. She says, "Sex was a beautiful gift" (Ishiguro, 2005, p. 46). She also guides students about safe sex and to avoid sexual diseases. She teaches them, "We'd have to be very careful to avoid diseases when we had sex…" (Ishiguro, 2005, p. 40). She also teaches them, "… be careful who we had sex with." "Sex affects emotions … We had to be extremely careful about having sex in the outside world" (Ishiguro, 2005, p. 41). Safe sex can save them from fatal diseases because every carer must be healthy. The carer who donates body organs must be free from sexual diseases like AIDS, syphilis etc.

Theme of "sex" is found 77 times in the novel. Sex in Hailsham is a prohibited act while sex is implicitly permitted in Cottages because in the former place, children live and in the latter place, adults live. Chapter 8 indicates first time that Mr. Jack sees Jenny and Rob having sex. Later Tommy and Ruth have sex as a couple. Likewise, Kathy has sex with different males, for instance, Keffers, Hughie and Oliver just to understand them and to experience sex; hence, she tells "I'd had a few one nighters" (Ishiguro, 2005, p. 59). Once "Harry … hinted a few times he'd like to have sex with me" (Kathy) (Ishiguro, 2005, p. 47). Later, when Kathy becomes Tommy's carer, she openly confesses having sex with Tommy thus: "For a start, Tommy and I finally started having sex" (Ishiguro, 2005, p. 113). These incidents refer to several sexual encounters for different purposes.

Kazuo Ishiguro mentions sex without naming genitals and he euphemises sex scenes with his artistic quill. He mentions the relationship of Tommy and Kathy thus: "I went over, sat on the edge of the bed, and slid a hand under his T-shirt. Pretty soon I was down around his stuff, and thought it took a while for him to get hard" (Ishiguro, 2005, p. 113). Ishiguro euphemistically uses word "stuff" for genitals. Furthermore, he mentions "porn (9)" mags in Hailsham and how old Keffers infuriates whenever he finds such magazines from any student's room.

Love: In the whole corpus of the novel, the word "love" is used 19 times but true and persistent platonic love is extinct among all characters. Tommy and Ruth; later, Tommy and Kathy; and Chrissie and Rodney become couples but their couples break up after a certain time. Another very good strategy to keep carers united is that if a couple lives for 3 years together, Hailsham staff will give money to the couples but in reality, it is just Miss Emily's technique to keep the couples united and sexually safe. Sometimes materialistic gains keep them as couples for some years and later they become accustomed for each other's company. In absence of children, pseudo financial reward scheme is introduced to keep them united and safe.

Tears: The theme of "tears" is used 14 times in the novel. The guardian Emily weeps for Kathy when she is listening to the tragic song "O baby, never let me go…" Emily weeps that Kathy is unable to become a mother while she aspires so. So pains of being issueless has been expressed.

Painting: Theme of "paint" is found 17 times in the novel. All students paint in Hailsham and later their paintings are carried away by Madame for her gallery. Tommy is not good at drawing yet he persists to improve his artistic skill. In fact, guardians of carers ask them to paint because art expresses innerself of the painter.

Moral Lessons: Students are not allowed to smoke and lectures are delivered against smoking cigarettes. Moreover, they are not permitted to hug each other or to close the door

completely when they are inside. Besides, there is mentioning of "donor (73)," "donation (29)" and "charity (1)" in the novel.

Negation: Theme of negation is also found in the words of "wasn't (215)," "don't (205)," "wouldn't (63)." There is a theme of positivity expressed in a negative form, for example, "I'm not at fault" means that the speaker is free from shortcomings. Apparent negative word refers to the positive semantic shade.

Know Vs Suppose: Theme of "knew (99)" and "suppose (98)" reveal the pattern that knowledge and supposition almost go side by side in the novel but theme of knowledge occurs more than supposition.

Literary Allusions: There are references to "Shakespeare (2)," Tolstoy's world renowned novel "War and Peace (1)" and George Eliot's last novel "Daniel Deronda (4)." So literature within literature and literary allusions make it a sublime work.

