A Comparative Analysis of Metal Subgenres in Terms of Lexical Richness and Keyness

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophie

an der Ludwig-Maximilians-Universität München

vorgelegt von

Volkan Kahraman

geboren in Frankfurt, Deutschland

München, Oktober 2020

Erstgutachterin: Prof. Dr. Angela Hahn

Zweitgutachter: Prof. Dr. Hans-Jörg Schmid

Drittgutachter: PD Dr. Peter-Arnold Mumm

Tag der mündlichen Prüfung: 14. Juli 2020

ACKNOWLEDGEMENTS

I have always heard from people around me about the difficulties of doing a PhD, most of which centered on the psychological damage it leaves on the scholars. Although I had anticipated similar effects, I was also confident that I would do it more easily since it is on a subject which I like the most – metal. By the end of the study, I was baffled by my own naiveté. Now I find it strange that I still love metal music. I resemble the effort and misery in this long-lasting study to the one in the recording session of *W.A.S.P.*'s *The Crimson Idol* album of which detailed account is provided by Blackie Lawless himself in the album sleeve as follows:

> When I started this record, I had no idea of how long it would take or the diverse elements involved, so we started keeping track. 425 drum heads, 54 sets of guitar strings, 75 guitar picks, over 24 miles of recording tape (48 reels of 2" tape), 150 cassettes, 315 gallons of beer (both domestic and import, Corona Extra owes us and endorsement), 27 gallons of Smirnoff vodka (Pierre was real popular on this one), 175 gallons of drinking water (this Hollywood shit isn't fit to swallow), 251 gallons of soft drinks, 286 rolls of toilet paper (no comment), 210 light bulbs, over 1,200 pinball games (Mikey's the champ), 15 bottles of aspirin (see Smirnoff), 32 bottles of Vitamin C, an unruly 24 track machine named "Otis" (who did not go the distance), one 6.1 earthquake, and a fucking major flood in Feb. '92 to top it all off. All this and a couple of dozen "knock-down, drag-out" arguments later, this monstrosity is finally complete.[1]

I personally did not keep such a detailed record probably because I had never thought it would last this long. Laid out in Lawless's fashion, the grand total for the present thesis consists of eight years of study, serious loss of hair, becoming a father, diagnosis of my dad with an incurable disease, a global pandemic (SARS-CoV-2), a three-month lockdown and the subsequent travel restrictions (still on), two car accidents, a broken computer, a broken tablet, nine colloquiums, over 700 articles and books read, 510 pages written, 180 pages deleted, countless sleepless nights, oceans of coffee and enough headaches to pass a mammoth out. All this and thousands of the infamous question "*when will you be done*?" later, this "monstro-study" is finally complete.

I would like to thank my supervisor (Doktormutter) Prof. Dr. Angela Hahn for her endless support and belief in my dissertation. This thesis would not have existed were it not for you. Your easy-going and positive attitude always kept me

[1] W.A.S.P. – The Crimson Idol (1992), Sanctuary Productions, Parlophone. p. 3.

on track whenever I felt astray or lost in the study. I feel myself honored to be your student. Prof. Dr. Hans-Jörg Schmid for his invaluable contribution especially in the analyses; PD Dr. Peter-Arnold Mumm for his positive feedback; Dr. Abdullah Can without whose books and first-hand support I would have a hard time; Dr. İsmet Öztürk – my MA supervisor – who encouraged and helped me to continue my academic career to a PhD; Dr. Meral Öztürk for being there when I needed some advice; late Prof. Dr. Şeref Kara for his unforgettable support (RIP); Prof. Dr. Aysan Şentürk, Prof. Dr. Sevda Gürsakal and Dr. Melih Engin for their most valuable assistance in statistics; Tom Cobb for his outstanding work at lextutor.ca and his prompt communication; Meik Michalke for his superb koRpus plugin and replying my emails quickly and positively; Dr. Monika Geist for her clear explanations and continuous help; my "classmates" Dr. Franz Steinberger, Dr. Jan Ullmann and Klára Klimčíková for their tremendous support – as well as all the fun; Dr. Umut Salihoğlu and Dr. Uğur Çetinavcı for always being there whenever I needed some help;

Prof. Dr. Müfit Parlak and Dr. Sıdıka Parlak for their endless support and encouragement; Prof. Dr. İ. Naci Cangül, Prof. Dr. İlkin Çavuşoğlu, Prof. Dr. Sezen Özeke, Prof. Dr. Ömer Kaynaklı, Dr. Bülent Şenay and Dr. Ozan Aşık – my former and current bosses – as well as my colleagues for their support and understanding; Lec. Mehmet Doğan for his outstanding support and German translations; Prof. Dr. Yunus Alyaz who have shown great interest in my work; Dr. Z. Ezgi Kara for all the help;

My dear wife İnci Kahraman who never gave up supporting and believing in me and always helped me up when I felt badly down; my dear son Doruk Kahraman, who was born during my PhD study, for being unexpectedly understanding when I had to write this dissertation during when I actually should have been spending time with him; my mom and dad, who have impatiently waited for this, for all their efforts, help and belief in me;

Last but not least, I would like to thank metal music for giving me a degree in academia as well as a lifestyle.

TABLE OF CONTENTS

I. FIGURES

II. TABLES

III. LIST OF ABBREVIATIONS

AAVE	African American Vernacular English
AWL	Academic Word List
B2K	Beyond 2,000 Words
BE06	The BE06 Corpus of British English
BNC	British National Corpus
CEC	Cambridge English Corpus
COCA	Corpus of Contemporary American English
CW	Content words
DMSC	Death Metal Subcorpus
EFL	English as a Foreign Language
EP	Extended Play
GSL	General Service List
HMSC	Heavy Metal Subcorpus
K1	1st 1000 words
K2	2nd 1000 words
KW	Key word
L1	First language, native language
L2	Second language
LD	Lexical Density
LFP	Lexical Frequency Profile
LO	Lexical Originality
LOB	The Lancaster-Oslo-Bergen Corpus

LQ	Lexical Quality
LS	Lexical Sophistication
LV	Lexical Variation
MC	Metal Corpus
NGSL	New General Service List
NS	Native speaker of English
NNS	Non-native speaker of English
NWOBHM	New Wave of British Heavy Metal
RIAA	Recording Industry Association of America
TESL	Teaching English as a Second Language
TMSC	Thrash Metal Subcorpus
TTR	Type/Token Ratio

IV. ZUSAMMENFASSUNG

Metal-Musik wird in einer Vielzahl von Subgenres ausgeführt, die jeweils alle ihre eigenen (oder gemeinsamen) Merkmale im Klang oder auch in den Songtexten besitzen. In der diesbezüglichen Literatur wird Metal-Musik hinsichtlich ihrer Subgenres und im Fokus verschiedener Musikrichtungen diskutiert. In den wissenschaftlichen Studien über Songtexte liegt der Schwerpunkt jedoch insbesondere auf vorhandenen Themenbereichen (Walser 1993, Weinstein 2000, Morrison 2006, Pieslak 2007, Buts und Buelens 2008, Philips und Cogan 2009, Strother 2013 u.a.). Dies erfolgt meistens im Rahmen eines semantischen Ansatzes und zieht soziologische oder psychologische Schlussfolgerungen. Erzielt werden dabei oft lediglich Verallgemeinerung über das erforschte Genre. Im Gegensatz dazu fokussiert sich die vorliegende Arbeit auf die linguistischen Merkmale der Songtexte, um objektive und messbare Befunde zu erzielen. Aus dieser Sicht kann diese Arbeit als eine Forschung eingestuft werden, die einen interdisziplinären Beitrag sowohl zum neuen akademischen Fachgebiet Metal-Musik als auch bei der Anwendung von bestimmten linguistischen Analysen auf unerforschte Kunstrichtungen – hier Metal-Songtexte – leisten soll.

Aus der linguistischen Perspektive setzt sich diese Arbeit die Entwicklung von Analyseverfahren als Ziel, wodurch die Subgenres von Metal-Musik angesichts der Songtexte kategorisiert werden sollen. Es ist jedoch zu bemerken, dass es sich dabei nicht um die Kategorisierung von Songtexten ohne Rücksicht auf die musikalischen Eigenschaften handelt. Ganz im Gegensatz ist hier das Ziel, die Wirksamkeit von Wortschatzanalysen bei der Unterscheidung von verschieden Genres festzustellen. Dies bedeutet, dass diese Arbeit die Legitimation von vorhandenen Genres nicht bezweifelt, sondern anerkannt.

In dieser Arbeit werden von allen Subgenres jeweils 200, somit ein Korpus von 600 Songtexten von Heavy Metal, Thrash Metal und Death Metal hinsichtlich des Wortschatzreichtums sowie der Schlüsselwörter analysiert und es wird versucht, eine Korrelation zwischen den Subgenres aus der Sicht der musikalischen Härte und des Wortschatzreichtums herzustellen. Die

Grundannahme hier ist, dass in verschiedenen Härtestufen von der Metal-Musik ein anderer Wortschatzreichtum und andere Schlüsselwörter vorhanden sind.

Die Forschungsfragen und die Schwerpunkte dieser Arbeit lauten wie folgt:

1. Gibt es zwischen den Wortschatzniveaus der Metal-Subgenres Unterschiede?
   a) Gibt es Unterschiede in den Metal-Subgenres in Bezug auf die lexikalische Dichte?
   b) Gibt es Unterschiede in den Metal-Subgenres in Bezug auf die lexikalische Spezifiziertheit?
   c) Gibt es Unterschiede in den Metal-Subgenres in Bezug auf die lexikalische Variation?
2. Gibt es Unterschiede zwischen den Schlüsselwörtern von Heavy Metal, Thrash Metal und Death Metal und falls ja, wie sieht die Beziehung zwischen den Subgenres in Bezug auf die Schlüsselwörter aus?
3. Können Analysen des Wortschatzreichtums und der Schlüsselwörter für einen Vergleich von Musikrichtungen verwendet werden?
4. Gibt es eine Übereinstimmung zwischen den existierenden Kategorisierungen und der lexikalischen Kategorisierung?

Auf der Suche nach Antworten auf diese Fragen wurde darauf abgezielt, die lexikalische Unterschiede der Subgenres zu finden. Die Hypothese lautet: „Je härter die Musik ist, desto größer ist der Wortschatzreichtum." Diesbezüglich wird erwartet, dass der Wortschatzreichtum von Heavy Metal an letzter Stelle, von Trash Metal an zweiter Stelle und von Death Metal an erster Stelle steht. Diese Theorie bezieht sich auf die Kategorisierung der Metal-Richtungen in *Dionysian* und *Chaotic* von Weinstein (2000), auf die allgemein anerkannte Auffassung, dass Thrash Metal und Death Metal wildere, brutalere und störendere Songtexte beinhalten (Walser 1993, Arnett 1996, Buts und Buelens 2008, Phillipov 2012), sowie auf die exemplarische Arbeit von Taina (2014) über die Metal-Subgenres.

In der vorliegenden Arbeit wurde die Fachliteratur ausführlich auf Genre, Korpus und Wortschatzreichtum untersucht und die angewandten Methoden detailliert dargestellt. Insbesondere wurden die Komplexität des Wortschatzreichtums und die Schlüsselwörter dargelegt und eine Schlussfolgerung gezogen. Der Wortschatzreichtum wurde dabei aus drei Perspektiven analysiert: (1) lexikalische Variation (2) lexikalische Spezifiziertheit und (3) lexikalische Dichte. Die lexikalische Variation wurde mithilfe von TTR, Guiraud, Uber und HD-D bewertet. Die lexikalische Spezifiziertheit wurde anhand von zwei Frequenzlisten – GSL und BNC/COCA – mithilfe des lexikalischen Frequenzprofils und in Bezug auf die Prozentzahl von Token und Type bewertet, dies sich nicht innerhalb der meist benutzten 2000 Vokabeln befinden (Laufer 1995). Des Weiteren wurde der P_Lex, ein weiterer Test zur Feststellung des Wortschatzreichtums mit Bezug auf die GSL-Liste, durchgeführt. Die Analyse der lexikalischen Dichte wurde als Proportion von Inhaltswortschatz zu Tokens berechnet. Um diese quantitative und datenorientierte Methode mit einer qualitativen Methode zu vervollständigen, wurde dazu noch eine Analyse der Schlüsselwörter ausgeführt.

Die Auswahl der Bands erfolgte durch Referenzen in der Fachliteratur zur Metal-Musik. Die recherchierte Fachliteratur über die Metal-Musik besteht unter anderem aus akademischen Büchern, Artikeln sowie aus bemerkenswerten Medienproduktionen wie *Metal Evolution*, *Encyclopaedia Metallum*, *Invisible Oranges* und aus Webseiten und Blogs. Die Songtexte wurden manuell bearbeitet und von vorhandenen *meta-data* und *mark-ups* befreit. Somit wurde verhindert, dass die verschiedenen Längen der Wiederholungen die Vergleichsanalysen beeinflussen. Außerdem sind die eingesetzten Analyseverfahren auf Wiederholungen sensitive, da sie die Frequenz der Wörter und das Verhältnis der Wiederholungen messen. Des Weiteren wurden Songtexte mit mindestens 100 und höchstens 400 Wörtern ausgewählt, weil manche Messmethoden sensitive auf die Textlänge sind.

Die Analyse zum Wortschatzreichtum wurde mittels vier verschiedenen Programmen ausgeführt. Die Wortschatzvariation wurde mit den Methoden – Uber und HD-D – mittels des Plug-In *koRpus* des Programms *R* von Michalke (2018) berechnet. LFP wurde mittels *AntWordProfiler*, eines von Anthony (2014) entwickelten Analyseverfahrens zur lexikalischen Spezifiziertheit, und P_Lex wurde mittels eines online Mediums von Meara (2018) bewertet. Die Analyse zur Wortschatzdichte wurde mit der online Version des *Lexical Complexity Analyzer* von Ai und Lu (2010) und Lu (2012) ausgeführt. Die Befunde wurden mittels Kolmogorov-Smirnov, Kruskal-Wallis H, Mann-Whitney-U, Welch Anova, Dunnett's C, Spearman's correlation coefficient und Quadratic Discriminant analysiert.

Da der Kolmogorov-Smirnov-Test ergab, dass die Daten mit Ausnahme von Guiraud keine Normalverteilung zeigten, wurde ein parameterfreier Kruskal-Wallis-H-Test durchgeführt, und es wurde ein Unterschied zwischen den Gruppen festgestellt. Um zu bestimmen, zwischen welchen Gruppen Unterschiede existieren, wurden Mann-Whitney-U-Tests durchgeführt und es ergab, dass es statistisch signifikante Unterschiede zwischen allen Subgenres gab. Obwohl die Guiraud-Werte eine Normalverteilung zeigten, wurde der Welch-ANOVA-Test durchgeführt, da die Varianzen nicht gleich waren und beim Dunnett's C-Post-hoc-Test ein Unterschied zwischen allen Subgenres bemerkt wurde. Als Ergebnis wurde festgestellt, dass Heavy Metal in allen Analysen den niedrigsten und Death Metal den höchsten Wortschatzreichtum aufwies. Bei der Korrelationsanalyse zwischen den Ergebnissen wurde konstatiert, dass alle Methoden zur Messung des Wortreichtums in positiver Korrelation stehen. Schließlich stellte die mit den normalisierten Daten durchgeführte *quadratic discriminant*-Analyse die Frage, inwieweit die aus der Wortanalyse erhaltenen Daten mit den vorhandenen Arten übereinstimmen und es wurde eine Übereinstimmung von 63,2 % aufgezeichnet. Weiterhin wurde beobachtet, dass zwei Listen, GSL und BNC/COCA, die für den Wortentwicklungstest verwendet wurden, sehr enge Ergebnisse lieferten.

Die Keyword-Analyse wurde mit *AntConc v 3.5.7* (Anthony 2018) durchgeführt und es wurden wie bei der Analyse des Wortschatzreichtum Unterschiede zwischen allen Subgenres festgestellt. Wie Weinstein (2000) bemerkt, sind die Schlüsselwörter im Heavy Metal dionysischer, während Thrash und Death Metal mehr chaotische Wörter beinhalten. Basierend auf diesen Erkenntnissen kann Folgendes vorgeschlagen werden: (1) Diese drei Subgenres unterscheiden sich nicht nur musikalisch, sondern auch in Bezug auf Wortschatzreichtum und Schlüsselwörter und (2) die Analysen zum

Wortschatzreichtum und Schlüsselwörter können die Genreunterschiede widerspiegeln. Nach den Ergebnissen der Diskriminanzanalyse des vorhandenen Korpus entspricht jedoch eine umgekehrte Auffassung zur Bestimmung der Genres anhand lexikalischer Eigenschaften nicht völlig der aktuellen Klassifizierung.

V. ABSTRACT

Metal music is realized under a vast variety of subgenres all of which have their unique (or shared) characteristics not only in sound but also in their lyrics. Much research has been done to distinguish or classify subgenres but little has addressed the linguistic differences across them. This study seeks to find out the lexical richness and keyness levels of heavy metal, thrash metal and death metal using a corpus of 200 songs from each subgenre with a total of 600 songs. The selection of the bands and songs was carried out finding references in the metal literature. The metal literature in the present study takes into account the academic books and articles on metal as well as noteworthy media productions, websites and metal blogs such as *Metal Evolution* and *Encyclopaedia Metallum*.

The song lyrics were manually processed and meta-data, mark-ups and repeats have been removed so that the differences in repeat lengths do not affect the comparisons. Furthermore, the analyses used in the study are sensitive to repeats as they measure the frequencies and repeat ratios of the words. The song lengths – after the processing – were limited to lower and upper thresholds of 100 and 400 words.

The songs were analyzed for their lexical richness levels in three aspects: 1) lexical variation, 2) lexical sophistication and 3) lexical density. Lexical variation was operationalized as TTR, Guiraud, Uber and HD-D. Lexical sophistication was measured using lexical frequency profile with two different frequency lists – the GSL and the BNC/COCA – by looking at the ratios of tokens and types which fell beyond the most frequent two thousand words (Laufer 1995). Another sophistication measure – P_Lex – which also runs on GSL, was applied. Lexical density analysis was based on the ratio of content words to all tokens in the texts. In order to complement this quantitative and data-driven approach, a keyness analysis was administered to add a qualitative dimension to the research.

All lexical richness analyses pointed out to statistically significant differences between all subgenres, marking heavy metal as the least and death metal as the most lexically rich one. Keyness analysis indicated differences among all three subgenres as well. Heavy metal key words tended to be Dionysian whereas thrash and death metal keywords were more Chaotic as proposed by

Weinstein (2000). Finally, a correlation analysis showed that all lexical richness measures were statistically significantly correlated to each other. Based on the findings, it could be claimed that 1) these three subgenres differ from each other not only in terms of music but also of lexical richness levels and key words and 2) lexical richness analyses, coupled with keyness, are capable of reflecting the genre differences in song lyrics. However, as a result of a discriminant analysis of the present corpus, a reverse approach whereby genres are attempted to be classified based on lexical features does not provide a pattern which fully corresponds to the existing classifications.

Key words: metal, heavy metal, thrash metal, death metal, genre, subgenre, lexical richness, lexical sophistication, lexical density, lexical variation, keyness.

1. INTRODUCTION

Metal is such a broad music genre that it could cover numerous bands and albums which in fact have very little shared features as opposed to massive differences they display both musically and lyrically. This study aims to point out that metal is not a monolithic genre and that its subgenres show different linguistic characteristics in addition to more overt musical ones. Thus, research at subgenre level might reveal more solid results regarding the linguistic differences.

Metal is obviously different from other musical genres and quite easy to distinguish even for the non-fans. As far as the non-fans are concerned, the common opinion regarding the formula of metal could be described as loud guitar distortion + shouting/yelling + silly lyrics. For the fans though, not all guitar music counts as metal. Furthermore, metal, being a subgenre of rock, is comprised of many subgenres bearing different characteristics in music, sound, vocals, technicality, speed, look, lyrics, etc. Some of these differences clarify the distinction of subgenres immediately whereas some cases (bands/albums) need additional criteria for an exact classification. Sound, vocals and speed are usually the best predictors on which most critics and academics base their classifications (see Walser 1993, Berger 1999, Weinstein 2000, Christe 2003, Stelzner, Morrison 2006, Philips and Cogan 2009, Frandsen 2011, etc.).

Identifying or labeling the genres and assigning the bands to one are hardly focuses of concern for the bands themselves. Nevertheless, media and fans constantly pursue labeling for informative purposes, sometimes at the expense of receiving heat or disregard from the bands. For academic genre studies, labeling is indeed important. Such a big genre as metal needs to be divided into subgenres reflecting similar characteristics and research must be carried out accordingly.

The subgenres in metal have been studied with focus on lyrics as well as other aspects of the music, yet research on song lyrics mostly center on the thematic aspects (e.g., Walser 1993, Weinstein 2000, Morrison 2006, Pieslak 2007, Buts and Buelens 2008, Philips and Cogan 2009, Strother 2013, etc.). This is mostly realized in an overall semantic approach yielding sociological or psychological inferences out of the song lyrics. Such studies often come up with generalizations with respect to the genres in question.

The present study, in contrast, specifically focuses on the linguistic features in order to achieve an objective and measurable outcome in terms of the lexical richness and keyness levels of the given metal subgenres. It could be regarded as an interdisciplinary contribution to the field of metal music studies on the one hand, which is an emerging academic field of study, as well as, on the other hand, an application of selected linguistic analyses on a work-of-art not analyzed so far – metal lyrics. The study tries to establish a correlation between so-called 'heaviness' of subgenres as defined by criteria applied in music studies, and the lexical richness of their lyrics. The underlying assumption is that metal at different levels of heaviness posits different levels of lexical richness and different key words as well. The study seeks to delve into the lyrical aspect of metal songs in a linguistic approach in order to obtain quantitative results with regard to the lyrical differences of metal subgenres.

From the perspective of linguistics, the study seeks to devise analysis methods at lexical level to distinguish metal subgenres based on their lyrics. However, it needs to be stressed that the aim is not to develop an alternative method to 'label' or 'categorize' bands or songs isolated from their musical traits. Instead, the aim is to test the distinguishing capabilities of certain lexical analysis methods across genres. This means that the study does not stand against the existing genres; on the contrary, takes them for granted. Following the analyses, the results will be compared to the existing genre distinctions and the extent to which the two divisions overlap with each other will be discussed. This will be realized with both a quantitative (lexical richness) and a qualitative (keyness) approach. In sum, the present study will offer a methodology to find lexical differences across different genres and try to identify whether the literature on metal could be endorsed with one on linguistics.

The theoretical background provides definitions on some key concepts on genres and subgenres. Then, research in metal genres is reviewed with respect to the three genres under investigation and the characteristics of metal lyrics in particular are discussed. As it is a corpus-driven study, following the genre-related section, a section on corpus linguistics will be provided shedding light on the important dimensions of corpus building. Then, lexical richness will be introduced on a broad scale with references to various theories and approaches. The chapter ends with the last analysis method – keyness.

Methodological chapter will provide details on the establishing of the corpus used in the study including the selection of the songs and the compilation of the lyrics. The chapter will follow to give a detailed account of the analysis methods used in the study. The results and discussion chapters have the same outline whereby the analyses are grouped separately. An overall discussion will be provided at the end and the findings, limitations and further study will be discussed in conclusion.

1.1. Research Questions

The present study focuses on the determination of the lexical richness of three metal subgenres using a specialized song lyrics corpus. The research questions, and thus the primary focus, of the present study are:

1. Are there any differences between the lexical richness levels of metal subgenres?
    a. Are there any differences between lexical density levels of metal subgenres?
    b. Are there any differences between lexical sophistication levels of metal subgenres?
    c. Are there any differences between lexical variation levels of metal subgenres?
2. Are there any differences between the keywords of heavy metal, thrash metal and death metal and if any, how do they relate to the respective subgenre?
3. Can lexical richness and keyness analyses be used in music genre comparison?
4. To what extent do the existing categorizations overlap with the lexical categorization?

By seeking answers to these questions, this study aims to find lyrical differences between the metal subgenres. The hypothesis of the study is that ‘the heavier the music, the more lexically rich the lyrics’. This suggests that heavy metal is expected to be the least lexically rich subgenre of the corpus, thrash metal the second and death metal the most. This hypothesis is based on Weinstein’s (2000) categorization of metal genres as Dionysian and Chaotic, acknowledgement of thrash and death metal to feature more violent, brutal and explicit lyrics (Walser 1993, Arnett 1996, Buts and Buelens 2008, Phillipov 2012) and Taina’s (2014) exemplary research on metal subgenres.

2. THEORETICAL BACKGROUND

The theoretical part of the paper will expand on the key aspects of the study. As the research brings together two different fields of study – music (or more specifically metal) and linguistics, the theoretical part will attempt to shed light on both areas in a detail level specific to the focus of the study.

As the study focuses on genre comparison, the theoretical chapter will start with a section on genre. Genre is a very broad field of study spanning across various disciplines. In the present paper, it will be introduced from a brief general point of view underlining the basic concepts. It will include the terminology and definitions as well as the classification and comparison of genres. Then, the genre concept will be integrated in metal in an attempt to clarify and justify genre/subgenre divisions in the said music style. The genre part will conclude with details on the selected subgenres and bands.

Being a newly-emerged field of study, metal is likely to be unfamiliar to other research areas. There are controversies even within the metal literature on various subjects which can be attributed to both a relatively limited amount of research in the field and its rather subjective and non-classifiable artistic nature. Hence, it cannot even be possible for metal fans to agree on some points. With these considerations in mind, care has been taken to provide information on metal on a non-fan level, where non-fans mean people who are not familiar with the genre and not those who do not favor it. In short, the first part of the chapter, genre, will start with the genre concept in general and narrow it down to the subgenres of metal.

Then, a detailed account of corpus linguistics, featuring necessary concepts, will be provided. Given the fact that the study is a corpus-driven one, the theory behind the formation of the corpus in the study will be presented from various viewpoints.

As the next key aspect, lexical richness will be presented at observational and operational levels (Bulté and Housen 2012: 28), and then numerous measurement methods will be introduced. It will, then, move on to a detailed account of the literature on lexical richness studies. This study seeks to apply the lexical richness measurement methods in a relatively rare area. It should be noted that lexical richness methods will be investigated based on notable scholars' theories and in relation to their plausibility for application on song lyrics. An in-depth account of the selected measurement methods will be provided in the respective subchapters. This will conclude the quantitative aspect of the study.

Finally, the qualitative element of the study, keyness, will be introduced. Although a relatively new field of study, keyness analyses have been rather quick to be acknowledged in genre-related studies (see Xiao and McEnery 2005, Scott and Tribble 2006, Pojanapunya and Todd 2015). This section will give information on the theory and measurement methods of keyness analysis. The method is expected to enrich the present study by contributing a qualitative value to support the quantitative findings, which will be obtained from the lexical richness analyses. Although a qualitative analysis, keyness is measured through dedicated software, which run on certain statistical parameters and in comparison to a reference list. This makes the analysis more reliable as the human factor is virtually zero.

2.1. Genres

The present study aims to compare the lexical differences of different genres. Thus, the starting point will be to give detailed information on genre. The concept will be discussed under three main headings: 1) The Notion of Genre which will feature overall key aspects, 2) Genres in Metal where the emergence and categorization of metal genres will be delineated, and 3) The Three Metal Subgenres Used in the Current Study where specific information on the chosen subgenres will be provided.

2.1.1. The Notion of Genre

In this section, the notion of genre will be introduced as the first step. It will be detailed in terms of terminology as many scholars have discussed it in different contexts and using various terms. Each field has their own definitions, approaches and discussions on genres. This section will feature some major approaches in the concept of genre by notable scholars. Then, the definitions of genre will be introduced in an attempt to shed light on characteristics of genres.

Finally, the classification and comparison of genres will be addressed with references to numerous publications.

2.1.1.1. Terminology

A genre is a category of artistic, musical, or literary composition characterized by a particular style, form, or content (Genre). Subgenre, on the other hand, is a subcategory within a particular genre. DiMaggio (1987: 441) defines genres as “sets of artworks classified together on the basis of perceived similarities”. Genres are mostly defined based on their content (e.g., themes or settings) and form (structure and style) shared across similar works of art.

Genre is a very broad term used in various fields of study. The concept has been defined and discussed by many scholars in different fields such as rhetoric, literature, linguistics and media. There is a maze of terminology used with the genre concept which varies depending on scholars and approaches. Lee (2001) gives a broad overview of the different terminology about genre mostly focusing on the terms, *genre, register, text type* and *style.* This paper will draw highly on his views as distinction criteria.

According to Lee (2001: 38), the distinction between *genre* and *text type* stems from the viewpoint of the classification. He bases his argument on that of Biber (1988: 70) where he uses the term *genre* to refer to external criteria, “such as intended audience, purpose, and activity type”, and *text type* for linguistic forms (ibid.). He also notes that there are no clear-cut definitions across text types and overlapping might occur, such as between novels and biographies, which are different genres, yet likely to share similar linguistic features (text types) such as the use of past tense, third-person use, etc. (p. 39). As the next step, he sheds light on the distinction between *genre*, *register* and *style*. *Genre* and *register* are used interchangeably most of the time and there are differences of opinion among scholars as to their distinction and hierarchy. In systemic-functional approach, while Halliday and Martin (1993) consider genre to cover register, Kress (1993 qtd. in Lee 2001: 45) nests genre under register. *Style*, on the other hand, is related to one’s use of language which is based mostly – but not limited to – on formality (p. 45). Lee (2001) summarizes the distinction between *genre* and *register* as follows:

> [R]egister [is] used when we are talking about lexico-grammatical and discoursal-semantic patterns associated with situations (i.e., linguistic patterns), and genre [is] used when we are talking about memberships of culturally-recognisable categories. […] genres are about whole texts, whereas registers are about more abstract, internal/linguistic patterns, and, as such, exist independently of any text-level structures (pp. 46-7).

He concludes his arguments stating that when comparing genres what is actually done is to compare their registral variation which occurs situationally or sociologically but not because they belong to a particular genre (pp. 47-48). To incorporate this approach into the present paper, it could be argued that the established metal subgenres will be compared to each other based on their linguistic features (register).

Biber's stance on the terminology revolves around the terms *genre* and *register* as he is more inclined to opt for the term *register*. He claims the distinction between these two terms (and also *style*) is quite vague and different scholars have different interpretations and he prefers to use *register* as a cover term (Biber 1995: 8-9). *Register* is the situational linguistic characteristics and they "are defined in non-linguistic terms, by differences in purpose, interactiveness, production circumstances, relations among participants, etc." (Biber 1995: 7). This view is opposite of Lee (2001). Biber defines *text type* more straightforwardly as it refers to the distinction based on internal (linguistic) criteria (Biber 1993: 244-5) which corresponds to Lee's (2001) *register* definition. According to Biber et al. (1998: 135), some registers could be very specific (e.g., particular sections of biology articles or novels written by Jane Austen) whereas other are very broad, such as conversations or student essays. However, his description (or understanding) of *genre* in Biber (1988) is that "*genre* […] refer[s] to text categorizations made on the basis of external criteria relating to author/speaker purpose" (p. 68). The similarity of the description of the two terms is noteworthy. While he uses the term *genre* to illustrate the differences between spoken and written language (Biber 1988) and to account for the appropriate corpus size (Biber 1990), he incorporates both approaches in Biber (2006: 11) and chooses to use the term *register* "to refer to situationally-defined varieties described for their characteristic lexico-grammatical features". He claims that the attempts at distinguishing between the terms *genre* and *register* have been dependent on either the object of the study or the characteristics of language (ibid.).

In the systemic functional approach, *genre* is interpreted as "cultural purpose" which helps identify similar texts (Eggins 2004: 54). The two definitions of *genre* in this respect by Martin (1984 and 1985 qtd. in Eggins 2004) are as follows:

> Firstly, "a genre is a staged, goal-oriented, purposeful activity in which speakers engage as members of our culture" (Martin 1984: 25). Less technically, "Genres are how things get done when language is used to accomplish them" (Martin [1985]: 248) (Eggins 2004: 55).

As a follow-up to these descriptions, Eggins (2004: 56) proposes the following division of genres as depicted in Table 1, which are purpose-based to a great extent:

Table 1: Classification of genres adapted from Eggins (2004: 56).

Genres	**Examples**
Literary genres	Short stories, autobiographies, ballads, sonnets, fables, tragedies
Popular fiction genres	Romantic novels, whodunits, sitcoms
Popular non-fiction genres	Instructional manuals, news stories, profiles, reviews recipes, how-to features
Educational genres	Lectures, tutorials, report/essay writing, leading seminars, examinations, text-book writing
Everyday genres	Buying and selling things ('transactional' genres), seeking and supplying information, telling stories, gossiping, making appointments, exchanging opinions, going to interviews, chatting with friends

In the systemic functional approach, these genres feature three dimensions as *register*, *schematic structure* and *realizational patterns*. The *schematic structure* refers to the way texts are organized depending on the genres such as a face-to-face dialogue commencing with the greeting sequence. *Realizational patterns* are the lexico-grammatical features that we use for specific genres, e.g., using noun phrases in titles or imperatives in cookbook recipes (Eggins 2004: 68). What needs a closer look here is the notion of *register*. In addition to the discussion by Lee (2001), the register theory in systemic functional approach is a part of the genre which has three sub-dimensions as *field*, *mode* and *tenor* (see

Halliday et al. 1964 qtd. in Halliday 1978: 33). Eggins (2004: 58) describes and exemplifies these elements as follows:

> A genre comes about as particular values for field, tenor and mode regularly co-occur and eventually become stabilized in the culture as 'typical' situations. For example, the transactional genre of buying your coffee from the corner café involves the field of 'coffee', the tenor of 'customer/provider' and the mode of 'face-to-face'.

It could be argued that the differentiation of these terms is clearer in systemic functional approach than in others. This is due to the hierarchical status of the two terms nested under one another. Systemic functionalists adopt both terms simultaneously and assign different notions to them. Still, the question of which one is the top-level category is open to discussion depending on the different viewpoints in the literature.

To sum up the difference between *genre* and *register*, it could be argued that *genres* are broader categories clustered based on social and cultural traits whereas *register* is the linguistic use of the genres. From this point of view, the present paper could be a *register* analysis of music genres; however, the terminology of choice will be *genre* as it encompasses the musical style as well as the linguistic features.

2.1.1.2. Definitions

Following the terminological clarification, the next step is to focus on genre as discussed by various scholars. Campbell and Jamieson (1978: 20-1) focus on literary forms and argue that genres are made up of "constellation of forms" which make them distinctive. These forms are substantive, stylistic and situational features and the same forms may occur in different genres (ibid.). However, this is not enough to disregard them as different genres as what is important is the connections of the forms to each other in some distinguishable way. They note that "a genre is given its character by a fusion of forms not by its individual elements" (ibid.) and such an approach is likely to be beneficial to understand the internal system of genres.

Yates and Orlikowski (1992) set out to devise their own concept on genre and they refer to the classic dual approach as 'literature' and 'rhetoric' as a starting point. To illustrate the difference, they remark that novels, comedy, etc.

are called literary genres, whereas sermons, inaugural addresses, etc. could be named among the rhetorical ones. This distinction shows that rhetoric genres are socially oriented. Genres are indeed social institutions which shape and are shaped by the communicative actions of individuals; thus, they propose the term "genres of organizational communication" (p. 301). They state that "[a] genre of organizational communication (e.g., a recommendation letter or a proposal) is a typified communicative action invoked in response to a recurrent situation" (ibid.). Here the "recurrent situation" refers to social relations and established practices. Genre emerges, in this respect, as a concept with two distinct features as 'substance' and 'form'. Here, 'substance' represents the theme, motives and topics; and 'form' the linguistic features. They claim that each genre has their rules in substance and form and that each genre has a different level of abstraction, which is realized in the notion of subgenres. They also mention 'normative scope' as a feature of genres which refers to how widely shared the elements of a particular genre are, e.g., whether they are shared across countries, specific domains or a very limited group of professionals. The last one of their arguments that needs mentioning here is that genres – and their substances and forms – are not static in nature and bound to be modified. Such modifications, and especially significant and persistent ones, might create new subgenres or in some extreme cases lead to emergence of new ones (pp. 306-8).

One way to identify genres is focusing on the characteristics of the texts. In this regard, "genre is a model for describing and classifying literary texts" (Moessner 2001: 131). Mittel (2001: 5) coins this approach of text-based genre analysis as "textualist assumption" and claims, from the viewpoint of media, that genre might not be limited to mere text attributes. He suggests referring to the audience and the industry for their perception of the genres as well (p. 8). This approach also manifests itself in the communicative aspect attributed to the genre concept. Genres are "a class of texts characterized by a specific communicative function that tends to produce distinctive structural patterns" (Holmes 1997: 2). It is the communicative aspect of the genre that makes it an interactive concept, whereby the producer of the text and the listener/reader find a common field and affect each other within the boundaries of the genre. As a result of such genre awareness, the readers are guided towards certain attitudes and assumptions to make sense out of a text (Chandler 1997: 8). For instance, a crime novelist is aware of the genre he writes in and should observe the rules and features of the genre. Similarly, the reader expects from the novelist a particular type of novel fitting the genre which eventually affects the way s/he interprets the novel. In short, as DiMaggio (1987: 445) notes, “genre classifications let consumers invest in specialized knowledge and permit artists to do their work”.

The communicative aspect is particularly evident in Swales (1990) where he defines genre as a set of communicative events sharing communicative purposes. The sharing Swales (1990) mentions takes place in “discourse communities” which are essentially groups of individuals who have a common set of goals. Discourse communities are communities where a language or discourse exists and whose members have the say to include or exclude others (Teubert and Čermáková 2004: 114-5).

Another important concept for genres, proposed by Swales (1990), is “prototypicality”. It basically refers to the level of genre membership whereby some members are considered rather straightforward ones whereas some could be disputable. Swales (1990: 52) states that “the most typical category members are *prototypes*” and explains this concept with an example of birds noting that a sparrow is a more typical member of bird species while an ostrich could be considered an untypical one.

Although many definitions and theories on genres have been proposed in the literature, it must be underlined that defining genres is quite problematic since not all of them are of static nature. They evolve and change over time as a result of several factors such as the social changes and the advance of technology. Some genres disappear and new ones emerge (Chandler 1997: 3). This dynamic nature of genres, as well as the other traits mentioned above, is one of the aspects to be taken into consideration in the classification of genres.

2.1.1.3. Classification and Comparison of Genres

Genres, by their nature, are bound for classification. This could be as a result of an aggregation principle to sort cultural products into categories as noted by DiMaggio (1987: 441). Classifying genres is useful as it allows us to discover what an item shares with others in the same genre or how it differs from the items in other genres (Aucouturier and Pachet 2003: 83). As one of the earliest examples, Aristotle describes three kinds of rhetoric as deliberative, forensic, and epideictic (Miller 1984: 152-3). Another rhetorical division is proposed as narrative, exposition, argument, description (Devitt 1993: 574). Genre is an indispensable aspect of literature, evidently noted in the broad division of literary texts into poetry, prose and drama, which are then further divided into more specific categories, such as drama branching as tragedy and comedy (Chandler 1997: 1). In media (and especially television), some of the main genres are comedy, commercials, game shows, documentaries, etc.

Although it may seem as a relatively easy task to assign or classify genres, genre is, as Devitt (1993: 573) puts it, a "trivial concept" and not necessarily dependent on clear criteria which is also the case in music genres (metal in particular) (see Classification of Metal Subgenres). It is not objective and it often gives way to disagreements between theorists, even to an extent where "one theorist's genre may be another's sub-genre or even super-genre" (Chandler 1997: 1).

Each genre displays similarities or differences to the other which could be analyzed by means of comparisons with various criteria. The differences between genres or a particular genre and the general language can surface in terms of the distribution and frequency of linguistic features (Aijmer and Lewis 2017: 2). Genre analyses could be carried out at macroscopic and microscopic levels (Biber 1988: 61). In macroscopic analyses, the aim is to discover the overall features of a particular domain/genre. Texts from different genres are compared to each other by means of various criteria in order to achieve a general distinction between them. Microscopic analyses, on the other hand focus on specific linguistic items. For example, the use of *may/might* could be investigated so as to find a pattern and/or support/refute a theory. Biber (2006: 62) states that these two approaches, although different, complement each other. It is through the administration of a microscopic analysis that important items could be identified and in-depth investigations could be possible.

As mentioned earlier, genre is a broad term and in order to employ a target-specific approach, examples of comparisons to be provided in this section will be limited to linguistics. To start with a major approach, Biber (1986) came up with a groundbreaking method of genre comparison called Multi-Dimensional Analysis (MDA). He compared mostly spoken registers/genres to written ones (see Biber 1986, 1988, 1992 and 1995) to find out discursive similarities and differences. MDA is based on grouping of similar linguistic features into clusters following a factor analysis and the resulting clusters are called dimensions (Biber 1986: 391). The three dimensions he came up with were "'*Interactive vs. Edited Text'*, '*Abstract vs. Situated Content'*, and '*Reported vs. Immediate Style'*" (p. 410). These dimensions were quantitatively analyzed based on 41 different language structures such as contractions, *if*-clauses, nominalizations, agentless passives, subordinator *that* deletion, 3rd person pronouns, present tense, etc. (p. 393).

Biber's method of analysis showed that genres do not have to be distinguished externally and that linguistic analysis can uncover a good deal of variation among traditional genres (Kennedy 1998: 198). MDA, which "is methodologically extremely distinct and statistically sophisticated" method (McEnery and Hardie 2012: 94), evolved over time and Biber labelled the dimensions in various ways, included more linguistic features and excluded some of them as well. He applied 67 features in Biber (1988: 74-5) and 33 in Biber (1992: 140). The number of identified linguistic features in Biber (2006) was 129; however, he reduced the number from 129 to 90 in the same study on grounds that some of them overlapped with each other (p. 182).

As mentioned above, Biber's focus was mainly on the difference between spoken and written registers and since they are quite different from each other, at the level of medium to say the least, it makes sense to apply a more comprehensive methodology. His method was applied by other scholars as well. Crossley and Louwerse's (2007) MD analysis for the classification of spoken and written registers, where they studied the frequency of bigrams, were found to be compatible with Biber's (1988) findings. They however caution that the analysis could not differentiate between similar registers (p. 475).

MDA has also been applied to popular song lyrics by Bértoli-Dutra (2014). She extracted two sets of dimensions as lexico-grammatical and semantical, which contain persuasion, interaction, narration, personal action, emotion and society and musical reference, and analyzed song lyrics in terms of these dimensions. She found that most of the results were in accordance with Biber's (1988) findings as well (p. 170). Among many music genres such as pop, country, alternative, grunge, rock, hard rock, etc., heavy metal got the highest negative dimension score for persuasion and narration but the highest positive score for the emotion and society dimension in the study.

Groundbreaking as it is, MDA is not without its criticism, though. In spite of the fact that MDA is a comprehensive analysis, it is quite complicated, difficult, time consuming (Kilgarriff 1995: 613 qtd. in McEnery and Hardie 2012: 110-111) and it requires a high level of expertise (Hardy 2015: 172). It has also been reported that replication of the MD analysis has not been quite successful due to the lack of relevant software and datasets (Doyle 2005: 4 qtd. in McEnery and Hardie 2012: 112).

Johansson (1978 qtd. in Kennedy 1998: 104-5) compared one section of the Brown Corpus (learned and scientific writing) to the whole of the corpus to study the distribution of particular words. He found that some words were more frequent in scientific texts (such as *discussion*, *probability*, *species*, etc.) while locative adverbs of time and space were more frequent in the general corpus. Rutherford (2005) studied a corpus of Operating and Financial Reviews (OFR) and looked at the relation of word choices of companies, which are grouped based on their financial strength and size, such as profit-making, least profitable, largest, etc. He grouped the 50 most frequent words from each company group and merged them into one list. Then, he split the list into two as positive and negative words. He interestingly found that loss-making companies used more references to profits than to losses.

Nippold et al. (2005) compared the syntactic complexity levels of expository and conversational discourse and concluded that expository discourse displays a higher level of complexity. Lu (2011: 49) compared argumentative and narrative essays and found that "argumentative essays generally exhibit higher syntactic complexity than narrative essays". His conclusion is that genre affects the levels of complexity. With regard to lexical studies, Chung and Nation's (2003) study on technical words looked at anatomy texts and applied linguistics texts and concluded that anatomy texts exhibited one technical word out of every three whereas the ratio was one in five in applied linguistics texts (pp. 109-10).

Comprehension and retention studies have been carried out mainly on students using different genres of texts since complexity of texts and text quality are dependent on text genre (Beers et al. 2009: 185). Graesser et al. (1980: 288) compared the retention rate of narrative and expository texts and found that the former is recalled better than the latter by far. Allen et al. (1988) also compared four different genres in terms of comprehension: friendly letters, general articles, business letters and newspaper articles. Their results were in the same order from the most comprehended to the least (p. 168). Kirkness and Neill (2009) examined the language used in textbooks and academic journals in a comparative manner. Their findings indicate lower academic word percentage in textbook and thus they claim that text book was easier to read than the academic text (p. 12). McNamara et al. (2011) carried out similar research and found out that expository texts were more difficult to comprehend than the narrative ones (p. 242).

As a final remark on the comparison of genres, Kneer et al. (2011) conducted an interesting research which is noteworthy in shedding light on the perceptive aspect of genre comparison. They compared the texts of Goethe and *Rammstein* (a German metal band) in an empirical manner to find out which one is perceived as more aggressive and brutal – or in other words, negative. The design of the study involved the switching of author names of the texts when presented to the participants, who consisted of two groups as metal fans and non-metal fans. Their findings showed that when the non-metal fans were presented the poem by Goethe but told that it was written by *Rammstein*, they found it more negative. The results reveal that the knowledge of metal lead to a more negative perception of the lyrics although that was not actually the case. The study did not involve any lexical analysis but was based on emotional judgements of participants reading texts in two different genres. The results are remarkable both in terms of its possible connection to Swales' (1990) "discourse community" theory and the subjective nature of the genre concept. The shared values of metal fans as a "discourse community" might have allowed them a more informed perception of the song lyrics. On the other hand, the non-metal fans might as well have judged it based on their own shared values. Following the brief example involving a metal band, the genres in metal will be discussed in the following section.

2.1.2. Genres in Metal

Metal is one of the subgenres of rock which again branches into many other subgenres. It moves in many directions by a thousand different artists (Harris 2013). Each band comes up with such new or mixed ideas both musically and lyrically that no two bands are exactly alike even if they claim (or are perceived) to belong to the same subgenre. In *Episode 4* "*Never say Die*" of BBC's *Seven Ages of Rock*, rock writer Seb Hunter notes,

> Metal always gets bloated and self-satisfied and bored of itself. And it kind of just sits around going, 'Well, what next?' And of course, something always does come up (*Seven Ages of Rock, Episode 4 "Never Say Die",* 00:48:08).

More than 20 metal subgenres can be identified (Klepper et al. 2007: 9) from ballads to the most extreme forms (Tsatsishvili 2011). Hickam and Wallach (2011: 260) state that metal now encompasses innumerable subgenres from folk metal to industrial metal, gothic metal, funeral doom, etc. Considering the crossbred versions of these subgenres, the alternatives are theoretically infinite. This means that, unlike the way it is perceived, metal is not a monolithic entity but one consisting of many subgenres which have their own philosophy and characteristics (Riches 2011: 316).