To summarize, statistical occurrence rightly extracts key themes and characters with text mining process. Textual references validate theme extraction and topic modelling from the novel. Word clouds not only enhance thematic and character knowledge in the previewing phase but also increase interest and aesthetic level of the readers.

Statistical Limitation to Extract Implied Themes of Cloned Beings and Dystopia: Voyant tools are useful for extraction of written themes but obviously digital tools can't extract hidden and implied themes which human mind and cognitive abilities can grasp. So Cirrus tool extracts aforementioned themes but fails to explore least mentioned and implied theme of "cloned (2)" beings and their troubles in English society. Another major limitation is found that word clouds are generated on the basis of statistics that the word enlarges owing to its frequent occurrence in the corpus. It is applicable on many corpora but this statistical technique is deficient to some extent in the novel *'Never Let Me Go'* because it centres upon lives of cloned beings without repeating any sexist words for them. Kazuo Ishiguro focuses on carer role, peculiar diseases, psychological state of mind, and tragedy of not becoming parents and perfect partners. This is his stylistic quality that he highlights their problems without mentioning derogatory names of cloned beings, rather he uses euphemised words such as "poor creature," "less than fully human." Computational theme extraction system in the form of word clouds is not perfectly capable to extract the hidden themes of the novel because pragmatics is the speciality of human cognition. Computational tools only grasp peripheral written themes which can be counted statistically from the corpus.

Dystopia, contrary to utopia, is an undesirable society in the form of Hailsham. Dystopian theme is prevalent throughout the novel without naming the word "dystopia." Therefore, statistical data based word cloud is unable to exhibit "dystopia" as a key theme.

Another limitation is that after publication of this novel, various critics point out its genre while the novelist never mentions its type in his work.

Conclusion

As a major finding, this study brings to the light all major themes of the novel which are extracted aesthetically in the form of word clouds except unmentioned and implied themes of cloned beings and dystopia. In fact, word clouds are generated on the basis of statistical occurrence. So Cirrus tool is fully successful in extracting delineated themes, for instance, nostalgic feelings, sex, donor, donation, carer, names of main characters. At the same time, this word cloud/Cirrus facilitates viewers to extract key ideas without thorough reading of the novel.

As far as academic implications are concerned, word clouds can be used for previewing, key word extraction, text analysis, comparison, prediction, vocabulary building, vocabulary teaching, website evaluation, exam revision, speech analysis, summary output, inclination of an author and bias analysis. However, there is a limitation that the word clouds don't present multi word phrases together. It counts each hyphenated part of a word as a separate word.

Future research should be conducted on the use of word clouds for vocabulary teaching, exploration of code mixed language in any multilingual corpus. Extending the same word clouds, opinion clouds and semantic clouds should also be analysed in future research projects.

References

Aggarwal, C. C., & Zhai, C. (2012). An introduction to text mining. In *Mining text data* (pp. 1-10). NY: Springer.

Burrows, J. (2002). Delta: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, *17*(3), 267-287.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering Data Mining: From Concept to Implementation*. Upper Saddle River, NJ: Prentice Hall.

Castellà, Q., & Sutton, C. (2014). Word Storms: Multiples of Word Clouds for Visual Comparison of Documents. *International World Wide Web Conference Committee (IW3C2)*. Retrieved from homepages.inf.ed.ac.uk/csutton/publications/castella14word.pdf

Clement, T., Plaisant, C., & Vuillemot, R. (2008). The story of one: Humanity scholarship with visualization and text analysis. Retrieved from http://www.cs.umd.edu/hcil/trs/2008-33/2008-33.pdf

Durant, G. B. (2004). A typology of research methods within the social sciences. *NCRM Working Paper*, 1-22. Retrieved from http://eprints.ncrm.ac.uk/115/

Fayyad, U. M., Shapiro, G. P., Smyth, P., & Uthurusamy, R. (Eds.). (1996). *Advances in knowledge discovery and data mining*. Menlo Park, Canada: AAAI Press.

Godwin-Jones, R. (2006). Tag clouds in the blogosphere: Electronic literacy and social networking. *Language Learning & Technology, 10*(2), 8-15.