2.1.2.1. Definitions

To begin with the definitions on metal genres, the term "genre" needs clarification in the metal context. Although the use of genre to refer to music styles is widely accepted, other views are present well. Gracyk (2016) deals with this question of appropriate terminology for metal in a comprehensive study. He delves into the use of the terms *genre*, *style* and *subculture* and argues that the genre approach, which is a social and subjective phenomenon, is not the most appropriate way to identify music styles. He proposes an historical approach which takes into account all the factors (inspiring bands, directions, etc.) affecting particular bands in the formation of a particular "genre". However, his argument might make more sense in the identification of the (usually pioneering) bands whose genres are harder to be distinguished, rather than well-acknowledged and least disputed bands. Out of the options he proposes, subculture seems more sociological whereas style is more musical.

The terminology of choice in the present paper will be *genre*, not necessarily because it fits the description of one particular scholar, but rather because it is a widely accepted term in both music and linguistics. Another reason is that *genre* could be regarded as a cover term to bridge the fields of linguistics and music (which is the focus of the study) since there are no musical "registers" in the generic sense (not to be confused with the term *register* related with the ranges of musical notes). *Subculture* also cannot be used synonymously with genre to refer to linguistic differences across texts. Texts produced by a particular subculture could feature various genres. *Genre*, in this sense, offers a more common understanding of classification or grouping of different kinds of music or texts.

The vast amount of subgenres metal music involves leads to diversity in definitions. Over the fifty years following its emergence, metal has become a big web of subgenres and the terminology with regard to the metal subgenres is much debated and often subjective. The term heavy metal embraces all kinds of different styles, lyrics and behaviors (Walser 1993: 3-4). Brian Turner, Music/Program Director at WFMU radio station, notes that metal always gets pushed into new and bizarre territories (*Heavy: The Story of Metal, Episode 4: "Seek and Destroy",* 00:27:42) which eventually complicates terminology as well as the musical styles. Attempts to establish solid definitions for heavy metal are quite controversial due to the changeability of definitions over time. Philips and Cogan (2009: XIX) state that there are disagreements between scholars, and also metal fans, about the exact definition of heavy metal and when it began.

Each band has a unique sound, which Weinstein (2000) calls "signature sound" and furthermore one band may differ across albums or even songs (p. 22). There may be different interpretations of the definition of metal in the course of time and across nations, publications and writers (Weinstein 2009: 20) as genre boundaries in popular music are not solid (Walser 1993: 4). Genres, over time, may merge with each other or split into other subgenres (Tsatsishvili 2011: 2). These statements could be viewed as a reflection of Chandler's (1997) remarks regarding the changeability of genres. The classification of metal subgenres, hence, rests on these aspects and may not be clearly and fully objectively specified.

In order to be able to grasp the dynamics and the nature of metal genres, their emergence might be of primary assistance. The emergence of genres involves a good deal of sociological and musical factors; however, Byrnside's (1975) theory, quoted by Deena Weinstein in her *Heavy Metal – The Music and Its Culture* (2000), on the emergence and aftermath of the genres is widely accepted by metal scholars. The theory lists the following three steps in genre development: "formation", "crystallization" and "decay" (p. 7).

In the formation phase, the differences between the newly emerged genre and its parent genre are not so clear. The distinction becomes clear in the crystallization phase, where a genre is acknowledged as a different one from the parent. The decay phase is where the style becomes "so predictable that both composer and audience begin to lose interest" (Byrnside 1975 qtd. in Weinstein 2000: 7). It is obvious from these remarks and the common perception of the world that virtually no band claims to name its music style before they begin to play. Weinstein's (2009) view of the naming of the genres is as follows:

> Genres begin before they are named. When several artists working in a new style find a common audience and mediators (such as the rock press, record companies and concert promoters) recognize the genre, it is in its period of crystallization. How long that phase lasts and what succeeds it vary. It may merely disappear from the scene, fragment into several other styles (p. 20-1).

This phenomenon is consistent with Swales' (1990) "discourse community" term as the emerging and the formation of the genres are facilitated by the individuals contributing to or taking part in them as shared values. One band can single-handedly take the credit for inventing a particular genre, such as *Venom* for black metal, whereas a group of bands, more often than not producing music in close proximities, collaboratively build a (sub)genre, such as *Swedish Death Metal*. Whichever way the (sub)genres might be created, they are not named 'at birth'. Virtually no band/artist go to a recording studio with a thought in mind to invent genre x, but rather the music is produced and the (sub)genre emerges subsequently. The classification of the genres is, then, realized in line with the steps and actors defined in the following section.

2.1.2.2. Classification of Metal Subgenres

Classifying music genres is a complex task since there are a lot factors at play some of which have been provided in the previous section. Although the genres are widely used by both the music industry and the consumers, no sound classification of musical genre has yet been proposed (Pachet and Cazaly 2000). In fact, the classification of genres is more necessary for the music industry than the fans (Walser 1993: 5) and music retailers are probably the most important producers of music taxonomies (Pachet and Cazaly 2000). Similarly, on the listener/consumer side, genre classification is the "most natural paradigm" to browse through music despite their ambiguity (Tsatsishvili 2011: 1).

This ambiguity is discernible in metal subgenres as well. Studies show that classification of metal subgenres requires insights as a fan. Weinstein (2000: 6) stresses that it is necessary to listen to metal through the codes with which competent listeners appreciate the music (see "discourse communities" proposed by Swales [1990]) in order to be able to conduct research on metal. She clarifies the arising question of who is 'eligible' for labeling the bands into this genre and gives a brief account of the characteristic of metal genres as follows:

> There is no legitimate, established authority, like the French Academy, that decides on the correct name for a musical style or genre, and there is no consensus on what to call the subgenres that crystallized within heavy metal.[…] (p. 45)

- Bands do not stay in the same genre in their careers – even sometimes in their albums.
- Each band, as well as each song, can be seen to have unique features.
- There are no councils or dictionaries to standardize the genres.
- No umpires to assign works to one category.
- Retail stores can place a record into any bin they like.
- Fans, publicists, and rock journalists have the same freedom to be arbitrary, inconsistent, or merely perverse (p. 287)

With regard to the same issue, Hickam and Atwood (2008) argue on what can and cannot be classified as metal and who has such authority and experience for this task. They suggest that this matter should be left to historians and other researchers to judge, prioritize, and classify.

As regards the research involving metal, Arnett (1996), for example, selected his own corpus based on his perception of the popularity of the bands and defended his selection by asserting that most people will accept them despite the never-ending arguments among the metal fans as to which band is metal and which is not. His study yielded different results compared to Weinstein (1991) and Walser (1993) since their classification of the bands were different in the first place. Arnett (1996: 174) remarks that the authors concerned considered the bands metal whereas he treated them as hard rock.

As another example, Logan et al. (2004) loosely assigned each artist to a particular genre based on the information on www.allmusic.com, a comprehensive online music website. This approach may yield different results during analyses especially where metal music is concerned. *Morbid Angel*, a world-renowned death metal band, is listed as "Pop/Rock" on www.allmusic.com. This is not completely wrong; yet, too general to allow listeners to distinguish the band from other artists or subgenres. Furthermore, the fact that they are a death metal band is observable in the academic literature (see Weinstein 2000, Piccoli 2003, Morrison 2006, Davisson 2010, Mishrell 2012, and Phillipov 2012). Bértoli-Dutra (2014) carried out an MD Analysis of pop and rock bands and her account of possible interpretations of the same artists is of note here:

> There is no consensus for the categorization of music genres or styles. For this study, I consulted six different sources (Charlton 2008; Starr & Waterman 2007; Billboard charts; Rolling Stone Magazine; allmusic.com; and musicimprint.com), and very often the same source offered different categorizations for the same artist. The band Aerosmith, for instance, is classified as hard rock, blues rock, and heavy metal (Billboard), hard rock (Charlton 2008; Starr & Waterman 2007), pop/rock (allmusic.com and musicimprint.com), and hard rock, heavy metal, album rock, pop rock, arena rock, and pop metal (Rolling Stone Magazine). The Beatles are classified as rock and roll and rock/pop (Billboard), rock (Charlton 2008; Starr & Waterman 2007), pop/rock (allmusic.com and musicimprint.com), and early pop/rock, rock and roll, Am pop, pop/rock, British psychedelic and folk rock (Rolling Stone Magazine). I chose the most recurrent classification for each artist. Therefore the band Aerosmith was considered hard rock and The Beatles pop rock for this study (p. 174).

It is true that genre boundaries are not solid or clear (Walser 1993: 4). There may not be a pure scientific division of subgenres; nonetheless, this does not mean that research cannot be made into these *de facto* music genres. The above-mentioned examples are significant indicators of the possible differences between the interpretations of music bands as far as their genres are concerned; thus, care must be taken to draw the line between the genres. In order to achieve more objective distinction, the emergence and history of (sub)genres, the categorization principles and related media and literature coverage should be taken into consideration. The classification of genres is usually carried out by "mediators" as proposed by Weinstein (2009: 20). There is another institution to name the genres, though, which is perhaps, more important than all the other mediators combined: the bands themselves.

2.1.2.2.1. Categorization by Bands

Bands usually focus on playing the music they want and refrain from categorizing it. However, it is rarely the case that they are not associated with any genres. Some agree with the mediators' labeling, some develop their own ones, and some reject the associations.

Metal bands are usually proud of being metal bands and bands like *Manowar*, for example, are never tired of overtly expressing their endless love for the genre through their lyrics. Rock writer Malcom Dome asserts that *Judas Priest*, another major metal pioneer, was the first band to embrace the term 'heavy metal' and Glen Tipton, the guitarist for *Judas Priest*, states that they have always been proud of the term (*Seven Ages of Rock, Episode 4: "Never Say Die",* 00:21:12-00:23:23).

Slayer is a good example of straightforward genre labeling as they are generally acknowledged as a thrash metal band. The band takes pride in the association and Kerry King, the guitarist of *Slayer*, clearly says: "Yeah, it [thrash metal] is us" (*Get Thrashed: The Story of Thrash Metal*, 00:06:44). At one with Kerry King, Dave Mustaine of *Megadeth* remarks that the term thrash metal fits them very well (*Metal Evolution, Episode 6: "Thrash",* 00:29:46). An example of a self-developed genre could be the one put forward by *The Dillinger Escape Plan* who prefers to be referred to as "a creation merging new-school hardcore, progressive metal, and free jazz (www.dillingerescapeplan.com)" (Pieslak 2007: 243).

However, genre associations are not always as easy for some bands as they are for *Slayer*, *Megadeth* or the self-labelled ones such as *The Dillinger Escape Plan*. This is even more the case when the bands concerned change their styles across – or even within – their albums throughout their careers. A perfect example for this would be *Metallica*, the very band which has produced both *Whiplash* and *Nothing Else Matters*, the former being a fast thrash metal song while the latter an acoustic ballad. James Hetfield, the frontman of *Metallica* acknowledges the phenomenon and states that they do not fall into any category as, for instance, "Fade to Black" are "Sanitarium quite different from each other (Gore 1989 qtd. in Pillsbury 2006: 33).

Bands like *Metallica*, which are the creators of the genres or those not fitting in one, suffer from mislabeling and some make this known through interviews and other means. This is not the only case of rejecting the association, though. Many metal bands shun the term 'heavy metal' and metal is sometimes regarded as "limiting" (Konow 2002: XII). John Lord of *Deep Purple*, for instance, underlines that he did not like the term heavy metal and never applied it to *Deep Purple* (*Heavy Metal Britannia,* 01:07:12). Phil Collen of *Def Leppard*, another British band which is considered among the first bands of metal, refuses this association explaining that they nothing to do with metal and that they were more *Duran Duran* than they were *Black Sabbath* (*Heavy: The Story of Metal, Episode 2: "British Steel",* 00:37:02).

As mentioned previously, bands' rejection of labeling usually occurs when their genres are not clear enough. This is mostly the case with the pioneering bands which happened to make their entries to the scene in the formation phase of a genre. Two examples for this case could be *AC/DC* and *Led Zeppelin*.

AC/DC's genre is debatable as the band stands on the borderline between hard rock and heavy metal (Moberg 2009: 110). As Christe (2003: 58) puts it "*AC/DC* became the stepping-stone that led huge numbers of hard rock fans into heavy metal perdition". He also names the band's *Back in Black* album, which was released in 1980, among "The Best 25 Heavy Metal Albums of All Time" (p.

339). Nevertheless, the band's own perception about their music is not the same. Angus Young, the guitarist and the founding member of *AC/DC*, says that he hates the term heavy metal (Walser 1993: 3).

There are many scholars and authorities proposing that *Led Zeppelin* belong to this genre or that (Buts and Buelens 2008, Morrison 2006, Klepper et al. 2007, Phillips and Cogan 2009). But what do the band members think about themselves? Apparently, they are not happy with being associated with heavy metal. Sam Dunn, the producer of *Metal: A Headbanger's Journey* and the *Metal Evolution* series, explains his meeting with Robert Plant and Jimmy Page that never happened as a result of the band's disinclination to be associated with the term heavy metal (*Metal Evolution, Episode 3: "Early Metal Part 2: UK Division",* 00:07:08).

As the creators of their own music, the bands rightfully have their say about which genre they belong to. However, ironically enough, they do not have the exclusive authority on their genres since they are not the only members of the "discourse community" of metal.

2.1.2.2.2. Categorization by the Public

"Artists boast about themselves, fans boast about artists, and artists boast about their appreciative and loyal fans" (Weinstein 2000: 142). Metal fans value and care about their music to a considerable extent. They follow their favorite bands to the end as long as the band in concern does not stray away from metal. They usually have zero tolerance for any kind of diversion and insist on the purity of metal. The fans also do not favor their favorite bands going commercial and/or too popular and might even give up listening to them if such a change occurs (Konow 2002).

As a reaction or response to such fan practice, the bands usually heed their fans' demand because it is important for the bands not to lose their fan base. "'That's not heavy metal' is the most damning music criticism a fan can inflict, for that genre name has great prestige among fans" (Walser 1993: 4). Some bands, however, choose not to consider their fans' expectations as much as their musical progression and brave their criticism or, in extreme cases, sever the ties between themselves and the metal community for the sake of new the dimensions they would like to explore in their career. These practices, whether positive or negative, demonstrate the fans' power over metal to an extent that they can not only name the genres but also affect the way they are produced. Therefore, fans' perception in genre categorization cannot be denied.

Fans are not the only members of the "discourse community" of metal and therefore, public categorization is not exclusively limited to their opinions, as music scholars, members of other bands and media ("mediators" [Weinstein 2009: 20]) are entitled to the notion as well. Although bands, which are relatively easier to categorize, do not cause much problems, the mediators naturally have their different opinions about controversial bands.

As a follow-up to the discussion of *Led Zeppelin*'s dissociation from heavy metal in the previous section, some mediators argue that the band is a heavy metal band (Buts and Buelens 2008, Phillipov 2012), influential on heavy metal (Morrison 2006: 105), the inventor of the genre (Asch 1999) etc. Bruce Kullick of *KISS* remarks that *Led Zeppelin* "are the godfathers of heavy metal" (*Heavy Metal: Louder Than Life,* 00:06:00). BBC's *Seven Ages of Rock*, on the other hand, uses a more specific genre to categorize the band: Stadium Rock (*Seven Ages of Rock*, *Episode 5: "We Are the Champions",* 00:00:46).

These examples display the active part mediators take in the categorization of metal bands. It is not a surprise, though, as most of the mediators are fans themselves. As the last actor in the classification of metal, a brief account of some of the notable scholars and mediators will be provided in the following section.

2.1.2.2.3. Notable Scholars and Mediators

The classification of metal subgenres is a blurry task as noted in the previous sections. The bands and mediators put forward their claims both in harmony or disagreement. In order to reach a consensus, it might prove useful to take into account the arguments of the mediators and scholars as well as the bands and the fans. Metal is a rising field of study in the scientific community and is mainly spurred by metal scholars who are fans themselves. Deena Weinstein's *Heavy Metal: A Cultural Sociology* (1991) is the first fully scholarly work on metal and the following one in Robert Walser's *Running with the Devil* (1993) (Brown 2011: 215-6). Spracklen, Brown and Kahn-Harris (2011: 211) name

Deena Weinstein, Nicola Masciandaro and Keith Kahn-Harris as notable metal scholars and point out to the increase in PhD researches in metal, which are conducted by young metalhead scholars.

More books and academic papers on metal have come out since the 2000s. Some of which that have been used in the present paper include, *Bang Your Head: The Rise and Fall of Heavy Metal* (Konow 2002), *Sound of the Beast: The Complete Headbanging History of Heavy Metal* (Christe 2010), *Death Metal Music: The Passion and Politics of a Subculture* (Purcell 2003), *Damage Incorporated: Metallica and the Production of Musical Identity* (Pillsbury 2006), *Encyclopedia of Heavy Metal Music* (Phillips and Cogan 2009), *Death Metal and Music Criticism: Analysis at the Limits* (Phillipov 2012), etc. Brown (2011: 231) associates the increase in metal studies with that in genre and subgenre specific terminology.

In addition to the academic publications on metal, there are some other remarkable mediators from the media that have been extensively referred to in the present paper. 2000s saw very pioneering and challenging movies and documentaries produced by a metal scholar – a title he deserves in spite of his anthropologist background – Sam Dunn. Sam Dunn is a Canadian anthropologist who has produced many documentaries and movies about the history and classification of metal, both in general and specific to some bands. Sam Dunn's first movie, *Metal: A Headbanger's Journey (2005)*, could be considered a milestone in metal history on account of its attempt to classify metal subgenres into clear-cut divisions. He proposes 24 different subgenres under which he cites bands ranging from 4 to 14 in number (see Table 2). The original genealogy showing the relationship of subgenres to each other can be found in Appendix A.

Table 2: Metal subgenres taken from Metal: A Headbanger's Journey

Subgenres	Bands
Early Metal (1966−1971)	Cream, Jimi Hendrix, Blue Cheer, Deep Purple, Led Zeppelin, MC5, Mountain, The Stooges, Black Sabbath
Original Hard Rock (1974−1979)	Thin Lizzy, Blue Öyster Cult, Aerosmith, Ted Nugent
Shock Rock (1968−1983)	Arthur Brown, Alice Cooper, New York Dolls, Kiss, Ozzy Osbourne, W.A.S.P.
Early Punk (1976−1979)	The Ramones, The Damned, Sex Pistols, The Clash, The Dead Boys
Power Metal (1976−present)	Scorpions, Judas Priest, Rainbow, Accept, Manowar, Dio, Yngwie J. Malmsteen, Helloween, Blind Guardian, HammerFall, Primal Fear
New wave of British Heavy Metal (1979−1983)	Motörhead, Saxon, Iron Maiden, Angel Witch, Girlschool, Tygers of Pan Tang
Progressive Metal (1970−present)	Uriah Heep, Rush, Queensrÿche, Savatage, Fates Warning, Voivod, Dream Theater, Meshuggah, Symphony X, Evergrey
Glam Metal (1973−1990)	Slade, Sweet, Hanoi Rocks, Mötley Crüe, Twisted Sister, Poison, Cinderella, Skid Row
Pop Metal (1978−present)	Quiet Riot, Van Halen, Whitesnake, Def Leppard, Europe, Dokken, Lita Ford, Ratt, Guns N' Roses, Winger, Warrant, Doro, The Darkness
Stoner Metal (1982−present)	Witchfinder General, Trouble, Candlemass, Cathedral, Kyuss, Today Is The Day
Original Hardcore (1980−1986)	Agnostic Front, D.O.A., The Exploited, Bad Brains, Misfits, GBH, Discharge, Dead Kennedys, Minor Threat
Thrash Metal (1983−present)	Metallica, Slayer, Anthrax, Megadeth, Exodus, Overkill, Kreator, Destruction, Sodom, Testament, Nuclear Assault, Death Angel, Pantera, Sepultura, Children of Bodom
First Wave of Black Metal (1981−1986)	Venom, Bathory, Mercyful Fate, Celtic Frost
Norwegian Black Metal (1990−present)	Mayhem, Darkthrone, Immortal, Gorgoroth, Emperor, Satyricon, Enslaved, Dimmu Borgir,

Subgenres	**Bands**
	Cradle of Filth
Grindcore (1987−present)	Napalm Death, Carcass, Repulsion, Exhumed, Extreme Noise Terror, Cephalic Carnage, Brutal Truth
Death Metal (1985−present)	Possessed, Death, Morbid Angel, Obituary, Deicide, Cannibal Corpse, Immolation, Autopsy, Nile, Dying Fetus
Swedish Death Metal (1990−present)	Grave, Entombed, At the Gates, Unleashed, Dismember, Arch Enemy, Soilwork, In Flames, Dark Tranquillity, The Haunted
Goth Metal (1990−present)	Paradise Lost, Tiamat, Therion, Type O Negative, My Dying Bride, Anathema, Theatre of Tragedy, Katatonia, Opeth
Metalcore (1985−present)	Corrosion of Conformity, Suicidal Tendencies, Dirty Rotten Imbeciles, Machine Head, Stormtroopers of Death, Hatebreed, The Dillinger Escape Plan
Grunge (1988−1993)	Green River, The Melvins, Soundgarden, Mudhoney, Nirvana, Alice in Chains, Mother Love Bone, Stone Temple Pilots, Pearl Jam
Industrial Metal (1988−present)	Ministry, White Zombie, Godflesh, Nine Inch Nails, Fear Factory, Marilyn Manson, Static-X
Hard Alternative (1985−present)	Faith No More, Jane's Addiction, Prong, Living Colour, The Smashing Pumpkins, Rage Against the Machine
Nu Metal (1994−present)	Biohazard, Korn, Slipknot, Limp Bizkit, Godsmack, Coal Chamber, System of a Down, Disturbed, Kittie
New Wave of American Metal (2000−present)	Shadows Fall, Lamb of God, Darkest Hour, Chimaira, Killswitch Engage, Unearth, God Forbid

Despite being a good classification attempt and probably a work which was the first of its kind, it has its shortcomings as well. For example, the German hard rock/heavy metal act *Scorpions* is listed under “Power Metal” whereas the literature mostly considers the band as a “Traditional Heavy Metal”, “Second

Generation of Heavy Metal", "Classical Heavy Metal" or even "Proto Metal" band (Strother 2013, Walser 1993 and Christe 2003). Another mistake was assigning the German thrash metal act *Sodom* under death metal.

Sam Dunn's *Metal Evolution* came out between 2011 and 2014. This series is sequel to his first movie, *Metal: A Headbanger's Journey*, and it initially consisted of 11 episodes each of which depicts a single subgenre. The episodes are *Pre Metal, Early Metal Part 1: US Division, Early Metal Part 2: UK Division, New Wave of British Heavy Metal, Glam, Thrash, Grunge, Nu Metal, Shock Rock, Power Metal and Progressive Metal.* This series definitely took *Metal: A Headbanger's Journey* many steps ahead by expanding the coverage in 11 episodes. Sam Dunn tries to be neat in the classification of genres, yet at the same time he admits that it is not an easy task. For example, he claims that it is very difficult to categorize *Pantera*, which is called thrash metal by some and groove metal by others (*Metal Evolution, Episode 8: Nu Metal*, 08:00). The series also features a major correction with regard to the German *Sodom*'s place. The band was categorized as death metal in *Metal: A Headbanger's Journey* whereas it was re-categorized under thrash metal, a correction that was necessary and accurate for the band concerned as it is widely acknowledged as a thrash metal band (see Weinstein 2000: 187 and Davisson 2010: 179).

Being one of the most elaborate works on the classification of metal, *Metal Evolution* series was an important resource for the present study. However, it still has its flaws, some of which are classifying *Rammstein* and *Slipknot* as "shock rock" and *Scorpions* as "power metal". *Slipknot* is commonly accepted as a nu metal act (Perrone 2010 and Ailes 2016). *Rammstein*, on the other hand, is a *Neue Deutsche Härte* / industrial metal band (Phillips and Cogan 2009, Van Elferen 2009). The case of *Scorpions* was discussed above.

Metal Evolution series featured detailed subgenre charts as an extension to the Metal Genealogy (see Table 2 and Appendix A) which was first initiated in *Metal: A Headbanger's Journey*. These charts contained some exemplary bands for each subgenre, which were either the pioneers or the most noteworthy bands of the said subgenre – or both. There was not a chart for each and every one of the subgenres, but only the most significant ones were covered. Table 3 shows the bands placed under the major subgenres, New Wave of British Heavy Metal, thrash metal and death metal. The original listings as screenshots from the documentary are available in Appendix B.

Table 3: New Wave of British Heavy Metal, thrash metal, death metal as placed in the Metal Genealogy in Metal Evolution series.

New Wave of British Heavy Metal	**Thrash Metal**	**Death Metal**
Motörhead	Anvil	Possessed
Def Leppard	Metallica	Death
Quartz	Slayer	Autopsy
Saxon	Anthrax	Morbid Angel
Iron Maiden	Megadeth	Obituary
Tygers of Pan Tang	Pantera	Cannibal Corpse
Diamond Head	Exodus	Deicide
Angel Witch	Overkill	Immolation
Girlschool	Kreator	Vader
Raven	Destruction	Six Feet Under
Fist	Sodom	Kataklysm
Holocaust	Sepultura	Dying Fetus
Tank	Testament	Nile
	Death Angel	Amon Amarth

Sam Dunn and *Metal Evolution* series is not a constant classification and it is open to changes and additions. Dunn fired up a new discussion in November 2015 and asked all the metal fans for their contributions to some subgenres. They referred to fan comments on Facebook and other platforms. Eventually, they ended up updating and adding the following subgenres in the metal tree: Metalcore, crossover, doom, industrial metal, mathcore, folk metal and progressive metal. The ‘evolution’ of metal tree can be seen in the photos shared by Sam Dunn on Banger Facebook page which are available in Appendix C. These updates and corrections apparently addressed to Rammstein’s disputed status, and the band was eventually moved from “shock rock” to “industrial metal” (see Figure 10 in Appendix C).

Despite not being a scientific website for metal, *Encyclopaedia Metallum*, residing at www.metal-archives.com, is perhaps the most comprehensive and reliable source for metal music. It was established by two Canadian metal fans, whose nicknames are Morrigan and Hellblazer, in July 2002. The website aims to be the biggest metal database and is in fact quite so despite the strict rules the admins apply regarding the admission of bands. This strict policing results in the exclusion of many bands which give rise to debates in online platforms. For example, the German act *Rammstein*, playing *Neue Deutsche Härte* / industrial metal, is not and will not be included as stated in the rules of the website. Some changes over the years occur, though, partly because the bands in concern change their styles into metal and partly as a result of discussions about 'metalness'.

An example could be *Buckethead*, a guitar virtuoso who is known for his anonymity since he prefers to wear a mask on his face and a bucket over his head. He is perhaps the most productive musician in the music industry let alone the metal community with 174 full-length albums in his career as of May 2015, 61 of which were released in 2014 according to Encyclopaedia Metallum. His music style spans from rock to hard rock, from heavy metal to shred which makes is very difficult to categorize. *Buckethead* was not included in the *Encyclopaedia Metallum* for years until his acceptance on 24/07/2014 (*Buckethead – Encyclopaedia Metallum*: The Metal Archives).

Despite the fact that the website has no voiced intention of being scientific, *Encyclopaedia Metallum* is one of the references used by the scholars researching metal music. Kahn-Harris (2013) presents the website to bear plentiful information on metal and Tsatsishvili (2011) also mentions the website in his study. The website lists 115,220 approved bands in its database as of March 28th, 2017, 4:14 am EDT (Stats – *Encyclopaedia Metallum*: The Metal Archives). In his paper on extreme metal, Davisson (2010: 180) describes death metal and its specifications through references to *Encyclopaedia Metallum* on the matter of providing the number of death metal bands.

It is aimed with the inclusion of noteworthy academic and non-academic mediators to reach a more balanced classification of subgenres. The three metal subgenres used in the present paper will be introduced and detailed based on references to the above-mentioned resources and mediators.

2.1.2.3. Comparison of Metal Subgenres

Comparing metal subgenres is quite a straightforward action for the metal community. As it is usually not the case with the non-metal community, some basic aspects regarding the comparison of metal genres need to be addressed. It can be argued that the comparison of metal subgenres, much like genres in general, is based on form and content. Here, form basically refers to the musical features and imagery, whereas content is related with lyrics.

The categorization of metal subgenres is usually based on musical form, which mainly manifests itself in different instrument playing techniques, speed and vocal types. Metal bands also possess different images mostly reflecting the characteristics of their subgenres. The difference is more observable in clothing and album covers. The difference between genres or bands can be observed in lyrics, too, although not as dominant or deterministic in genre separation as the form-related aspects. A death metal band might write lyrics on war while another one writes on gore. However, this is not enough to categorize these bands as belonging to different genres. If their guitars are tuned alike, they play lightning fast, and growl rather than sing, they are both death metal bands regardless of their choice of lyrical themes. Similarly, a thrash metal band and a death metal can both write war-themed lyrics, but this does not qualify them into the same genre if they sound different from each other.

On a broader scale, Fell and Sporleder (2014: 624) point out that some genres display lexical differences among each other in their lyrics, e.g., “Rap (dominant slang use), Reggae (Jamaican slang, Rastafarian terms), Religious (religious terms) and Metal (death, violence)”. Metal lyrics were highly controversial until recently. They were perceived as “blatant and violent” (Hinds 1992: 154), “serious and pessimistic” (Arnett 1996: 44). With the advance of studies in metal and broadening of the scope with the inclusion of subgenres rather than treating metal as a whole, a deeper understanding and differentiation of lyrics have emerged.

Many indications of differences between the lyrics of different metal subgenres in particular can be found in the literature (Walser 1993, Morrison 2006, Buts and Buelens 2008, Strother 2013, etc.). Metal lyrics are typically divided into two categories as “Dionysian” and “Chaotic” (Weinstein 2000: 35). Dionysus is the God of wine and of harvest in the Greek mythology (Roman and Roman 2010: 137). Therefore, Dionysian themes mostly consist of sex, drugs and rock and roll – the three major themes in hard rock and heavy metal as well as the appraisal of rock music. Such lyrics focus on pleasures and daily life. On the other hand, Chaotic themes, as the name suggests, deal with the darker aspects of life such as “disorder, conflict, opposition, […] contradiction […] monsters, the grotesque, mayhem, […] disaster, injustice […] resistance, rebellion, and death” (Weinstein 2000: 39). Such lyrics have more oppositional stance and usually confront the norms of the society through the expression of taboos and rebel ideas. Chaotic themes are highly dominant in extreme metal subgenres such as thrash metal, death metal, black metal, etc.

Arnett (1996) compared mainstream metal bands (*Judas Priest*, *Iron Maiden* and *Ozzy Osbourne*, which fall under the heavy metal category in this paper) with speed/thrash metal bands suggesting that these genres have lyrical differences. Buts and Buelens (2008: 38) point out that “thrash metal and extreme metal do emphasize different lyrical, visual and idiomatic elements than the [New Wave of British Heavy Metal]” and add that these genres could be considered the extremer versions of the New Wave of British Heavy Metal.

As mentioned above, although the lyrics show differences across genres, they cannot be the sole means of comparison as there are many variations available. For example, the British death metal band *Bolt Thrower* and the German thrash metal act *Sodom* both mostly write war-themed lyrics. Another possible problem with comparison based on lyrics could be the level of seriousness and priority dedicated to lyrics by the artists themselves. Some bands excel with their lyrics as much as their music, such as *Iron Maiden* and *Death*, while some others do not care about lyrics at all and write them simply to match with music.

Harris (2013) believes metal to be a very literate culture when compared to other pop genres and asserts that there are many literary references in metal songs. *Iron Maiden* and *Rush*, for example, have “eloquent and meaning-charged lyrics” (Weinstein 2000: 123). On the other hand, some other acts do not prioritize their lyrics and focus more music. *Van Halen*, for example, states that he had no idea what their lyrics were (Interview with *Van Halen* in Musician 1987 qtd. in Walser 1993: 26). In a 1989 interview, Ozzy Osbourne, the singer of *Black Sabbath* which is one of the founders of heavy metal, expands on the subject as follows:

> [singing] “Love grows where my Rosemary goes...” – (dann… [mimics guitar sound]). It just doesn’t fit. It’s heavy so if it’s what you call “heavy metal”, then you’ve got to put a pretty heavy lyric to it. I suppose writing about the darker forces and the darker sides or whatever fits the music. You would hardly write about a love song to that kind of heaviness (*Heavy Metal Britannia,* 01:01:19).

In summary, it is usually the music that determines the metal subgenres and therefore the comparison of metal subgenres is mostly carried out based on musical features. It is quite observable in the fact that the members of the metal community can distinguish genres by sound without even knowing the lyrics of the songs.

Specification and comparison of metal subgenres, both across and within bands, entail a great number of factors. Every subgenre differs from each other by overt or slight details in music. Some bands even play in different subgenres across their albums over time, which makes categorization even more complex and the genre names longer. As the comparison of the subgenres in the present study will be based on the lyrical aspects, musical details will be kept to a minimum and focus will be on bands with no or very little genre shifts in their careers. The differences of the metal subgenres used in the present study with regard to their content and form are detailed in the following section.

2.1.3. The Three Metal Subgenres Used in the Current Study

Three major subgenres of metal music used in the present study, heavy metal, thrash metal and death metal, will be introduced in the following subsections. The main focus will be on the emergence and main characteristics of the subgenres. Exemplary bands for each subgenre will also be provided. The representatives of the respective subgenres and the methodology in their selection will be discussed in METHODOLOGICAL APPROACH.

2.1.3.1. Heavy Metal

Heavy metal is a subgenre of rock. The musical features described to be possessed by heavy metal are “strong melodic strain”, an “emphasis on heavy, rhythmic bottom sound” (Weinstein 2000: 55); “fiery instrumental technique, guitar riffs, powerful vocals, a warrior-like energy” (Morrison 2006: 105); “speed, the particular harmonic language emphasizing tritones and flatted seconds” (Pillsbury 2006: 4).

The origins and pioneers of heavy metal are often disputable although there is a general consensus on a few bands. Philips and Cogan (2009), for instance, consider *Led Zeppelin* and *Black Sabbath* to be the founders of heavy metal in their *Encyclopedia of Heavy Metal Music*, and disregard some other acts which are considered so in other resources. The case of *Led Zeppelin* was discussed in the previous sections mostly stressing on their undefined generic position. Excluding *Led Zeppelin*, *Black Sabbath*, formed in 1969 and released their self-titled first album in 1970, could be considered the pioneer of heavy metal (Berger 1999, Walser 1993 and Christe 2010).

Heavy metal got bigger, driven by other British bands of the time which formed the New Wave of British Heavy Metal (NWOBHM). The terms NWOBHM, heavy metal and metal was briefly introduced earlier in Classification of Metal Subgenres and it is necessary to shed some light on these terms as the first step. NWOBHM exclusively represents the British bands playing a heavier form of rock in the 1970s. It is known for its melodic structure and punk influence (Klepper et al. 2007 and Buts and Buelens 2008). Rock journalist Geoff Barton, who used the term NWOBHM in music press, if not actually coined it, explains what it means:

> I’d like to take credit for inventing the term “NWOBHM”--but once again I believe it was a “Big Al” Lewis brainstorm. […] in truth the NWOBHM was a catchall title for a melting pot of bands with an immense variety of musical styles. (Ward).

NWOBHM was the “real boom” in heavy metal (Walser 1993: 11). Christe (2003: 46-7) lists *Motörhead*, *Saxon*, *Iron Maiden*, and *Judas Priest* under NWOBHM section. His next list is titled “Classic Heavy Metal” explained as “the definition of heavy metal with a capital HM—bands that were the cream of the

NWOBHM" (p. 72). This statement acknowledges the terms 'Heavy Metal' and NWOBHM synonymously. The listing underneath "Classic Heavy Metal" includes *Iron Maiden*, *Def Leppard* and *Saxon*, which were also mentioned in "The New Wave of British Heavy Metal" chapter, suggesting that the terms are intertwined at some point. The first usage of heavy metal was in 1984 whereas metal as a standalone term came up in 1990 (Brown 2011: 221). Weinstein (2000: 55-6) stands a similar ground with her following proposition:

> By spawning subgenres heavy metal became part of a wider cultural complex, just as it had earlier erupted from a wider complex. The new formation, called "metal" here, includes both heavy metal and subgenres within itself, each of its members being defined by distinctive codes that contain some or all of the elements of heavy metal's code. Heavy metal persists, after the subgenres crystallized, as classic metal, carrying on the core of the musical culture.

Strother (2013: 11) draws the line by stating that the term heavy metal has recently been reserved for classical metal and "metal" itself refers to all kinds of metal. This approach was adopted by Buts and Buelens (2008) and they used the term metal to cover all subgenres of metal and restricted the use of heavy metal to early pioneering bands and the NWOBHM.

To sum up, heavy metal came up as a distinct genre in 1970s as a result of early pioneering bands and the NWOBHM bands. It, then, became a subgenre of metal following the emergence of various other metal subgenres. The term 'heavy metal' is now used to define those bands and various others which were formed later playing in the same style.

2.1.3.2. Thrash Metal

Thrash metal could be explained as the next step from heavy metal in terms of speed, heaviness and lyrics. Thrash metal is often also referred to as speed metal (Arnett 1996, Weinstein 2000, Morrison 2006, Buts and Buelens 2008 and Strother 2013), but this study will refrain from this usage due to the fact that speed metal is acknowledged as a different genre at the same time (Pieslak 2007, Klepper at al. 2007 and Philips and Cogan 2009).

Musically, thrash metal is a faster version of heavy metal and power metal, bearing resemblance to hardcore and punk with regard to tempo, featuring complicated arrangements and changes in meter (Walser 1993, Christe 2003, Philips and Cogan 2009). Lyrically, thrash metal bands address violence, death and madness in a way usually less mystical than *Iron Maiden* (Walser 1993: 157). In Episode 6 "*Thrash*" of Sam Dunn's *Metal Evolution*, thrash metal lyrics are described as graphic and dealing with warfare, human suffering and even serial killers (*Metal Evolution, Episode 6: "Thrash",* 00:02:29).

The emergence of thrash metal in early 1980s (Morrison 2006: 107 and Strother 2013: 220) is interpreted in two different ways: a) as a continuum or an expansion of NWOBHM, b) as a reaction to it. Buts and Buelens (2008) hold on to both interpretations. It is quite often associated with punk due to its speed, noise and violence; however, the two genres differ a lot in lyrics and music (Walser 1993: 14). The subgenre is an "American-made punked-up response to [NWOBHM]" (Weinstein 2000: 288) and the major difference of thrash metal as compared to heavy metal is the increase in tempo. In short, it is widely accepted as a mixture of heavy metal, hardcore and punk (see Strother 2013, Buts and Buelens 2008, Hickam and Wallach (2011).

Konow (2002) proposes England as the birthplace of thrash metal and his claim is based on his argument that *Motörhead* is the pioneering speed metal band (cf. Christe 2003: 46-7); yet it might not hold true when a distinction between speed metal and thrash metal is observed as is the case with the present study. Christe (2003: 223) remarks on the geography by pointing out that "heavy metal was born in England and thrash metal had its heart in San Francisco". In this respect, the geographical origins of the genre are San Francisco, Bay Area and Los Angeles and the pioneers are *Metallica*, *Slayer*, *Testament*, *Exodus*, *Megadeth*, and *Possessed* (Walser 1993 and Arnett 1996) although *Possessed* fits better into the genre of death metal. *Metallica*, *Megadeth*, *Slayer* and *Anthrax* are referred to as the "Big Four" and considered the originators of the genre (Weinstein (2000, Christe 2003, Davisson 2010 and Phillipov 2012).

Whether it was a continuum of or a rejection to NWOBHM, thrash metal has earned its place in the history of metal. One of the main reasons for this is the fact that the acts that are cited as the pioneers/originators of thrash metal, *Metallica*, *Megadeth*, *Slayer* and *Anthrax* – the Big Four – in particular, are at the same time universally accepted as the greatest metal bands, not only thrash. The

Big Four showed the way of thrash metal (Piccoli 2003) and that way led to a huge fanbase, success and eventually history.

2.1.3.3. Death Metal

Death metal is one of the heaviest music genres in the world. The main features of death metal are down-tuned and/or detuned guitars (Piccoli 2003, Morrison 2006: 107, Philips and Cogan 2009: 62), ample bass (Mishrell 2012: 15-6) tremolo-picking and unexpected changes in riffs (Frandsen 2011: 39), fast and hectic drumming (Morrison 2006: 107 and Frandsen 2011: 39), double bass drums (Morrison 2006: 107), blast beats (Frandsen 2011: 39 and Kitteringham 2014: 82) and screaming, growling, guttural (Piccoli 2003 and Philips and Cogan 2009: 63), non-pitched and unintelligible vocals (Berger 1999: 163, Stelzner, Morrison 2006: 107) which may sound like roaring to non-metal fans (Frandsen 2011: 37).

Death metal could be considered a reaction to or an extension of thrash metal much like thrash metal's relation to heavy metal. To put it very briefly, "death metal was created out of thrash metal" (Harris 2000: 17). Morrison (2006: 107) claims that it was *Slayer* who gave the inspiration to the genre. Death metal emerged during the heyday of thrash metal as one step forward down the lane of musical complexity of thrash (Buts and Buelens 2008: 68) "elevating the extremity of bands like *Venom* and *Slayer* to another level" (Strother 2013: 223). Christe (2003: 222) defines a distinctive aspect of death metal as "death metal bands used speed and intensity to squeeze an album's worth of ideas into a single song" as the bands constantly try to outdo each other in terms of extremity (Purcell 2003: 48).

The name of the subgenre appears to come from the band *Death* and/or the song *Death Metal* by *Possessed* in their 1984 demo and 1985 album *Seven Churches* (Philips and Cogan 2009: 62-3). Monte Conner, the president of Nuclear Blast Entertainment (one of the major heavy metal record companies based in Germany), notes in the interview made by Sam Dunn that Chuck Schuldiner is the artist who started death metal with his band *Death* (*Metal Evolution: "Extreme Metal: The Lost Episode",* 00:24:19).

Death metal is now the biggest metal genre in the world (Weinstein 2000: 288-9) and it is "arguably the most musically varied" (Kitteringham 2014: 82) owing to the vast number of its subgenres. Davisson (2010: 180) indicates the same phenomenon with reference to *Encyclopaedia Metallum* which listed 11,756 active death metal bands as of 2010. This number is 20,919, as of March 2017, and counting (Advanced Search - *Encyclopaedia Metallum*: The Metal Archives).

Possessed and *Death* are considered the pioneers of death metal (Strother 2013: 224). Other exemplary bands that are cited in the literature are *Morbid Angel* (Weinstein 2000, Piccoli 2003, Morrison 2006, Davisson 2010, Mishrell 2012, Phillipov 2012), *Dismember* (Weinstein 2000), *Obituary* (Weinstein 2000, Piccoli 2003, Morrison 2006, Davisson 2010, Mishrell 2012, Phillipov 2012), *Cannibal Corpse* (Weinstein 2000, Christe 2003, Piccoli 2003, Morrison 2006, Philips and Cogan 2009, Davisson 2010, Mishrell 2012, Phillipov 2012, Strother 2013), *Entombed* (Weinstein 2000) and *Six Feet Under* (Weinstein 2000, Piccoli 2003). In her *Death Metal and Music Criticism Analysis at the Limits (2012)*, Michelle Phillipov's corpus of death metal bands consists of the significant acts in the genre: *Carcass*, *Cannibal Corpse*, *Death*, *Deicide*, *Morbid Angel*, *Napalm Death*, and *Obituary*.

Death metal offers the most obscene and explicit lyrics in popular music. Purcell (2003: 39) states that death metal lyrics are very offensive yet often poorly written (especially by non-native English speaking bands). This argument is, in fact, acknowledged by some of the musicians, too. For example, John Tardy – the vocalist for *Obituary* – admits to making up the lyrics to match the song (Mudrian and Peel 2004: 144-5). On the other hand, Phillipov (2012: 89) gives credit to death metal lyrics as they play an important role in the aesthetics of the subgenre which is observable in the fact that death metal lyricist often try to outdo each other.

Death metal lyrics have a narrower range of themes compared to thrash metal (Weinstein 2000) and they usually deal with morbidity (Piccoli 2003), violence (Arnett 1996, Buts and Buelens 2008, Philips and Cogan 2009, Frandsen 2011), gore (Weinstein 2000, Purcell 2003), decay (Weinstein 2000), murder (Frandsen 2011) injury (Buts and Buelens 2008), death (Weinstein 2000, Arnett

1996, Stelzner, Buts and Buelens 2008, Frandsen 2011), suffering (Philips and Cogan 2009), anger and dominance (Stelzner).

Writing lyrics in these themes obviously result in lower word frequency levels since they are not about everyday topics which could be expressed in a moderate language. As a result, some death metal bands take this to extreme ends and create extremely complex lyrics. Doug Moore addressed the issue in a post on *Invisible Oranges*, a metal blog residing at www.invisibleoranges.com, through an example by *Vastum*'s *Incel*:

> *Deprivations of the flesh / A quarantine in abject absence / Euphoria denied: dripping, unsated phantom limb / Suffocated, stifled mass can feel the temperature rise / By knowledge unrealized: a tumorous, parasitic grief* (Moore 2013).

He calls this kind of language "Death Metal English" and following the quoted lyrics, lists the common treats of Death Metal English as: polysyllabic words, a good deal of adjectives, prepositional phrases, progressive tense, passive voice, archaic or pseudo-Biblical verbiage, grandiloquent metaphor and illogical or meaningless sentences (Moore 2013). These features of death metal lyrics have been hugely influential in forming the basis of the present study.

Three major metal subgenres have been introduced so far providing information on their brief historical backgrounds, musical and lyrical traits and naming some exemplary bands for each one of them. In line with the focus of the present study, these subgenres will be investigated by means of a series of corpus analyses. Thus, the next section will provide details on corpus linguistics.

2.2. Corpus Linguistics

A corpus is basically a collection of texts. Even though, "the very notion of what constitutes a valid corpus can still be controversial" (Kennedy 1998: 2), it is not a mere random collection, but a carefully built one taking various aspects into consideration. Kennedy (ibid.) differentiates between *archive* and *corpus*, emphasizing that the former is a non-structured, randomly or opportunistically collected bodies of texts, whereas the latter is built according to a system and plan (p. 4). He quotes Leech (1991: 11) and states that "the difference between an archive and a corpus must be that the latter is designed or required for a particular 'representative' function" (ibid.). It is not the sheer size of the texts that constitute a corpus but a body of texts that can be subject to particular investigations. "A corpus is defined not by what it contains but by how it is used" (Hunston 2012: 243). However, there is another view that any collection of texts can be considered corpus regardless of their sampling method (Kübler and Zinsmeister 2015: 9) Kübler and Zinsmeister (2015: 4-8) highlight the necessity of corpus to be available in electronic medium so that it can be searched and analyzed.