GitHub. (n.d.). Semantic word cloud visualization - Description. Retrieved January 23, 2018, from http://wordcloud.cs.arizona.edu/description.html

Graham, S., Milligan, I., & Weingart, S. (2013). Voyant Tools | The Historian's Macroscope: Big Digital History. Retrieved August 11, 2017 from http://www.themacroscope.org/?page_id=639

Hayes, S. (2008). Toolkit: Wordle. *Voices from the Middle, 16*(2), 66-68.

Ishiguro, K. (2005). *Never let me go* (1st Ed.). London, United Kingdom: Faber and Faber.

Jockers, M. L., & Mimno, D. (2013). Significant themes in 19th-century literature. *Poetics*, *41*(6), 750-769.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60-70.

Lohmann, S., Heimerl, F., Bopp, F., Burch, M., & Ertl, T. (2015). ConcentriCloud: Word Cloud Visualization for Multiple Text Documents. *19th International Conference on Information Visualisation*, 114-120. Retrieved from https://puma.ub.uni stuttgart.de/bibtex/2fb567abbd01d10cb6efc8a59105f8114/visus

McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *The Qualitative Report Volume*, *15*(3), 630-643.

Miley, F., & Read, A. (2011). Using word clouds to develop proactive learners. *Journal of the Scholarship of Teaching and Learning*, *11*(2), 91-110.

Milgram, S. (1976). Psychological maps of Paris. In *The individual in social world* (2nd ed., pp. 104-124).

Mohammad, S. M. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, *53*(4), 730-741.

Nejad, A. M., Nariman, Z., & Hosseinkhan, N. (2013). *Next generation sequencing and sequence assembly* [pdf]. doi:10.1007/978-1-4614-7726-6

Nguyen, T. T., Chang, K., & Hui, S. C. (2011). Word cloud model for text categorization. In *11th IEEE International Conference on Data Mining* (pp. 487-496). Singapore: IEEE.

Prensky, M. (2001). Digital Natives, Digital Immigrants. *On the horizon*, *9*(5), 1-6.

Ramsden, A., & Bate, A. (2008). Using word clouds in teaching and learning. Retrieved January 24, 2018 from http://opus.bath.ac.uk/474/1/using%2520word%2520clouds%2520in%2520teaching%2520and%2520learning.pdf

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students* (6th ed.). Pearson Education Limited.

Seyfert, M., & Viola, I. (2017). Dynamic word clouds. *Proceedings of the 33rd Spring Conference on Computer Graphics - SCCG '17*. doi:10.1145/3154353.3154358

Sinclair, S., & Rockwell, G. (2015). Principles of Voyant Tools | Voyant Tools Documentation. Retrieved May 29, 2017, from http://docs.voyant-tools.org/CONTEXT/PRINCIPLES/

Sinclair, S., & Rockwell, G. (2015, June 29). Voyant Tools Documentation. Retrieved March 1, 2018, from http://docs.voyant-tools.org/?p=0&scripto_action=transcribe&scripto_doc_id=944&scripto_doc_page_id=945

Schöch, C., & Schoech, C. (2012). Topic modeling genre: An exploration of French classical and enlightenment drama. *Digital Humanities Quarterly*, *11*(2). Retrieved from http://digitalhumanities.org/dhq/vol/11/2/000291/000291.html

Scrivener, O., & Davis, J. (2017). Interactive Text mining suite: Data visualization for literary studies. Retrieved January 27, 2018, from http://ceur-ws.org/Vol-1786/scrivner.pdf

Steele, J., & Iliinsky, N. P. (2010). *Beautiful visualization*. Beijing: O'Reilly.

The Writepass Journal. (2012, June 5). How to write a dissertation: Methodology - the writepass journal: *The WritePass journal*. Retrieved August 13, 2017 from https://writepass.com/journal/2012/06/how-to-write-a-dissertation-methodology/

Wang, J., Zhao, J., Guo, S., North, C., & Ramakrishnan, N. (2014). *Graphics Interface Conference.* Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.9008&rep=rep1&type=pdf

Yeates, R. (2013, May 2). Voyant Tools | Post-Apocalyptic Cities. Retrieved March 25, 2017, from https://postapocalypticcities.wordpress.com/2013/05/02/voyanttools/