Corpus studies expand in two dimensions as ***mega corpora***, which consist of multi-million words and ***specialized corpora*** which are used mostly for genre studies (Flowerdew 2005 qtd. in Handford 2012: 14-5, Handford 2010: 256). ***General corpora*** (or ***reference corpora***) are carefully sampled to achieve utmost representativeness (Kübler and Zinsmeister 2015: 10). One of the largest general corpora, The British National Corpus (BNC) is a general, multilingual, synchronic corpus, built to represent the English language in general (http://www.natcorp.ox.ac.uk/corpus/index.xml). It was built in 1994 and the second and third (the latest) editions came out in 2001 and 2007 respectively. It consists of 100 million words from various genres in written and spoken language. The BNC contains 4,049 texts, 90% of which are written and 10% are spoken (http://www.natcorp.ox.ac.uk/corpus/index.xml?ID=numbers). 3,887 of the texts (98.18%) were produced between the years 1960 and 1993). The production date of the remaining 162 texts, amounting to 2.09% of the corpus, is unknown (http://www.natcorp.ox.ac.uk/docs/URG/BNCdes.html#BNCcompo). The BNC has been used in various studies over the years. As the time passes, though, it ages as it is a static corpus.

The Contemporary Corpus of American English (COCA) was released in late 2008 by Davies (2009) in an attempt to substitute the BNC and the American National Corpus, which was aimed to contain 100 million words in size but only achieved 22 million (Davies 2009: 160). The COCA contains 560 million words in 220,225 texts as of December 2017 (https://www.english-corpora.org/coca/). Unlike the BNC, the COCA is a monitor corpus and is constantly updated. It started with 385 million words and has increased by half over the course of approximately ten years. Davies (2009: 161) gives an account of the COCA's comparability to the BNC as follows,

> [COCA] was designed to be roughly comparable to the BNC in terms of text types. In the BNC, approximately 10% of the texts come from spoken, 16% from fiction, 15% from (popular) magazines, 10% from newspapers, and 15% from academic, with the balance coming from other genres. In the COCA, texts are evenly divided between spoken (20%), fiction (20%), popular magazines (20%), newspapers (20%) and academic journals (20%).

Corpus compilation has been made easier thorough the advance of technology and the Internet in particular. Modern mega corpora are based on online resources, some of which are immense in size, such as iWeb: The Intelligent Web-based Corpus, News on the Web (NOW), Global Web-Based English (GloWbE), Wikipedia Corpus, etc. By way of comparison, iWeb: The Intelligent Web-based Corpus, the largest corpus of English, has 14 billion words which makes is 25 times larger than the COCA (https://www.english-corpora.org/). Table 4 shows the details of the largest corpora of English.

Table 4: Mega corpora adapted from https://www.english-corpora.org/ as of 13.04.2019.

Corpora	# words	Language/Dialect	Time Period
iWeb: The Intelligent Web-based Corpus	14 billion	US/CA/UK/IE/AU/NZ	2017
News on the Web (NOW)	7.64 billion+	20 countries / Web	2010-last month
Global Web-Based English (GloWbE)	1.9 billion	20 countries / Web	2012-13
Wikipedia Corpus	1.9 billion	English	2014
Hansard Corpus	1.6 billion	British (parliament)	1803-2005
Early English Books Online	755 million	British	1470s-1690s
Corpus of Contemporary American English (COCA)	560 million	American	1990-2017
Corpus of Historical American English (COHA)	400 million	American	1810-2009
The TV Corpus	325 million	US/CA/UK/IE/AU/NZ	1950-2018
The Movie Corpus	200 million	US/CA/UK/IE/AU/NZ	1930-2018
Corpus of US Supreme Court Opinions	130 million	American (law)	1790s-present
TIME Magazine Corpus	100 million	American	1923-2006
Corpus of American Soap Operas	100 million	American	2001-2012
British National Corpus	100 million	British	1980s-1993
Strathy Corpus (Canada)	50 million	Canadian	1970s-2000s

Corpora	# words	Language/Dialect	Time Period
CORE Corpus	50 million	Web registers	2014

The sizes of the largest corpora range between 14 billion and 50 million words. They also differ in their coverage of periods. Monitor corpora such as News on the Web (NOW) are constantly updated to include the most recent texts. Some of them include one dialect or genre (Corpus of American Soap Operas) whereas others are more comprehensive (Wikipedia Corpus).

Specialized corpora are those which are built for specific research projects (Kennedy 1998: 20). Unlike general corpora, which feature as many genres as possible, they usually contain texts from specific genres, fields, etc. They are better structured for genre studies as they are built according to the set of rules reflecting the contextual features of genres (Koester 2012: 48). It has become common practice to design a specialized corpus for genre studies and their sizes vary (Warren 2013: 4). There is not a standard recipe for a specialized corpus as it depends on the research design and the hypotheses (Teubert and Čermáková 2004: 120). A study might target an analysis on texts which are not part of any corpora – or at least not in a sufficient manner – which requires the researcher to build their own specialized corpus Kennedy (1998: 70).

One of the most important aspects of corpora is the ***size***. Corpus size, in fact, is a controversial issue, which mostly depends on the purpose of the corpora. First, it must be noted that no matter how big any corpus is, it cannot be more than a very small sample of the language used on a single day (Kennedy 1998: 66). There is not a universal method of calculating the ideal corpus size (Eiter 2017: 7). In general, the larger corpus, the better; nevertheless, the composition of the corpus, i.e., the diversity of genres it offers is a more preferable trait (Meyer 2002: 44).

Each study might look into different genres and the availability of the texts for a study is a key determinant in the size of the corpus (Hunston 2012: 243). For example, while a study on the newspaper articles on economy can be realized with a million-word corpus, another one on love poems written in a specific period of time might be carried out with a 100,000-word one. “A corpus is defined not by what it contains but by how it is used" (Hunston 2012: 243). Tagg (2012: 161) advocates the plausibility of small-sized corpora if they are suitable for the desired analyses. Tagg (2012: 161) claims that "depending on the purpose and nature of the data, size is not necessarily a problem".

Similarly, Baker (2006: 24) mentions Stubbs (1996: 81-100) and Shalom (1997) while making a case on the effective use of small corpora. The former author compared two short letters by the same author (330 and 550 words each) and found differences in repetitions while the latter studied 776 newspaper ads to achieve noteworthy results on word choices of heterosexual and homosexual people looking for partners. Biber (1990: 261) mentions that 1,000 word samples are mostly enough for investigating many linguistic features. However, for lexicographical purposes, larger corpora are necessary as collocations could appear in low frequencies. Although size is a crucial issue, it cannot compensate a deficiency in diversity. There is a limit for studies and it is important to be realistic in corpus design. However, if well-designed, small corpora are also useful in many linguistic analyses (Biber et al.1998: 250).

Representativeness is a feature of corpus to represent the target language (or desired genre[s]) as best as it can. An ideal corpus is supposed to reflect the language on a broad scale, from general to specific covering as many genres, periods, etc. (Gellerstam 1992: 154). Kennedy (1992) advocates a careful selection of texts stating that "[t]exts selected without awareness of how typically they represent salient features of the language can present a chaotic picture of the language" (p. 366). Although it is not completely clear how to measure representativeness (Kennedy 1998: 62), having a wide range of genres in a corpus could be regarded as a solid attempt at maintaining representativeness; however, the question of "which genres to include" is the flip side of the medal, which needs further consideration (Kennedy 1998: 62). Biber et al. (1998: 246) remarks on representativeness as follows:

> A corpus is not simply a collection of texts. Rather, a corpus seeks to represent a language or some part of a language. The appropriate design for a corpus therefore depends upon what it is meant to represent. The representativeness of the corpus, in turn determines the kinds of research questions that can be addressed and the generalizability of the results of the research.

Biber (2006) puts forward two important features, which affect the representativeness of a corpus: *size* and *composition*. Size is important especially to study on occurrences of specific items. Composition refers to the inclusion of various registers so that a balanced and diverse corpus is attained (pp. 251-2). If the composition of the corpus is not carefully planned, the analyses may be severely compromised (Meyer 2002: XIV). Miller and Biber (2015) emphasize two dimensions of corpus representativeness as *internal* and *external* as proposed by McEnery et al. (2006: 14). They relate these two dimensions to Biber's (1993: 243) interpretation of those corpus features: linguistic representativeness and situational representativeness. Internal (linguistic) representativeness is to which extent the individual pieces of the corpus yields the same (or similar) results upon replication, while external (situational) representativeness is about correct sampling of text so that they are capable of reflecting the characteristics of that particular field/genre.

These notions are mostly discussed in terms of general corpora which include many texts from various genres (see: Kennedy 1992, 1998, Kübler and Zinsmeister 2015); however, they also apply to specialized corpora. An important point noted by Kübler and Zinsmeister (2015: 10) is to pay attention not to have "rogue texts" which do not belong to the respective genres. This could be attained by clear-cut genre distinction. Biber (1993: 243) explains it by emphasizing that it is important to draw the lines of the target population from which the texts will be selected so that the texts to be included or excluded could be made clear. Furthermore, the corpus builders must refrain from opting for texts which they think will better fit the hypothesis or eliminate others which may run counter to it Despite the fact that these considerations have to be observed, it needs to be pointed out that no corpus can be ideally representative and it can only be a matter of degree (McEnery and Hardie 2012: 10-5).

Corpora also differ from each other depending on the date of the texts they contain. There are two main types of corpora with relation to time as ***synchronic*** and ***diachronic*** corpora. Synchronic corpora cover language samples from a specific time while diachronic ones cover a period of time. Synchronic corpora covers a relatively limited period, i.e., close to the data collection time, yet diachronic ones cover longer periods. (Kübler and Zinsmeister 2015: 13).

Rissanen (1992) defines four key features of a diachronic corpus. A diachronic corpus needs to have a) chronological coverage of the intended period(s), b) regional coverage to feature as many varieties of the language, c) sociolinguistic coverage inclusive of members of the society from diverse backgrounds, age groups, sexes, etc. and 4) generic coverage to cover a wide array of different genres (p. 189).

Another time-aspect of corpora is determined based on the size and growth, which manifests itself as ***static*** or ***monitor*** corpora. Static corpora, as with most linguistic ones, are built once and not changed afterwards. Of course, they may be updated at times (see the BNC), but they are still considered static. Monitor corpora (term coined by Sinclair 1991: 26), on the other hand, is open-ended, regularly updated and can constantly grow on a certain basis of time, such as newspaper corpora growing every day (Teubert and Čermáková 2004: 121, Kübler and Zinsmeister 2015: 13-4). This kind of a corpus has both a historical dimension and the most up-to-date language samples as a result of extensive record keeping (Sinclair 1991: 25).

Corpus studies sometimes involve the tagging of particular items of language. Untagged corpora are usually used for searching words or word sequences, which is frequently undertaken by means of concordances (Biber et al. 1998: 257). As corpus tagging is a time-consuming work, corpus studies usually focus on untagged corpora (Kennedy 1998: 90). Whether or not a corpus has to be tagged depends on the research setting. If the aim is to look for frequencies an untagged corpus can be used – although Biber et al. (1998: 31) thinks otherwise. If, however, the aim is to find out the lexico-grammatical features, a tagged corpus will be necessary. In some cases, both tagged and untagged versions of the same corpora could be used as in the case of Taina (2014: 37), who studied frequency levels of and some grammatical patterns in metal song lyrics.

The most common annotation methods are word-based and syntactic tagging. As per the focus of the present study, only word-based annotation will be introduced. In word-based annotation, or tagging, the words are tagged based on their classes – parts of speech, and this process is called Parts Of Speech (POS) tagging. It allows the researcher to count specific parts of speech or distinguish *chair (n)* from *chair (v).*

There are numerous tools for POS tagging – usually called taggers or POS taggers - and it seems like an easy task to tag all the words in a text. However, there is naturally a limit to how accurately the computers can tell apart words. Is *like* a noun, an adjective or a verb? In such cases, human intervention in necessary; nonetheless, it is not the best alternative. As Kennedy (1998: 5), notes, “the analysis of huge bodies of text ‘by hand’ can be prone to error and is not always exhaustive or easily replicable.”

Taggers have different aims depending on what they have been built for (Sinclair 1992: 385) and it is not easy to say which tagger works best for a particular task. In order to be able come up with such a claim, the same texts should be analyzed by all the taggers and the results must be compared to each other, which is hardly plausible given the fact that the some studies are carried out by self-developed software (Kennedy 1998: 226).

Based on these views, Kübler and Zinsmeister (2015: 24) assert that annotations could be made in three ways as, *manual*, *automatic* and *semi-automatic*. Manual annotation is likely to yield fewer errors than automatic annotation. However, the errors in automatic processing are usually more consistent (ibid.). Manual annotation is bound to be inconsistent and contain errors as a result of the human factor (McEnery and Hardie 2012: 32). It seems here that semi-automatic annotation is the best choice to achieve the least error rate. If an automatic annotation is not followed by a manual correction, the rate of error can be quite high (Meyer 2002: 89). However, manual annotation – even when it is performed semi-automatically – is highly time consuming and difficult (Kübler and Zinsmeister 2015: 24-33) and it will continue to be so in the foreseeable future (Meyer 2002: 99).

Another way to process corpora is ***lemmatization***. “A lemma is the basic form of a word, as it is represented in a lexicon” (Kübler and Zinsmeister 2015: 45). Kennedy (1998: 97-9) quotes Francis and Kucera’s (1982: 1) definition of lemma as “a set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling.” Lemmatizing is grouping of the inflected versions of the same head word. For example, *listens*, *listened*, *listening* can be lemmatized under the head word *LISTEN*. Biber (2006), for instance, carried out his research on the frequency levels of different registers at lemma level (p. 35). This process allows researchers to search for headwords and find all versions. For example, in order to find the occurrences of the verb *be*, it would be possible to find all forms (*am*, *is are*, etc.) by searching the verb *be* in a lemmatized corpus. In an unlemmatized one, each form has to be looked for separately (Meyer 2002: 116).

There are tools for this purpose, which are called lemmatizers, and they are quite reliable as they are rule-based (Kübler and Zinsmeister 2015: 47). However, lemmatizing is not as easy as it may seem (Sinclair 1992: 391). There are some intermediate cases/words, such as considering contractions one or two words, treating the colloquialisms, etc. (Kübler and Zinsmeister 2015: 46), which require a certain path to be chosen.

An alternative approach would be to use word families, as proposed by Nation (2001). The word family approach groups the closely derived words together, which are transparently related to the core sense of the headword (Nation 2001 qtd. in Biber 2006: 242). In this approach, for example, the words *care* and *careful* could be grouped together even though *careful* belongs to a different part of speech (*adj*.) since the derivation allows a connection between the headword and the derivative – hence transparency is maintained. Biber (2006) prefers to follow the lemma approach instead of word families as it may not be clear which items are transparent and which are not (pp. 242-3).

2.2.1. Corpus Analysis

Corpus studies and analyses differ depending on the research setting. Tognini-Bonelli (2001) proposes two approaches in corpus studies as *corpus-based* and *corpus-driven.* Corpus-based approach is a top-down process which justifies or refutes an existing theory through the use of a corpus. Corpus-driven approach, on the other hand, is a bottom-up one which seeks to devise theories based on the findings on corpus studies (McEnery and Hardie 2012: 6, 150). McEnery et al. (2006) discuss these two approaches and argue that corpus-driven approach is not easily attainable as it requires zero previous knowledge of linguistics. They state that the distinction between the two approaches are quite fuzzy and prefer to use "corpus-based" as an umbrella term to refer to corpus studies in their book regardless of which approach they follow (pp. 10-1).

There is a third method, a combination of the two methods, which involves both methods. Rayson (2008: 4) calls this *data-driven* approach or *Type III*. In this approach, the focus is first on the whole texts, and then detailed studies are carried out. He outlines the steps as follows:

1. Build: Corpus design and compilation
2. Annotate: Manual or automatic analysis of the corpus
3. Retrieve: Quantitative and qualitative analyses of the corpus
4. Question: Devise a research question or model (iteration back to Step 3)
5. Interpret: Manual interpretation of the results or confirmation of the accuracy of the model (Rayson 2008: 4).

Rayson (2008) mentions Ringbom (1998), Hoffmann and Lehmann (2000), and Leech and Fallon (1992) as similar studies involving Type III approach. Taina (2014) also follows the same approach in his analysis of metal lyrics. Although the approaches have different names, they hardly differ in methodology and as McEnery et al. (2006) puts it they are quite fuzzy.

Corpus analyses can discover if a particular item or pattern exists in a particular corpus or it can tell us how frequently those items or patterns occur. However, corpus studies do not provide negative evidence, i.e., it cannot be inferred from any corpus analysis that a particular pattern does not exist since a corpus can only cover a limited portion of a language and if that corpus is, perhaps, enlarged by a small margin, there might be a chance that that particular pattern shows up (Kübler and Zinsmeister 2015: 164-6).

Biber et al. (1998) argue that, in corpus analyses, it is of utmost importance to specify the unit of analyses. The unit of analyses in corpus studies could either be the occurrences of a specific linguistic feature or the text. If the unit is the linguistic features, then the focus will solely be on that particular feature and each occurrence of the item must be observed. On the other hand, if the unit of study is the text, counts or rates of linguistic items are taken into consideration and each text (or register/genre) is assigned a score based on the counts and rates of chosen linguistic items, which, in turn, allows a comparison or different registers (Biber et al. 1998: 269-73). Corpus analyses are used to study variation between different corpora as well as investigating the distribution and frequency of specific linguistic elements (Kennedy 1998: 4).

Handford (2012: 17) highlights four potential aspects in corpus linguistics which could be identified as important: a) frequent items, b) statistically significant items (compared to some norm), c) stylistically salient/culturally key items and d) items found important in other related studies. Frequency of particular items is a distinctive feature of a specialized corpus that can be analyzed through software (Koester 2012: 49). Some of this kind of software has built-in statistical algorithms, which provide statistical significance reports of the results (Meyer 2002: 120).

Corpus analysis methods are frequently used in genre/register studies, mostly through the use of specialized corpora. This way, distinctive properties of genres and the reasons for their use could be identified in addition to establishing the structural identity of genres through their lexico-grammatical patterns (Hyland 2012: 31). According to Biber et al. (1998: 136-7) genre studies through corpus have three indispensable requirements: 1) "inclusion of a large number of texts", so that the studies are carried out accurately, 2) "consideration of a wide range of linguistic characteristics" to achieve accurate generalizations, 3) "comparison across registers" to set a reference point for the frequencies of particular items.

Corpus size, inevitably, has an effect on some analyses. Biber (2006) points out that the distribution of word types (occurrences of different words) are not linear. Thus, the fact that there are 500 word types in a 1,000-word text does not mean that 5,000 word types can be identified in a 10,000-word corpus. He illustrates the issue with the coverage of T2K-SWAL Corpus in full and in half as shown in Table 5.

Table 5: Comparison of full and half corpus figures from the written texts from T2K-SWAL Corpus – adapted from Biber's (2006: 253-4).

		Full Corpus	**Half Corpus**	**Percentage Representation**
Spoken Texts	# of texts:	291	146	
	# of words	1,665,624	806,023	
	# of word types	27,312	19,342	70.8%
Written texts	# of texts:	172	86	
	# of words	1,073,508	512,865	
	# of word types	39,053	27,409	70.2%
Written Social Science Texts	# of texts:	35	18	
	# of words	262,707	139,23	
	# of word types	17,935	12,641	70.5%

The ratio of word types range within a very limited scale – less than 0.3% – and accumulate around 70%, regardless of the corpus medium and content. Due to this non-linear distribution, it is advised to work with normalized ratios rather than raw counts of occurrences.

Corpus analysis methods are usually quantitative ones (e.g., frequency, ratios, etc.); yet, they can be of qualitative nature (e.g., concordances, lexico-grammatical analyses, etc.). However, they may not be enough to put forward the differences across registers without qualitative interpretation (Biber et al. 1998: 139). Seidlhofer (2012: 142) cautions that quantitative methods might offer too little where qualitative ones may be misleading if not treated with caution. A good corpus offers a balanced distribution of quantitative and qualitative methods (Meyer 2002: 123).

Corpus analysis methods, more often than not, feature a comparison of a specialized corpus to general ones where general (or reference) corpora are used as benchmarks (Hunston 2012: 244, Teubert and Čermáková 2004: 119). O'Keeffe (2012: 119) reminds that the use of different reference corpora yields different results; hence, the selection of the reference corpus must be carefully thought out. She compared a TV interview with a corpus of media interviews and an academic corpus. As expected, the results were different from each other as the academic corpus was not relevant to the purpose of the study (pp. 119-120). Some studies may employ more than one reference corpora, as in the case of Werner

(2012) where he compared 1128 song lyrics to the BNC, American National Corpus (ANC), CIC and some other minor ones (p. 23).

The use of corpora is quite common for purposes other than linguistics as well (Kennedy 1998: 3-4). O'Keeffe (2012: 119) suggests that a corpus analysis of frequency, keywords and concordancing can offer much to media studies. The interdisciplinary nature of corpus studies is also mentioned by McEnery and Hardie (2012). They suggest that corpus studies may (and should) be applied in disciplines other than linguistics (p. 227). Such studies shed light on specific features of genres (Hyland 2012: 31) and researchers working on particular genres may compile their own corpora (Kennedy 1998: 70). Working on genres, however, requires a careful selection of corpora as the results will be heavily dependent on the type of corpora to be used and genres differ not only in terms of vocabulary but also grammatical structures and other linguistic properties (Kübler and Zinsmeister 2015: 11).

2.2.2. Lyrics Corpora and Studies

Song lyrics have been within the focus of academic studies for quite a long time. They have been studied in the fields of pedagogy, psychology, sociology, etc. Studies not only vary depending on their academic fields but also on their choice of music genres. While some studies focus on specific genres, others follow a comparative approach. Kreyer and Mukherjee (2007: 32) argue that "[t]he discourse and language of pop song lyrics has attracted much interest in the field of cultural studies and in the English language teaching community".

Werner (2012) observes that despite the fact that song lyrics are a significant part of daily life they are not included in many of the general corpora, such as the BNC, ICE, CIC, Brown family, etc. Due to the fact that general corpora of the English language hardly contain song lyrics, studies on genre comparisons and lyrics analysis have developed their own specialized corpora in varying numbers of songs per genre.

Miethaner (2005 qtd. in Kreyer and Mukherjee 2007) built one of the largest lyrics corpora, which is limited to one music genre: blues. He studied the use of African American Vernacular English (AAVE) in blues lyrics. The

Giessen-Bonn Corpus of Popular Music (GBoP) was built by Kreyer and Mukherjee (2007) to study the linguistic variations of pop songs. It comprises 27 albums, 442 songs and around 176,000 words. They state that pop song lyrics could be considered "written-to-be-spoken" and they make use of "deviant spellings" (Kreyer and Mukherjee 2007: 37-8). What they term as deviant spelling are the words such as, "*Synkronized*, *Afrika Shox*, *C U When U Get There*, *Money Don't Matter 2 Night*", etc. (ibid. p. 34).

Murphey's (1992) study features a corpus of 50 songs totaling 13,161 words. He analyzed 50 pop songs in terms of five different aspects, namely word counts, *you* and *me* references; time, place and gender references; words per minute; and readability and human interest. He summarizes his findings as follows:

> [Pop songs] offer short, affective, simple, native texts with a lot of familiar vocabulary recycled, yet vague. They are dialogic and engaging auditorily but, because of our narrative expectations, they are probably not very interesting as reading material (Murphey 1992: 773-4)

Eiter (2017) compiled the Innsbruck Corpus of English Pop Songs (ICEPS) containing 303 songs released between the years 2012 and 2016, totaling 119,982 tokens. He investigated the non-standard English use, such as *ain't*, third person singular *don't*, negative concordance, etc., in order to decide whether lyrics are of spoken or written nature. His findings suggest that lyrics stand somewhere in between both registers as written-to-be-sung texts (Eiter 2017: 46).

Sophiadi (2014) compared the lyrics of rock and pop songs across five decades (1960s-2000s) by choosing 25 rock and 20 pop songs for each decade. Her aim was to find out how the rock and pop lyrics changed over time depending on the historical changes. She found that rock lyrics responded more to the changes than pop lyrics and commented that "rock is more pop than pop is" (Sophiadi 2014: 138).

In her diachronic investigation of rock song lyrics, Falk (2012) compiled a corpus of 185 songs released between the 1950s and 1990s. She analyzed the use of particular words across decades. For example, the use of *baby* was more frequent between 1950s and 1960s while it decreased over the years (Falk 2012: 22). Another important finding in her study shows that the song lyrics are neither spoken nor written, but they constitute a special category (p. 21), as also indicated by Kreyer and Mukherjee (2007) and Eiter (2017).

Dukes et al. (2003) also carried out a diachronic semantic research on song lyrics. They investigated the use and change of love and hurt words in popular music over time. Their corpus consisted of 100 songs taken from the charts between the years 1958 and 1998. They found that female singers used more sexual references in song lyrics in early seventies and late eighties whereas male singers used more in the nineties. Pettijohn and Sacco (2009) analyzed the No. 1 song of each year between 1955 and 2003 to shed light on the changes of lyrical content of the songs. Their findings indicate that the song lyrics are more meaningful during times of social and economic threat (p. 297).

Petrie et al. (2008) studied *The Beatles* song lyrics to find out the stylistic differences between the songs written by John Lennon, Paul McCartney and George Harrison. They also observed the differences in lyrics over the course of the band's career from a diachronic viewpoint. They compiled a corpus of 185 songs, which are written by these three songwriters between 1960 and 1970 and analyzed them using the software Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2015). They built another corpus consisting of approximately 100 songs form US Billboard charts to set up a comparison group for *The Beatles'* songs. The results showed that Lennon's songs display more negative emotion than McCartney's and Harrisson's lyrics are associated with a more intellectual approach (p. 200).

Olivo (2001) investigated rap songs lyrics for the use of AAVE through a corpus of 18 albums by 13 different artists. He notes that the use of non-standard spelling is used deliberately as well as due to AAVE. Such spellings are a distinctive feature of the rap culture and their use helps sustain it (Olivo 2001: 67). These studies, however different in their size, content and purposes, all contain one specialized corpus of song lyrics. There are other studies, which contain two or more sets of corpora, for contrastive analyses.

Werner (2012), for instance, compiled two corpora for his study, one for American English and another for British, consisting of a total of 1,128 songs. The study focused on historical change in the lyrics and the style; hence, the

British corpus contained songs released between the years 1952 and 2008 and the American between 1946 and 2005. He posits that, unlike the classical division of registers as spoken and written, song lyrics do not fit into either category but stand as a distinctive genre of its own (pp. 19-21). His findings suggest that the language of lyrics resembles spoken language in that they cover a limited range of topics and high level of self-referencing. However, a difference is observed between the lyrics and spoken language when it comes to the usage of certain elements, such as interjections, which are quite frequent in spoken language whereas scarce in lyrics (pp. 33-4).

Tegge (2017), too, compiled two corpora for comparison. The first corpus, the Wellington Corpus of Popular Songs (WOP), has 408 pop songs and the second one, the Wellington Corpus of Popular Songs in English Teaching (WOPET), 635. The WOP corpus contains songs from the years 2008-2014 while WOPET has songs spanning from 1900s to 2011. Tegge's aim was to find out the lexical demand of song lyrics for classroom applications. Her findings indicate that the songs from the charts share the same characteristics as scripted and non-scripted spoken language and that the teacher-selected songs require the knowledge of fewer words (p. 95).

Öztürk (2017) tested the vocabulary load of popular songs using a corpus of 177 songs by four different artists released between 2005 and 2014. Her analysis was on the frequency levels of the lyrics and the details on the study will be provided in Studies on Lexical Richness. Finally, the study which is so far the most similar one to the present one is Taina's (2014) analysis of five metal subgenres in terms of keyness and co-keyness. His corpus consisted of 200 songs in total (40 from each subgenre) spanning the years between 1970 and 2010. He notes that black metal and death metal have more lexical variance compared to traditional heavy metal. He also remarks that keywords reveal stylistic features for each subgenre (pp. 85-6).

As can be seen in the brief summaries of the studies, they are quite different from each other, mostly due to their research questions. The corpus details of all the above-mentioned studies are summarized in Table 6 below:

Table 6: Details of selected lyrics corpora

Corpus Builder	Corpus	Song selection method	Songs	Words	Av. Words per songs	Year span	Coverage of Years
Eiter (2017)	Innsbruck Corpus of English Pop Songs (ICEPS)	Charts	303	119,982	396	2012-2016	4
Öztürk (2017)	Rock and pop songs	Personal taste - Most recent 4 albums by each artist	177	54,661	309	2005-2014	9
Tegge (2017)	The Wellington Corpus of Popular Songs (WOP)	Charts	408	180,892	443	2008-2014	6
	The Wellington Corpus of Popular Songs in English Teaching (WOPET)	Criteria-based	635	177,384	279	before 1900s to 2011	111
Sophiadi (2014)	Rock	Charts	125	Not specified	N/A	1960s-2000s	50
	Pop	Charts	100	Not specified	N/A	1960s-2000s	50
Taina (2014)	Metal Lyrics Corpus (METAL)	Criteria-based	200	40,915	205	1970-2011	41
Falk (2012)	Rock Lyrics Corpus (ROLC)	Charts	300	52,907	176	1950s-1999	49
Werner (2012)	British Chart Corpus (BCC)	Charts	1,128	170,000	301	1952-2008	56
	American Chart Corpus (ACC)	Charts		170,000		1946-2005	59
Petrie et al. (2008)	The Beatles Songs	Complete discography	185	N/A	161	1960-1970	10
Pettijohn and Sacco (2009)	Billboard No. 1 songs	Charts	49	N/A	N/A	1955-2003	48
Kreyer and Mukherjee (2007)	Giessen-Bonn Corpus of Popular Music (GBoP)	Chart - Albums (27)	442	176,000	398	2003	1

Corpus Builder	Corpus	Song selection method	Songs	Words	Av. Words per songs	Year span	Coverage of Years
Miethaner (2005)	Blues Lyrics Collected at the University of Regensburg (BLUR)	N/A	7,341	1,490,000	203	1920s-around 1939	19
Dukes et al. (2003)	Popular Songs	Charts	100	N/A	N/A	1958-1998	30
Olivo (2001)	Rap Songs	Albums (18)	Not specified	Not specified	N/A	1991-1997	6
Murphey (1992)	Pop Songs	Charts	50	13,161	263	1987	1

The selection of songs for most of the corpora is done through the media charts, listing the most popular songs/albums for a particular year. Out of the 17 corpora in Table 6, 11 are created with the songs and albums from charts. GBoP, Rap Songs and *The Beatles* Songs feature an album-based selection, contrary to the other 13 chart-based corpora, which include songs only. Only two corpora are based on informed selection of songs. WOPET features songs, which are mentioned in an online survey carried out by Tegge and taken by ESL/EFL teachers, and other songs are taken from ESL textbooks and websites (Tegge 2017: 90). Taina (2014) compiled his corpus based on websites on metal. In an attempt to have an objective methodology, he used the search option of *Metal Storm* (a metal website) to choose the songs to be included (pp. 33-4). However, his attribution of specific bands to certain subgenres is questionable (e.g., *Gojira* as death metal, *Death* as thrash metal, *Ghost*, *Kiss*, *Led Zeppelin* and *Mercenary* as heavy metal, etc.). Petrie et al.'s (2008) corpus is unique in that it features the complete discography of *The Beatles* except for the songs which contained fewer than fifty words and those written by Richard Starkey (Ringo Starr) (p. 198). Öztürk (2017) compiled her corpus based on her personal music taste. The methodology of BLUR is not known as the source material is unavailable to the author of the present study.

These corpora show differences in their sizes. Their size could be compared based on the number of songs and words. The largest lyrics corpus, by all standards, is BLUR prepared by Miethaner (2005) which contains 7,341 blues song lyrics and a total of 1.49 million words. This size is unmatched as the closest ones contain around 170,000 words (GBoP, WOP, WOPET, BCC and ACC). These corpora contain 442, 408 and 635 songs respectively. BCC and ACC contain 1128 albums in total, and if equally distributed, correspond to 564 albums each. ICEP contains 303 songs and 119,982 words. The smallest corpora are Murphey's (1992) and Pettijohn and Sacco's (2009), with the former featuring 50 pop songs which contain 13,161 words and the latter 49 songs. Sophiadi's (2014) corpora of rock and pop lyrics contain 125 and 100 songs lyrics; however, the numbers of words are not provided in the study. Similarly, Olivo's (2001) study does not mention the number of songs or words. The average number of songs per corpus – except for the outliers (BLUR and Murphey [1992]) – is approximately 280.

The number of songs and words were rationalized to have a better view of the average number of words per song. The numbers range between 161 and 443 words/songs. A closer look at the number of songs and words reveal that they are not quite consistent with each other. For example, although ICEPS and ROLC have almost the same amount of songs (300 and 303 respectively) the former contains 119,982 words whereas the latter 52,907. These numbers are reflected in the average number of words per song as 396 and 176, with a difference of more than two times. This difference obviously stems from the way the lyrics are processed. While Falk (2012: 10) removed the repeats for ROLC, Eiter (2017: 14) included them all in ICEPS. Petrie et al. (2008) removed the repeats at the third occurrence. Tegge (2017) also removed the repeats at the fade-out parts in WOP; nevertheless, this does not seem to affect the average as much as it did on Falk (2012) and revealed a value of 443. This indicates that there may be other factors playing a role in the average number of words per song (perhaps, genres, artists, etc.). To sum up, excluding the outliers (ROLC and WOP), the average number of words per song is around 250-300.

The last aspect of the selected lyrics corpora to be investigated is their coverage of periods. As can be seen in Table 6, the coverage of corpora ranges between one and approximately 111 years. The corpora of the five diachronic studies, Dukes et al. (2003), Petrie et al. (2008), Pettijohn and Sacco (2009), Falk (2012) and Sophiadi (2014), cover an average year span of around 40 years. The coverage of other corpora is more or less the same excluding the outliers.

Collecting the lyrics for a corpus is more often than not a problematic issue. Most albums have their lyrics printed on the album sleeves but there are cases where this does not apply. The most commonly used method is to access them on websites which provide song lyrics, but not necessarily the artists' official pages. Many studies make use of lyrics downloaded from such third-party websites such as www.azlyrics.com, www.lyricsdepot.com, etc. (see Dukes et al. 2003, Kreyer and Mukherjee 2007, Petrie et al. 2008, Pettijohn and Sacco 2009, Falk 2012, Werner 2012, Taina 2014, Eiter 2017, Öztürk 2017 and Tegge 2017).

Olivo (2001: 72-3), on the other hand, acquired the song lyrics from album sleeves. Knees et al. (2005) warn that even though such websites have a broad range of lyrics, none are complete and accurate.

It is not easy or even appropriate to use the lyrics exactly the way they are published. Even on official resources (album sleeves and official web sites) there could be typos or spelling errors (mostly with, but not limited to, NNS artists). Lyrics used in academic studies are usually collected from various sources and this may cause some challenges. Knees et al. (2005: 565) point out to six main problematic areas in collecting lyrics from various sources as follows:

- Different spellings of words
- Differences in the semantic content
- Different versions of songs
- Annotation of background voices, spoken text, and sounds
- Annotation of chorus, verses, and performing artist
- References and abbreviations of repetitions

In order to avoid these shortcomings, Knees et al. (2005) chose to double check the lyrics from online sources with the original ones in the CD sleeves. Werner (2012) also used the same method and checked the lyrics from the artists' homepages and other online databases as well. Nevertheless, this is not always the ultimate solution because there may be typos also in the booklets (Knees et al. 2005: 568). The next step after acquiring the lyrics is to process them. In many papers, which are based on text analysis, alterations or removals are made. Below is a list of some lyrics processing methods:

1. ***Removing markups:*** In this method, the markup lines such as 'repeat', 'chorus', etc., which are not primary parts of the text, are removed. Meara (1993), for example, analyzed BBC English series aiming at a lexical profile and he had to manually process the text to omit parts such as 'stage directions' and similar other details to be able to obtain an automatically analyzable text. Referring to the studies listed in Table 6, Werner (2012), Taina (2014) Eiter (2017) and Öztürk (2017) and removed such metadata. Other studies do not mention a specific procedure regarding this issue. In addition to the above-mentioned markups, Werner (2012: 23) and Taina (2014: 36) removed character names or narrated/unsung parts from the corpus. Werner (2012: 23)

and Eiter (2017: 14) mention the omission of artist names and album/song titles from the texts.

2. ***Removing markups and inserting corresponding text:*** This method could be seen as a complimentary to the first method. The principle is to remove the markup lines and to add the respective text in their stead. In their study on textual analysis of song lyrics, Mayer et al. (2008) retrieved the lyrics for 397 songs from the Internet. Then, they manually processed all the songs and removed markup lines such as '2x' or 'chorus'. Those parts were replaced by the respective corresponding text. Taina (2014), Eiter (2017) and Öztürk 2017 followed a similar path and spelled out all repeats. Knees et al. (2005: 566) looked for markups such as *chorus*, *refrain* etc. written before the paragraphs and inserted them at the next occurrence of the same markups.
3. ***Modifying/correcting the text:*** In some cases, lyrics are modified for correctional purposes or to fit in a specific study setting. This method is usually carried out by manually checking the lyrics from album sleeves or checking for inconsistencies by listening to each song. Petrie et al. (2008) converted the lyrics they used in their study into American English spelling. Lightman et al. (2007) compared the lyrics by suicidal and non-suicidal songwriters and their corpus consisted of 35 songs of which lyrics were retrieved from "websites devoted to that artist or his band" (p. 1219) meaning they had no concern for accessing the lyrics through the official websites of the artists or from CD booklets. However, their manual processing afterwards was highly thorough and time-demanding. They listened to all the songs to check for errors and omissions since not all websites provided completely accurate lyrics (ibid.). Taina (2014: 36) corrected obvious errors and added a space after the punctuation marks where missing. Tegge (2017: 91) grouped colloquial words such as *ya* under *you* when lemmatizing the corpus.

   Compound words are another point of discussion. It is uncertain whether words and sequences such as *lunch-time*, *CD*, *English-*

speaking, or *corpus linguistics* count as one word or two (Halliday et al. 2004: 1, Teubert 2004: 86). They may be treated as one word or two (or more) depending on the focus of analysis. Similarly, Meyer (2002: 73) gives an account of how colloquial and linked expressions can be treated by stating that if expressions such as *gotta*, *hafta* are regarded merely as phonetic merging, all instances could be transcribed as *got to* and *have to*. However, if they are treated as single lexical units, then they must be treated different from their correct spellings. Although Meyer's remarks are on spoken language, it is not unrelated to songs as lyrics stand somewhere between spoken and written language (Kreyer and Mukherjee 2007, Werner 2012 and Eiter 2017). Öztürk (2017) has tackled the issue in more detail by changing all colloquial expressions and number to their written forms (e.g., *gonna* becomes *going to*, *17* becomes *seventeen*) and writing the contractions in full (e.g., *I'll* become *I will*) so that the analysis software recognizes these words and assign them to their respective frequency bands instead of grouping them as off-list words (p. 60).

4. ***Removing repeats:*** Even if the repeats are not shown as 'repeat' in the text, these parts are removed. Falk (2012: 10-11), for example, removed the repeats from the texts. Petrie et al. (2008: 198) removed the parts which were repeated three times or more to maintain only one repetition in each song. Tegge (2017: 91) only removed the repeats at the fade-out parts at the end of the songs.

5. ***Removing a part of the main text:*** This is not a common way of processing the lyrics. In their analysis of expository texts, Nippold et al. (2005) ignored fragments – an approach that affects the integrity of a text. In fact, their study was on the transcriptions of conversations and since the focus of their study was on T-units, they had to eliminate all fragments which did not qualify as sentences or clauses. Tegge (2017) and Öztürk (2017) applied this method on song lyrics by removing some (or all) exclamations from the texts. Tegge (2017: 91) removed words like "*ooh, hmm, doobee, shoobee, na*" which she calls

"non-lexical vocables" but retained "*shh, oops, tada* or *wow*". Öztürk (2017: 60) similarly removed such words as "*oh, mmm, ha, na, whoa, yeah*"

6. ***No modification:*** Lyrics are processed the way they are retrieved from the source without any additions or deletions in the text. Although, Pettijohn and Sacco (2009) applied such a methodology, they acknowledge that the text length variations the repeats may interfere with the results. Kreyer and Mukherjee (2007: 39) and Pettijohn and Sacco (2009: 305) chose to retain all the existing repeats, typos and punctuation. In an unusual methodology, Olivo (2001: 72-3) added the "song titles, liner notes, comics, and shout-outs" in the corpus.

2.3. Lexical Richness

Lexis is a field of linguistics which is increasingly used for distinguishing genres (Hyland 2012: 2). Studies on lexis focus on various aspects of vocabulary using numerous methodologies and terminologies. Despite the fact that some terms are well-established, it is not uncommon that they are also used in different senses. The terms *lexicometry* and *lexical richness* refer to themes, which are shared or excluded from each other depending on various interpretations. Lexicometry (or lexicometrics) is basically the analysis of language to reach quantitative results with regard to its lexical and semantic properties. It provides "systematic and objective results, and this contributes to an objective presentation of the quantified linguistic data" (Assunção and Araújo 2019: 3). It involves computer analysis of frequency of specific items, collocation, keywords (Tristl et al. 2015, Breyer and Schemmann 2018). It was pioneered by scholars such "George Yule (1944), George Zipf (1929, 1935) Gustav Herdan (1964, 1966), [and] Pierre Guiraud (1954, 1960)" and the term was coined by Tournier (1975) (Scholz 2019: 125-6).

Lexicometry differs from lexical richness in that the former entails semantic aspects and lexico-grammar; and it could be used to investigate phenomena such as context-dependent meaning (Breyer and Schemmann 2018: 753), the quantitative aspects of dictionaries (Findler and Viil 1974: 16) or recurrent language patterns so that information could be obtained regarding social groups (Scholz 2019: 124). Scholz (2019: 127) further remarks that it is a data-driven approach to analyze "mostly political" language.

Lexical richness, on the other hand, is mostly related to the quantitative aspect of word use and exclusive of the properties lexicometry possess such as semantics, lexico-grammar, etc. As Miller and Biber (2015: 36) note, lexical studies often involve type-based analyses to investigate the number/ratios of different words in a text, which recalls lexical richness. The discussion on lexicometry will be limited as its only aspect to be used in the present study is the keyness analysis, which will be shed light on following the part on lexical richness. This study will use the term lexical richness as the umbrella term and will attempt to provide details on its sub-elements and measurement methods.

2.3.1. Existing theories and classification

Before going into the details of lexical richness, a clarification of terminology is in order. Therefore, the existing theories and classifications will be introduced as the first step. Then, an integrated schema will be provided which will form the skeleton of the lexical richness analysis of the present study.

> […] lexical richness (e.g. Singh 2001; Daller et al. 2003), lexical density (e.g. O'Loughlin 1995), lexical sophistication or rareness (see Read 2000: 203), lexical variation or variety (e.g. Hyltenstam 1988; Granger and Wynne 2000), lexical individuality or originality, lexical complexity or simplicity, vocabulary diversity (e.g. Ciani 1976; Johnson 1979), lexical range and balance (Crystal 1982), vocabulary richness (e.g. Härnqvist et al. 2003; Sokolova et al. 2006), and vocabulary density are frequently used interchangeably with lexical diversity (e.g. Malvern and Richards 2002), but sometimes exclusively to each other, and sometimes hierarchically (e.g. Read (2000) defines lexical diversity, sophistication and variation as different aspects of lexical richness). Further complications arise when the same term was conceptualized and quantified differently in different studies (Yu 2009: 238).

Yu's (2009) brief account of the related terminology summarizes the problems and includes all possible terms that have been used in lexical studies. The introduction of the terms will follow a top-down approach; hence, it will start with lexical complexity/richness. ***Lexical complexity*** is a term which is used to refer to the vocabulary quality. This usage is more preferred in L2 and SLA research (see: Robinson 2007, Pallotti 2009, Ågren et al. 2012, Bulté and Housen 2012, etc.). From a broader perspective, lexical complexity is a branch of linguistic complexity. Bulté and Housen (2012) define complexity theory at two levels as *global/systemic complexity* and *local/structural complexity*. Although their research is on L2, the notions can correspond to L1 as well. They state that global/systemic complexity is about the size of a person's vocabulary or the structures he knows whereas structural complexity refers to the breadth of the knowledge (p. 25).

Based on this distinction, they summarize the complexity levels according to forty different research papers in an attempt to clarify and illustrate the notion of lexical complexity. These dimensions can be regarded as a guide to navigate through the jungle of terminology and approaches. They propose three levels of complexity as *theoretical*, *observational* and *operational* which is given in Figure 1. It is seen that the systemic lexical complexity is observed in three levels as lexical density, lexical diversity and lexical sophistication. Lexical density is operationalized (or measured) by lexical word/function words formula, TTR related indices or number of types and lemmas. Lexical diversity is also operationalized in these two methods plus LFP or Lambda (P_Lex). Lexical sophistication is measured by LFP or Lambda as well. The fourth observational level and the corresponding operational methods morphemes/words and syllables/words are introduced by Bulté and Housen (2012) but they are beyond the scope of the present paper.

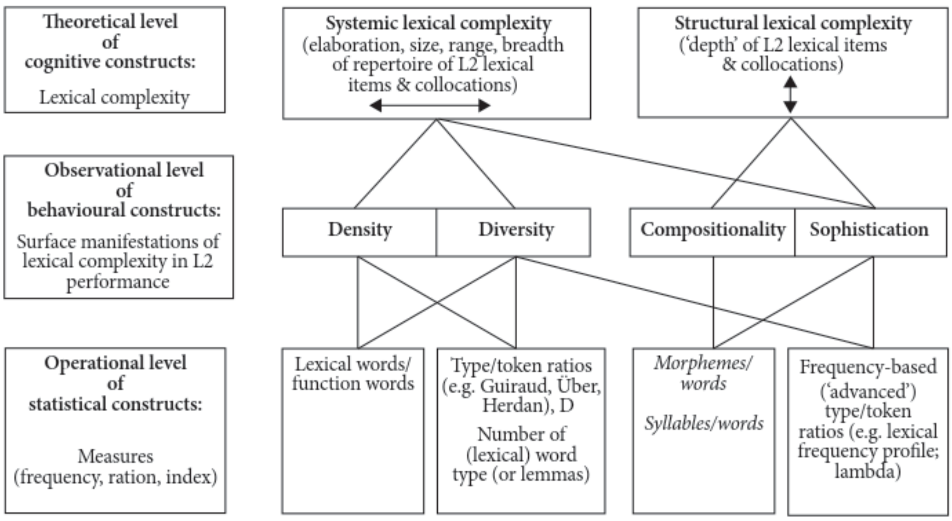

Figure 1: Lexical complexity at different levels of construct specification (Bulté and Housen 2012: 28)

Despite the fact that the diagram given in Figure 1 is a very beneficial attempt in distinguishing the terminology and clarifying the levels, more confusion is also likely. Lexical density is shown to be operationalized in TTR and this makes is synonymous with lexical diversity. On the other hand, lexical diversity is shown to be measured in lexical words/function words formula which, in turn, deems lexical density and diversity interwoven or even similar. This, coupled with the fact that LFP is used both for lexical sophistication and diversity, renders the schema quite complex for the intricate network it has.

They also provide a list of lexical complexity measures which are simplified and grouped under "appropriate" headings. Their list is adapted into Table 7 combined with that of Tonkyn (2012). It must be noted that they do not provide these list as the complete measurement methods but rather as their choices in their research.

Table 7: Measures of lexical complexity proposed by Bulté and Housen (2012) and Tonkyn (2012).

	Bulté and Housen (2012)	**Tonkyn (2012)**
Diversity	Number of word types	D
	TTR	Word types
	Mean segmental TTR	Word families
	Guiraud Index	
	(Word types)2 / words	
	D	
Density	Lexical words / Function words	
	Lexical words / Total words	
Sophistication	Less frequent words / Total words	'Rare' word tokens
		'Rare' word types
		'Rare' word families

While lexical complexity is used synonymous to lexical richness, the latter is used synonymous to other terms more than the former. ***Lexical richness*** is "the quality of vocabulary in a language sample" (Malvern and Richards 2013). Lexical diversity and lexical richness are often used interchangeably (see: Meara and Bell 2001, Daller et al. 2003, Gregori-Signes and Clavel-Arroitia 2015) as well as vocabulary richness (Arnaud 1984, Wimmer and Altmann 1999). "Lexical richness is often measured either by the traditional type-token ratio (TTR) or by its square root variant, the index of Guiraud" (Daller et al. 2003: 197).

Laufer and Nation (1995) define lexical richness as "the degree to which a writer is using a varied and large vocabulary" (p. 307). They do not limit lexical richness exclusively to TTR but consider it an umbrella term, which covers various other measurement methods such as "lexical originality (LO), lexical density (LD), lexical sophistication (LS), and lexical variation (LV)". Although there are various aspects of lexical richness which could be assessed through certain formulas, it also varies depending on the "familiarity with the topic, skill in writing, and communicative purpose" (Laufer and Nation (1995: 308-9), etc.

While Laufer and Nation (1995) propose a broad and inclusive coverage for lexical richness, Gregori-Signes and Clavel-Arroitia (2015) and Daller et al. (2003) use the term interchangeably with lexical diversity and mostly narrow it down to TTR and TTR-related measures. Malvern et al. (2004: 5) place lexical diversity and lexical sophistication under lexical (vocabulary) richness. Guo et al. (2013: 225) subsume lexical diversity under lexical sophistication as a feature of the software Coh-metrix (Graesser et al. 2004, McNamara and Graesser 2012). In a clearer classification, Read (2000: 200-1) identifies four components of lexical richness as, lexical variation (measured through TTR), lexical sophistication, lexical density and number of errors. Since the discussion here is limited to native speakers, number of errors will be disregarded. The measurement methods will be discussed in the following sections.

2.3.2. Observational level

According to Bulté and Housen's (2012) classification, lexical complexity (i.e., lexical richness) could be observed in some surface manifestations. This level could be regarded as the realization of lexical richness across different dimensions which are hierarchically sub-elements of lexical richness.

Lexical diversity may refer to the proficiency level of a language user or the quality of the user's oral or written products (Yu 2009: 238). Essentially, it is an index of the amount of different words used in a text or speech (Johansson 2008: 61). Lexical density refers to the "range and variety of the vocabulary" produced by a person but not that person's potential vocabulary (McCarthy and Jarvis 2007: 459).

> Lexical diversity is often used as an equivalent to lexical richness (e.g., by Tweedie and Baayen 1998, Daller, van Hout & Treffers-Daller 2003). However, Malvern et al. 2004 […] [state] (along the lines of Read 2000) that the lexical diversity measure is only one part of the multidimensional feature of lexical richness (Johansson 2008: 62).

Malvern et al. (2004: 5) use lexical diversity to mean ***lexical variation***. Lexical diversity is also called as *lexical variation*, *lexical variety* (Vögelin et al. 2019: 52, Jarvis 2002: 57) or *lexical specificity* (Biber 1988: 238). Toolan (2009: 34-5) goes as far as to state that lexical diversity could be called as lexical density. Skehan and Foster (2012) remark the use of TTR to measure lexical diversity and Gregori-Signes and Clavel-Arroitia (2015: 548) state that "traditionally, lexical diversity has been measured using the TTR". Lexical variation is synonymous to lexical diversity and "usually measured by the type-token ratio" (Read 2000: 200). Kuiken and Vedder (2007), on the other hand, use the term to encompass lexical sophistication and Johnson (2011) addresses lexical variation as lexical sophistication. These definitions and approaches show that lexical richness, lexical diversity, lexical variation, lexical sophistication and lexical density seem to intersect at some point.

Lexical individuality is another measure of lexical richness proposed by Linnarud (1986) which indicates the use of words only by one person (unique words) (Daller et al. 2003: 202). It is basically the same notion as Laufer and Nation's (1995) ***lexical originality (LO)***. It is measured by counting the words used only in one text across others in a group of text writers. It takes into account the number words which are used by only one person and no one else in a group. It is not a stable measurement method, though, as the results may change when the group members change. Laufer and Nation's (1995: 309) lexical originality formula is:

$$LO = \frac{Number\ of\ tokens\ unique\ to\ one\ writer\ \times 100}{Total\ number\ of\ tokens}$$

Lexical sophistication (rareness/rarity) is "a measure of the proportion of relatively unusual or advanced words" (Read 2000: 203). These less frequent or rare words are called *advanced* (Daller et al. 2003: 202-3), *difficult* (Meara and Bell 2001: 9) or *peripheral* or *marginal* words (Holmes 1994: 87). Lexical sophistication is "the percentage of 'advanced' words in the text" and it is a rather subjective method of measurement as the definition and levels of sophistication are subject to variation from one researcher to another, which eventually requires a standard by which to specify the advanced words (Laufer and Nation 1995).

The general assumption is that frequent items are relatively common or basic and infrequent items are relatively rare or advanced (Daller et al. 2003: 203). Lexical sophistication has been found to increase, thus causing a lower proportion of most frequent words, as the proficiency level increases, which indicates that the measure is capable of differentiating well between the levels (Vidaković and Barker 2009). Similar results have been found by Read and Nation (2006: 11) indicating that low frequency words are used by more proficient speakers in higher proportions and that there is a gradual decrease in low frequency vocabulary use as the proficiency level decreases.

Lexical density, in its basic definition, is the ratio of content words to all tokens in a given text. In other words, it is a measure of the proportion (or percentage [Laufer and Nation 1995: 309]) of lexical items (i.e., nouns, verbs, adjectives and some adverbs) in the text (Johansson 2008: 61) which was introduced by Ure (1971 qtd. in Johansson 2008). It determines between the orality and literacy of the texts as a higher level of lexical density is indicative of a more literate text (O'Loughlin 1995, Read 2000). Hence, it can be used to discriminate between registers and written and spoken language (Read 2000: 200, Camiciottoli 2007: 73). Conversational language contains fewer content words than written language due to the fact that the latter is more informative than the former. This is reflected in the degree of the concentration of ideas and information (Read 2000: 200). In order to clarify the terminology involved, content words will be expanded on in the following section.

2.3.3. Operational Level (Measurement Methods)

The operational level of Bulté and Housen's (2012) classification is actually the measurement methods of various dimensions of lexical richness. There are numerous formulas and calculation methods for different dimensions of lexical richness. The present paper will limit the measures to the most commonly known or used ones. Before going into details of individual methods, some broader views could be cited. For instance, Vermeer (2000: 66) illustrates the measures of lexical richness as shown in Table 8:

Table 8: Measures of lexical richness (Vermeer 2000: 66).

Measure	Label	Formula
1. Total number of words	Tokens	N
2. Number of different words	Types	V
3. Number of different dictionary entries	lemmas	
4. Number of types occurring only once	hapaxes	
5. Type/token ratio	TTR	V/N
6. Corrected TTR	TTR (c)	$V/2\sqrt{N}$
7. 'Indice de Richesse'	Guiraud	$V/\sqrt{N}$
8. Index of Herdan	logTTR	$_{\log}V/_{\log}N$
9. Uber index	Uber	$(_{\log}N)^2/(_{\log}N{-}_{\log}V)$
10. 'Theoretical vocabulary'	Menard	

Her definition of lexical richness assumes a synonymy to lexical variation. The first four measures are not ratio based and therefore is not comparable across texts of different lengths. The remainder of the measures is ratio-based which are realized in different formulas claiming to be independent of text length; however, that is seldom the case (Vermeer 2000: 67). She defines the theoretical vocabulary of Menard (1983) as equalizing the text lengths based on the shortest one which is, in fact, a threat to text integrity and originality.

Table 9: Measures of lexical richness (van Hout and Vermeer 2007: 94).

Lexical Measure	Label	Formula
1. Number of words	Tokens	N
2. Number of different words	Types	V
3. Number of types occurring only once	V_1	V_1
4. Type/token ratio	TTR	V/N
5. 'Indice de Richesse'	Guiraud	$V/\sqrt{N}$
6. Herdan's Index	logTTR	$_{\log}V/_{\log}N$
7. Uber's index	Uber	$(_{\log}N)^2/(_{\log}N{-}_{\log}V)$
8. VOCD	D	$TTR= [D/N][(1+2N/D)^{1/2} - 1]$
9. Zipf' Z		
10. Yule's K		

Van Hout and Vermeer (2007) updated the measures of lexical richness by removing number of lemmas, CTTR and Menard's "theoretical vocabulary" and adding VOCD, Zipf's Z and Yule's K, which will be briefly discussed below.

Meara and Bell (2001: 6) provide a similar table of measures, which excludes "number of different dictionary entries", Corrected TTR and Uber index. They define these seven methods as ***Intrinsic Measures of Lexical Variety*****,** which means that the focus is on the words as they appear without any kind of categorization. Instead, they propose that ***Extrinsic Measures of Lexical Richness*** are necessary for a better assessment of text quality – such as LFP and their proposal P_Lex. With this distinction, Meara and Bell (2001), in a way, reshape the terminology by suggesting that lexical variety refers mostly to TTR and related indices whereas lexical richness is measured by LFP, P_Lex and similar methods. They show the difference through the following sentences:

Example 1: The man saw the woman.

Example 2: The bishop observed the actress.

Example 3: The magistrate sentenced the burglar.

Meara and Bell (2001: 6)

TTR or any related measure (such as Guiraud [Daller et al. 2003: 203]) will inevitably give the same result for each one of these sentences as they all contain the same amount of types and tokens. However, there is a clear difference of difficulty between the words *man* and *magistrate* or *woman* and *burglar*. According to Meara and Bell, the most ideal way to differentiate the richness levels sentences like these is the implementation of Extrinsic Measures of Lexical Richness. Skehan (2009) observes a similar distinction and proposes the terms ***text-internal*** (measured through TTR or a related index) and ***text-external*** (measured through LFP or Lambda) measures. The latter type is observed as lexical sophistication (Skehan 2009: 514). Kuiken and Vedder (2012) use LFP and corrected TTR in two different studies but group these two measures under the title of lexical variation, which blurs the terminology distinction even more. Finally, Daller and Xue (2007: 153) address the same distinction using the terms ***word-list-free*** and ***word-list-based*** approaches.

2.3.3.1. Measures of Lexical Variation (Lexical Diversity):

There are various measurement methods of lexical variation from basic divisions to very complex formulas. All the work in these complex calculations have been developed to eliminate (or minimize) text length effect. However, all LV measures have been found to be affected by text length in varying degrees and thus should be used with caution (McCarthy and Jarvis 2007, cf. McCarthy and Jarvis 2010). In their seminal work, Tweedie and Baayen list some of these measures as follows:

Table 10: Measurement Methods of Lexical Richness (adapted from Tweedie and Baayen 1998)

Basic measures	Vocabulary size
	Hapax legomena
	Mean word frequency
	Type-Token Ratio (reciprocal of mean word frequency)
Measures based on simple transformations	Guiraud
	Herdan (1960, 1964)
	Rubet
	Maas (1972)
	Dugast (1978, 1979)
	Luk'anenkov and Nesitoj (1975 qtd. in Tuldava 1977)
	Brunet (1978)
Measures Using Elements of the Frequency Spectrum	Honoré (1979)
	Hapax dislegomena Sichel (1975)
	Reciprocal of hapax dislegomena Michéa in 1969 and 1971
	Good (1953)
	Entropy
	Yule's K (1944)
	Simpson's D
Parameters of Probabilistic Models	Orlov's generalised Zipf model (1983)
	Sichel's generalised inverse Gauss-Poisson model

They divide the methods into four categories based on their complexity. They tested these formulas and found that all are text-length dependent. It needs to be noted that the models in the last three categories – excluding Guiraud – described in Table 10 are based on the urn model, which is derived from the randomness theory. Basically, the model considers all words as marbles in an urn (text) and the calculations are made through random selection of words from the urn. This method seems to be mathematically appropriate as a more sophisticated calculation approach. Nonetheless, it disregards the text cohesion and integrity as a side effect. As Herdan (1966: 96) notes, “A contiguous passage seems very far from a random sample: the words hang together round the same sequence of ideas and therefore particular words tend to be repeated”. Tweedie and Baayen (1998) tested the validity of the urn models and found that they are not stable across different text lengths and that the discourse structure has an effect on these measures (p. 349). Van Hout and Vermeer (2007: 105-6) note that “a lot of unwanted fluctuations enter the measurements, […] especially when the lexicon possesses realistic, human-like properties.” Covington and McFall (2010: 95) state that complex measures such as Yule’s K, *Vocd-D*, Tuldava, etc. “[incorporate] statistical assumptions and is not directly equivalent to type–token ratio.” Torruella and Capsada (2013: 448-9) group a condensed amount of these measures into three categories as follows:

> **First class** of indices based on the direct relationship between the number of terms and words (type-token): TTR, RTTR, CTTR.
>
> **Second class** of indices has been developed using formulae based on logarithmic function: Herdan, Summer, Maas, Dugast, Tuldava.
>
> **Third class** of indices is formed by a group of indices obtained from more complex calculations: MSTTR, MTLD, HD-D.

Some of the most widely used lexical variation methods will be introduced in this section. ***TTR*** (type-token ratio) is used as the measurement method of lexical variation (or lexical diversity). TTR only distinguishes different words but not their rarity. Therefore, the results will be of quantitative nature (Laufer and Nation 1995: 310). It is measured by dividing the number of different words by the number of total words. TTR can be formulized parallel to Laufer and Nation’s (1995) formula of lexical variation (LV).

$$LV = \frac{Number\ of\ types\ \times 100}{Number\ of\ tokens}$$

It demonstrates the ratio of different words used in a given text. Informal speech, for instance, is considered to feature a lower rate of different words than written language (Read and Nation 2006: 5). It was also an element of Biber’s (1988) MDA under the category of lexical specificity. Baker (2006: 52) remarks that low TTR levels may indicate simple language which in turn suggests that

TTR "can give an indication of the linguistic complexity or specificity of a file or corpus".

Though it may seem a good method of assessing vocabulary variation based on this formula, it is not without limitations. In fact, its biggest deficiency is its sensitivity to text length, which places the method into the crosshairs of numerous criticisms. Laufer and Nation (1995: 310) criticize LV values on the basis of its instability on short texts and variations in the results from texts of different lengths as the measure is bound to be affected by text length despite the fact that different formulas have been devised to overcome this problem, one of which is to test texts of equal lengths.

Longer texts have lower TTR levels than short texts since as the text becomes longer more function words are used. This renders the use of TTR for developmental purposes questionable. "TTR is only possible to use when comparing texts of equal length" (Johansson 2008: 63). Park (2013) compared the lexical variation levels of university students and found that TTR levels are inversely correlated with the proficiency level. The reason is that more proficient students tend to write longer and thus the number of new types decreases, affecting the TTR values negatively.

Vermeer (2000: 69) strongly criticizes TTR and asserts that it is the worst measure of lexical richness. She even goes as far as to say "never apply TTR or logTTR (p. 79). She reports Van Helvert's (1985) study where the TTR scores of the researcher and the children who learn Dutch as Second Language were found similar. The research showed that TTR performs especially poor in longitudinal studies (p. 67). Jarvis et al. (2003) compared the texts produced by students of different proficiency levels using 21 linguistic criteria and the TTR results were not found to be statistically significant. Vermeer (2000) reports that TTR (as well as Guiraud) does not perform well beyond 3,000 words, although it provides relatively reliable results at the earlier stages of vocabulary acquisition. She attributes the low reliability level beyond 3,000 words to the more frequent use of function words (p. 78).

The fact that TTR is heavily dependent on text length does not necessarily disqualify its use in text analysis. In some analyses, this could be the very thing to look for. For instance, a person's vocabulary diversity could be tested by watching the TTR curve reaching a flat line – thus indicating the saturation of the word types for a particular theme (McCarthy and Jarvis 2010).

The index of **Guiraud** differs from the standard TTR in that it is calculated by dividing the total number of types to the square root of the number of total tokens (G = types /√ tokens). Therefore, it is also called as Root TTR (Park 2013). This formulation enables the calculation to be stable across longer texts. Daller et al. (2003: 200) report van Hout and Vermeer's (1988) and Vermeer's (2000) findings that Guiraud has been found to be stable and performing better than TTR. De Jong et al. (2012) used this method to measure lexical diversity and found that native speakers outperformed non-native ones. Guiraud measure seems to be a better method than TTR but it is found to fail beyond 3,000 words (Vermeer 2000: 65). It needs to be noted that her study was a longitudinal one, which focused on the language acquisition of children between the ages of 4 and 7. Daller and Xue (2007: 151) note that it performs differently in different situations and therefore maintain that researchers can never be sure if Guiraud is a valid option in a particular study. ***Corrected TTR (CTTR)*** is another square-corrected measure based on the "number of types divided by the square root of twice the number of tokens" (Carroll 1964 qtd. in Malvern et al. 2004: 189). Levkina and Gilabert (2012) used CTTR for the measurement of lexical variation in their research on L2 oral production. Park (2013: 141-3) report that Guiraud and CTTR are more reliable measures of lexical variation than TTR. These indices have been found to increase as the proficiency level increases.

There is another method of measurement for lexical richness called "***D***" which was developed by Malvern and Richards (1997). It is based on Sichel's (1986) formula and is calculated through the *Vocd* software which was developed for this purpose (Malvern et al. 2004: 56-60, Malvern and Richards 2012: 2). It runs on random sampling method and determines the TTR values of random samples. Then it turn creates a curve determined by the parameter "*D*" which is indicative of the lexical richness of the given text (Daller et al. 2003: 200).

Daller et al. (2003: 200) gives an account of the comparison of TTR, Guiraud and *D* and reiterate Jarvis's (2002) finding that "*D*" offers a more accurate calculation as Guiraud still falls short of dealing with varying text lengths. They also point to Hoare's (2000/2001: 38) finding that "*D*" seems to be a better predictor of the TTR curve. Vidaković and Barker (2009) used *D* to check lexical development across different proficiency levels and found that lexical variation increased with proficiency. Read and Nation (2006: 10) also found positive correlation between *D* scores and proficiency levels. The method has also been found to give similar results in Cantonese with regard to language development of children (Klee et al. 2004: 1407).

McCarthy and Jarvis (2007) criticize *D* on grounds that it is still sensitive to text length. They state that "the method of measurement employed by *vocd* is intended to compensate for the overly strong effect of word repetition on TTR, but the end result is an overcompensation" (p. 474). They report the software to work well with texts between 100 and 400 tokens.

HD-D (Hypergeometric Distribution) is a similar measure to *vocd-D* developed by McCarthy and Jarvis (2007) and coined by McCarthy and Jarvis (2010) in an attempt to simplify the *vocd-D* measure. They claim that *vocd* generates results based on the *D* coefficient and that "HD-D is what *vocd-D* approximates" (p. 383). The reason for their claim is that it would be a more direct measure to calculate the sums of probabilities (HD-D) rather than calculating them and converting them to a TTR value and then converting back to a *D* value (*vocd*) (ibid.). The index calculates the probability of finding a word type from a sample of random 42 words and the final HD-D result indicates the sum of probabilities for finding that type (Lissón and Ballier 2018: 8). HD-D and *vocd* generate similar results and could replace each other (Koizumi and In'nami 2012, Fergadiotis et al. 2013).

In ***Mean Segmental TTR (MSTTR)***, the lengths of the texts to be compared are standardized based on a fixed number of tokens taking into account the shortest text size and applying it to the others (Malvern et al. 2004: 25). It is also called as Split TTR (Jarvis 2002: 59). Ellis and Juan (2004) applied this method to students' narratives by splitting each essay into 40-word segments and found that the results were stable across all three groups. This finding suggested that planning time on tasks had no effect on the lexical diversity of the products as measured by MSTTR. Ellis and Yuan's (2005) follow-up study compared written texts to oral ones using MSTTR and the results indicated significant differences between these registers, yielding higher scores for the written language. Frear and Bitchener (2015) used this method to compare tasks of different complexity levels. They found differences among the tasks albeit not statistically significant.

Moving-Average Type–Token Ratio (MATTR) is proposed by Covington and McFall (2010) in an attempt to eliminate the text length effect as is the case with all the other measures. The method could be visualized as a window of a fixed size (e.g., 500 words) which moves across the text one word at a time and calculates the TTR within the window. The mean score that is obtained at the end of the text is the MATTR value. The difference between MATTR and MSTTR is that MSTTR uses random window frames, which do not overlap with others. Hence, Covington and McFall (2010: 96) argue that MATTR yields a value for every point in the texts and that MSTTR could only be viewed as an approximation of MATTR. They conclude that "MATTR is better for tracking changes within texts, and MATTR is not affected by accidental interactions between segment boundaries and text unit boundaries" (ibid.). Fergadiotis et al. (2013: 11) remark that "a great advantage of MATTR is its face validity because it is equivalent to TTR and fairly straightforward to grasp and explain".

Measure of Textual Lexical Diversity (MTLD) is another segment-based measurement method, developed by McCarthy (2005 qtd. in Šišková 2012: 30), which aims at comparing text parts of equal characteristics. The calculation is made on software developed for this purpose, which divides the given texts into a number of segments each of which has a TTR value of 0.72. The total number of words in the text is divided by the total number of segments. A higher value indicates a more complex text (Johnson 2011).

McCarthy and Jarvis (2010: 390) claim that there is a high correlation between the results of MTLD and other sophisticated methods and that it is not affected by text length. It is considered a reliable measure as all the segments "reach the stabilization point of TTR" (Lissón and Ballier 2018: 7). Fergadiotis et al. (2015: 849) consider it a useful tool and advocate its use. However, Koizumi and In'nami (2012: 554) assert that MTLD should be used with more than 100 words and that “MTLD values can be compared between texts across 100 and 200 tokens”.

Indicated as W, ***Brunette’s index*** (1978) is one of the simple transformations of the classical type/token based calculation methods (Tweedie and Baayen 1998). “W is a measure of lexical richness that relates how varied the vocabulary is for a given piece of text”. The values usually range between 10 and 20 and a lower value indicates a more lexically rich language (Singh 2001: 254-5). Tweedie and Baayen (1998: 332) found that the W value is monotonically increasing with the text length. ***Yule’s K*** method is based on a Poisson distribution assuming that the occurrence of a word is by chance (Holmes 1992: 94). Relying on hypergeometric distribution, the method determines the sums of probabilities with a sample size of two words (McCarthy and Jarvis 2010: 384). It is based on and sensitive to the repetition of words (Lissón and Ballier 2018) and one of the methods which is less affected by text length and operates reliably with texts containing between 100 and 500 tokens (McCarthy and Jarvis 2007: 481). The lower the value, the richer the vocabulary (Singh 1994: 489). For example, Oakes (2008) notes that “using Yule’s own data, works by Gerson had K = 35.9, works by Kempis had K = 59.7, and “De Imitatione Christi” had K = 84.2”.

Maas (1972) is based on the idea of fitting the TTR curve to a logarithmic one. As a result, the TTR curve falls as a function of sample size (Fergadiotis 2011: 24-5). Although McCarthy and Jarvis (2007) found it to be among the five best performing method out of a total of 14, Fergadiotis (2011: 101) claim that it is not a good predictor of lexical diversity. ***Uber index (U)*,** proposed by Dugast (1978), was found to decrease monotonically with text length which makes it a text-length-dependent measure (Tweedie and Baayen (1998: 328-32). It is also reported to have weak correlations with lexical sophistication measures (Šišková 2012: 33). However, Šišková’s another finding indicates that Uber correlates highly with Guiraud and *vocd-D*. ***Honoré’s (1979) Statistic (H)*** is a measurement method for lexical diversity, which “tests the propensity of an author to choose between the alternatives of employing a word used previously or employing a new word” (Holmes 1992: 93). The resulting values are usually between 1,000 and 2,000, where a higher value means more complex language. However, H has been found to decrease as the text length increases (Tweedie and Baayen 1998 and Singh 2001).

As can be noted, the measurement methods span from simple counts to highly complex models. There are other versions of TTR such as Herdan's index, Rubet's K, Simpson's D, etc. (Holmes 1994, Tweedie and Baayen 1998), which are developed using different formulas in an attempt to prevent the inverse curve of type distribution and subsequently eliminate the text length effect. However, they do not address the issue as efficiently as the claim but "merely change the shape of the curve or alter the scale" (Malvern et al. 2004: 30). It mostly depends on the aims of the research to choose a method. It still remains unclear whether such advanced measurement methods are indeed necessary or valid. Daller et al. (2007: 114) note that,

> Guiraud's Index is often a better transformation, at least from the perspective of concurrent validity. It suggests that taking a square root is a happy medium between doing nothing to the number of tokens (TTR) and applying too strong a transformation (as in Herdan's logarithm) that levels out all relevant differences.

2.3.3.2. Measures of Lexical Sophistication (rareness/rarity)

Rarity measures can be classified as *intrinsic* and *extrinsic* – in the same vein as lexical variety (Meara and Bell 2001). An intrinsic measure of rarity could be *hapax legomena* (words occurring only once in the text), *hapax dislegomena* (words occurring twice in the text) (Malvern et al. 2004: 125) or mean word length. Hapax legomena are usually larger in number than hapax dislegomena (Baayen and Lieber 1997: 281). Mean word length is based on the fact that shorter words are more frequently used than the longer ones; hence, they are indicative of a more lexically sophisticated language (Yoon 2017: 134).

As for the extrinsic measures, two methods mentioned by Daller et al. (2003: 203) are Laufer and Nation's LFP and Meara and Bell's P_Lex, which are both frequency based methods. Basically, lexical sophistication is the ratio of advanced words in a text as shown in the formula by Laufer and Nation (1995: 309) below.

$$LS = \frac{Number\ of\ advanced\ tokens\ \times 100}{Total\ number\ of\ lexical\ tokens}$$

As the vocabulary knowledge increases, so does the use of infrequent words (Laufer and Nation 1995: 317 and Meara and Bell 2001: 9). Lexical variation measures (TTR and Guiraud) are of quantitative nature whereas lexical sophistication measures provide qualitative results (Daller et al. 2003: 203). Therefore, lexical sophistication (rareness) measures can be an alternative for a deeper and quantitative analysis of lexical richness. They can provide more reliable results "in relation to different types of lexical frequency benchmarks" (Tonkyn 2012: 224). Lexical sophistication is usually measured based on the ratio of words beyond the 2,000 words in the GSL as proposed by Laufer and Nation (1995: 150) (Malvern et al. 2004, Kuiken and Vedder 2012, Read 2000, Read and Nation 200). In other words, it is often measured through LFP (Park 2013: 135).

2.3.3.2.1. Lexical Frequency Profile (LFP)

Perhaps the most commonly used measure of lexical richness is the lexical frequency analysis, or Lexical Frequency Profile (LFP) as proposed by Laufer and Nation (1995). LFP basically demonstrates the frequency levels of words in a given text based on various points of reference, such as the BNC, the COCA, General Service List, etc.

It is obvious that some words are used more commonly than others, especially in daily speech. Approximately 2,000 words constitute 80% of all the words we meet in everyday language (Meara 1993: 3). 2,000 words are usually set as the limit for the high frequency words (Nation 2001: 22). Low frequency words, on the other hand, are proper names and those which are rarely used such as *eponymous*, *gibbous*, *bifurcate*, *plummet*, *ploy* (Nation 2001: 28-9). The idea behind the lexical frequency analysis is to determine the most frequent words so that a text could be assessed in terms of its lexical richness since a text with a high ratio of low-frequency words is associated with a high level of lexical proficiency (Bardel et al. 2012: 273). The benefits of lexical frequency analysis are that a) it provides reliable quantitative data regarding lexical richness (Laufer and Nation 1995), b) it is fully computer-based and leaves less margin of error and c) an analysis at lexical level would be applicable to most literary genres disregarding the syntactic structures and any complications resulting thereof.

Frequency levels could be found out through judgements and frequency lists. While Ringeling (1984: 61) claims that subjective estimations could be a better way of determining word frequencies than objective counts and Carroll (1971: 12) acknowledges them psychologically more relevant, frequency lists are considered more objective as indicated by Alderson (2007). He compared both methods in his study and his conclusion was that human judgements on frequency cannot substitute corpora based frequency counts (p. 407). Schmitt and Dunham (1999: 407) also point out that frequency intuitions are not sufficient to be able to reach clear-cut results, especially to differentiate between native speakers (NS) and non-native speakers (NNS).

Laufer and Nation (1995: 311) claim that "the LFP shows the percentage of words a learner uses at different vocabulary frequency levels in her writing – or, put differently, the relative proportion of words from different frequency levels". They state the superiority of LFP over LD as the former not being syntax-dependent as the latter. In this respect, LFP is a purely lexical method. This measure could be viewed similar to LS due to its qualitative dimension. However, LFP is more advantageous than LS as it does not simply divide the words as frequent and rare but also classifies them on a scale of frequency. Even though LFP is listed under lexical sophistication here, it is a more revealing measure of lexical sophistication at its core.

Another advantage is that LFP leaves no question as to the reliability of the definition of rare words since it is based on established and standard word frequency lists rather than human judgement (Laufer and Nation 1995: 313). They conclude that LFP has been found to be a valid and reliable measure of lexis, and is able to differentiate between different proficiency levels (p. 319). LFP is a unique method of measuring lexical richness as it is computer based and both quantitative (percentage of frequency levels allow comparative numerical assessments) and qualitative (the rarity/sophistication of vocabulary is measured) at the same time.

Vermeer (2000: 65) reports that lexical richness measures should be based on frequency of the words rather than the relation of types and tokens. This definition provides support for the usefulness of the LFP method. Meara and Bell

(2001: 8-9), on the other hand, criticize LFP on grounds that it is not reliable on short texts which contain fewer than 200 words and that the division of infrequent words in the results are not detailed enough. Laufer (1995: 267) also notes that the LFP method was found to be stable for texts which have between 200 and 400 words. Crossley et al. (2013) advocate a count-based index, as proposed by Tuldava (1996 qtd. in Crossley et al. 2013), rather than a frequency band index as in LFP in the calculation of lexical sophistication. They exemplify their method as follows:

> *The* (BNC frequency = 6,041,234) *cat* (3844) *sat* (11,038) *on* (729,518) *the* (6,041,234) *mat* (569) works out to an average word frequency value for the text of 2,137,906.17 (SD = 3,036,494.47). In contrast, the text *The* (6,041,234) *lizards* (196) *basked* (47) *in* (1,937,819) *sunshine* (629) *on* (729,518) *igneous* (129) *rocks* (2864) amounts to a word frequency value for the text of only 1,089,054.50 (SD = 2,114,424.55) or about half that of the first text (Crossley et al. 2013: 967).

This method separates lemma frequencies so that each one is assessed based on its frequency range in the corpus. Although it may seem a useful and more detailed method, as the authors acknowledge as well, the results are not as comparable as those obtained by LFP.

"Beyond 2,000" (B2K) is a condensed version of LFP, as coined by Laufer (1995), and it is essentially a grouping of the first two thousand words and the remaining ones (UVL/AWL and Off-List). In such an analysis, the ratio of words which fall outside the first K2 words are taken into consideration. Laufer (1995: 267) asserts that the advantage of this method is the ability to correlate with other studies. Another advantage of B2K is that the obtained scores could be used in statistical calculations and correlations for other lexical proficiency measures (Laufer 2012: 2).

2.3.3.2.1.1. Word Families and Lemmas

An important point of consideration with LFP is the unit of counting. It needs to be emphasized that the unit of measurement is very important in assessing lexical richness. Calculations based on word families, lemmas, types, tokens will inevitably give different results. Therefore, the unit of counting should be clarified prior to any kind of measurement.

In order to set the base, a distinction between the terms is in order. Tokens are all the running words in a text regardless of how many times they are repeated.

Types are the “different lexical items in a text” (Daller et al. 2003: 199). They are identified by their different spellings. For example, *realize* and *realise* will be considered different types in an analysis of corpus which contains texts in American and British English. Lemmas and word families are the groups of words which are packed based on different affixation criteria. Brezina and Gablasova (2013) define the distinction as,

> [A] lemma with the headword *develop* (verb) includes also the inflectional forms *develops, developed*, and *developing*. A word family with the same headword would in addition include adjectival derivatives *undeveloped* and *underdeveloped* as well as the nominal forms *development, developments, developer*, and *developers* (p. 4).

Word families consist of a base word and all its closely related derived and inflected forms regardless of the parts of speech (Bauer and Nation 1993: 253 and Nation 2004: 6). They rely on the transparency of words but sometimes it may not be relevant as in the case of *train* and *trainers*. The term “closely related derived and inflected forms” may not be a clear enough; yet, Bauer and Nation (1993) offer a levels scale for the affixes which can be seen in Table 11:

Table 11: Word Family levels (adapted form Bauer and Nation 1993: 258-62).

Level 1	A different form is a different word. Capitalization is ignored.
Level 2	Regularly inflected words are part of the same family. The inflectional categories are - plural; third person singular present tense; past tense; past participle; *-ing*; comparative; superlative; possessive.
Level 3	-able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-, all with restricted uses.
Level 4	-al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in-, all with restricted uses.
Level 5	-age (leakage), -al (arrival), -ally (idiotically), -an (American), -ance (clearance), -ant (consultant), -ary (revolutionary), -atory (confirmatory), -dom (kingdom; officialdom), -eer (black marketeer), -en (wooden), -en (widen), -ence (emergence), -ent (absorbent), -ery (bakery; trickery), -ese (Japanese; officialese), -esque (picturesque), -ette (usherette; roomette), -hood (childhood), -i (Israeli), -ian (phonetician; Johnsonian), -ite (Paisleyite; also chemical meaning), -let (coverlet), -ling (duckling), -ly (leisurely), -most (topmost), -ory (contradictory), -ship (studentship), -ward (homeward), -ways (crossways), -wise (endwise; discussion-wise), ante- (anteroom), anti- (anti-inflation), arch- (archbishop), bi- (biplane), circum- (circumnavigate), counter- (counter-attack), en- (encage; enslave), ex- (ex-president), fore- (forename), hyper- (hyperactive), inter- (inter- African, interweave), mid- (mid-week), mis- (misfit), neo- (neo-colonialism), post- (post-date), pro- (pro-British), semi- (semi-automatic), sub- (subclassify; subterranean), un- (untie; unburden).
Level 6	-able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-, re-.
Level 7	Classical roots and affixes.

These levels “are based on frequency, regularity, productivity and predictability” (Nation 2016: 37). Level 1 corresponds to word types and Level 2 to lemmas. Levels 3 onward are derivational suffixes. The GSL, the BNC 2,000 and the BNC/COCA 2,000 lists are based on Level 6 word families (Dang and Webb 2016: 155). Bauer and Nation (1993: 257) admit the list to be arbitrary and a certain level could be used depending on the proficiency levels of the learners. The example in Table 12, taken from Nation and Webb (2011: 136), shows the lemmas and word family members of *access*:

Table 12: Lemma and word family members of access adapted from Nation and Webb 2011: 136).

Lemma	**Word Family**
Access (Base word)	Access (Base word)
Accessed	Accessed
Accesses	Accesses
Accessing	Accessing
	Accessible
	Inaccessible
	Accessibility
	Inaccessibility

As can be seen in Table 12, the lemmas of access only include inflectional forms whereas the word family members contain all possible noun, verb and adjective combinations. Not all words can be assigned to lemmas in such a straightforward fashion. For example, whether irregular forms such as *mouse-mice*, *good-better-best* should be included in the same lemma structure is debatable (Nation 2001: 11).

Table 13 exemplifies the distinction among tokens, types, lemmas and families with the example sentence, "*She promised me that she would not do what she had done before and kept her promise*".

Table 13: Tokens, word types and word families.

Tokens (17)	**Types (15)**	**Lemmas (13)**	**Word Families (12)**
She	She	She	She
promised	promised	promised	promised
me	me	me	me
that	that	that	that
she			
would	would	would	would
not	not	not	not
do	do	do	do
what	what	what	what
she			
had	had	had	had
done	done		
before	before	before	before
and	and	and	and
kept	kept	kept	kept
her	her		
promise	promise	promise	

The sentence contains 17 tokens. This corresponds to the total number of words irrespective of their affixes or how many times they are repeated. There is only one word in the sentence which is written three times in the same spelling – *she* – therefore the second and third occurrences of the word are not considered as different types resulting in a total word type number of 15. As far as lemmas are concerned, the inflected forms of *do* (*done*) and *she* (*her*) are counted as one. Although *her* is not an affixed version of *she*, it is considered a lemma of the headword *she* as an inflectional variant (Leech et al. 2001: 5-101). The same rule is applicable for *done* even though *do* is not regularly suffixed with the participle –*ed*. Finally, 12 word families can be identified in

Table 13. The sentence has *promise (v)* and *promise (n)* which can be grouped under the same headword in a word family. Hence, in addition to the lemmas, *promise (n)* is removed as it is the base word regardless of part of speech resulting in a total number of 12 word families.

In the light of this information on word types, lemmas and word families, the question of which level to use arises. The aim of the research usually entails a specific approach in which a preference needs to be made among types, tokens, lemmas and word families. A major approach is that lemma should be preferred for productive knowledge and word families for receptive knowledge (Nation 2016: 25–30).

Lemmas are preferred more for learners with lower proficiency levels (Brezina and Gablasova 2015, Gardner and Davies 2014, Dang and Webb 2016 qtd. in Nation 2016: 31-2). He proposes that for productive purposes, lemma approach could be used but also cautions that this argument needs investigation (p. 34). Another important aspect of the word families is that the frequency levels of individual words differ to a great extent. For instance, the frequencies of the word family members of *agree* based on the BNC are as follows: *agree* (8,057), *agreeable* (394), *agreeably* (70), *agreed* (14,390), *agreeing* (813), *agreement* (13,254), *agreements* (2,704) and *agrees* (938). Thus, in a word family approach, *agreeably* will be considered a high frequency word in spite of the fact that it has only 70 occurrences in the BNC. In short, deciding on the word units depends on whether the focus is on productive or receptive skills and the level of the learners' proficiency.

2.3.3.2.1.2. Frequency Lists

Even as Meara (2005) acknowledges, frequency lists are widely agreed upon as standard practice in categorizing vocabulary. Hence, such analyses can be carried out through frequency software, which runs on pre-established frequency lists. Lexical frequency analyses are carried out based on established frequency lists such as the GSL, the AWL, the BNC and the COCA, which are the most commonly used ones. Nation and Waring (1997) mention other lists of note in addition to the GSL, such as The Teachers Word Book of 30,000 words (Thorndike and Lorge 1944), The American Heritage Word Frequency Book

(Carroll et al. 1971), LOB (Johansson 1978) and Brown (Francis and Kucera 1982). The GSL is one of the oldest frequency lists; yet, the list of the first three thousand words in the BNC covers most of the words in the GSL and AWL combined (Nation 2004: 9). Even more specialized lists are being built such as Nursing Academic Word List (NAWL) (Yang 2015), Medical Academic Vocabulary List (MAVL) (Lei and Liu 2016), etc. Most frequently, lists are based on corpora of which majority consist of written language as it is more labor-intensive to collect and store spoken language samples (Read 2000: 236).

The General Service List (GSL) was developed by West in 1953 and it has been in use ever since. It contains the most frequent 2,000 word families in English which are "selected to be of the greatest 'general service' to learners of English" (Bauman and Culligan 1995) out of a 5 million-word corpus (Coxhead 2000: 213). West's (1953) book is currently out of print but the GSL can be accessed at various online sources such as http://www.sequencepublishing.com/academic.html (Gilner 2011: 66). Frequency was one of the main factors in the selection of the GSL (Nation 2004: 7). West (1953), however, did not base his list only on the frequency but also other criteria to be comprehensive for classroom use (Nation and Kyongho 1995: 37) such as frequency, ease of learning, coverage of useful concepts, and stylistic level (Coxhead 2000: 213).

The GSL has a wide range of utility despite its age and it offers coverage of over 80% in written texts and 90% in spoken English (Neufeld and Billuroğlu 2005: 4, Gilner and Morales 2008b: 518) which shows that "the GSL has stood the test of time remarkably well" (Neufeld and Billuroğlu 2005: 4). Reda (2003) analyzed English textbooks and found out that the GSL offers major coverage including scientific and technical vocabulary. Sutarsyah et al. (1994: 46) investigated general academic texts in comparison to academic texts on economics and found that the GSL offered coverage of 78.43% in general academic texts and 82.5% in texts on economics. Hancıoğlu et al. (2008: 461) remark that 82.84% of the words used up to the point in their paper where they mention the coverage are in the GSL. They comment on this result as "impressive".

In addition to appraisals, the GSL has received critical feedback as well, most of which are directed at its age and not reflecting current usage and high-frequency English (Hyland and Tse 2007: 249, Gardner and Davies 2013: 5). Cobb (2013: 82) state that the list is small, intuitive and mostly for pedagogical purposes, thus excluding vulgarities. Gilner and Morales (2008a) compared the GSL with the BNC and stressed the advantage of the latter to the former. They reason their claim based on the size of the BNC being ten times larger and the GSL's shortcomings in profiling. Gilner (2011: 70) asserts that although the GSL is a comprehensive list, it does not mean that it "is either perfectly adequate or superior to all other word-lists". Another reason for its criticism is the criteria used for the selection of the GSL words. Due to the fact that they were not only selected based on their frequencies but also some additional qualitative data such as ease of learning, necessity, etc., its objectivity is questioned (Brezina and Gablasova 2013).

As a result of the GSL's old age and the criticisms it has received, the recent years have witnessed two major individual attempts to update – or upgrade – the GSL. Brezina and Gablasova (2013) came up with a new list, titled the new-GSL, as result of a comparative research on 12 billion running words acquired from The Lancaster-Oslo-Bergen Corpus (LOB) (Leech et al. 1970-1978), The British National Corpus (BNC) (2007), The BE06 Corpus of British English (BE06) (Baker 2009), and EnTenTen12 (2017). Although they give credit to West's GSL by acknowledging that it has been the most widely used and influential word list in English, they criticize it on the basis of the inclusion of subjective criteria in the development of the GSL and propose an absolute quantitative approach based on frequency, dispersion, and distribution across language corpora.

Another concern they put forward is the fact that the GSL was built on the word family principle. Brezina and Gablasova (2013) developed the new-GSL based on the lemma principle. They argue that the lemma principle distinguishes between parts of speech since it excludes the derivational forms. Word families rely on the transparency of words but sometimes it may not be relevant as in the case of *train* and *trainers,* as mentioned earlier. It also aims the learners of

English at beginner level as they may not be able to distinguish the morphological differences resulting from derivation (Brezina and Gablasova 2013: 5).

Unaware and independent of Brezina and Gablasova's (2013) new-GSL, Browne (2013) also established a new list called the New General Service List (NGSL), which he claims to have a broader coverage than the GSL by incorporating 273 million words from Cambridge English Corpus (CEC) (Browne 2014). The NGSL was also compiled by the lemma principle, which Browne (2014: 6) coined as "modified lexeme approach".

Both Brezina and Gablasova's (2013) and Browne's (2013) lists have recently been published and have therefore not been mentioned and tested as much as the GSL and the AWL. Moreover, in their comparative study on word lists, Dang and Webb (2016: 153) point out that Browne's (2013) NGSL has the lowest coverage compared to the GSL, the BNC, the BNC/COCA and the new-GSL. The GSL continues to be a reliable source and it is still commonly used. It has undeniable influence on the emergence and content of other lists; and while it might need some modifications it does not necessarily have to go under a massive change (Neufeld and Billuroğlu 2005: 9).

The GSL is commonly complemented with the Academic Word List (AWL), developed by Averil Coxhead as her master's thesis at School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand in 1998 (Coxhead 1998). The AWL features 570 word families which are taken from Humanities, Science, Commerce and Law texts based on their frequencies and range (coverage across various types of texts) (Nation 2004: 7).

The 570 words the list contains exclude the most frequent 2,000 words in the GSL. The fact that the AWL excludes the GSL words is an indicator of their interdependency (Brezina and Gablasova 2013: 5) in a way that the AWL is an extension to the GSL (Gilner 2011, Brezina and Gablasova 2013, etc.). The list consists of 10 frequency bands, the first nine of which have 60 words families each and the last 30 (Yang 2015: 29). Hancıoğlu et al. (2008: 462) remark that the GSL and the AWL words cover 92.71% of the tokens in their paper up to the point of mention. Vongpumivitch et al. (2009: 38) tested the coverage of the

AWL on articles in applied linguistics and found that the list offers a considerable coverage in the said field (11.17%) which is even greater than some of the fields (e.g., Art [9.3%]) which Coxhead (2000) herself analyzed.

On the other hand, Neufeld and Billuroğlu (2005: 7) criticize the addition of the AWL on top of the GSL and assert that the AWL has broader coverage than GSL K2 band in academic texts. They also criticize the AWL on the basis of its coverage not actually being solely restricted to academic purposes but also offering more commonly used words in English. Hyland and Tse (2007) also observe that the AWL does not address all academic fields equally and that more specific, discipline-based lists ought to be created for better precision on learning the exact vocabulary required by each field. Hancıoğlu et al. (2008: 475), on the other hand, criticize the division of a word lists as general and academic and propose a unified one which is equally banded based on their frequencies.

Although the AWL is based on words from academic texts, it has been tested on non-academic texts as well (see Cobb and Horst 2004) and the words in the list could be considered "advanced words" as suggested by Laufer and Nation (1995). It contains important vocabulary for high school and university students (Nation 2004: 7).

Reflecting on his study 11 years later, Coxhead (2011: 356) states that building a word list to complete the GSL was a controversial decision because of the criticisms it was subject to due to its age and asserts that the GSL is yet to be replaced. Gardner and Davies (2013: 5) support this criticism claiming that the AWL already contains high frequency words from the BNC and that it should not be regarded as an appendage to the GSL.

Billuroğlu and Neufeld created a frequency list named after themselves, the Billuroğlu Neufeld List (BNL) which consists of the words in the GSL (1,983) and the AWL (570) complemented by 183 new "contemporary" words added by the authors. They tested the BNL against GSL+AWL on the same texts and found that the coverage reaches to approximately 90% with the BNL compared to ca. 86% with GSL+AWL. It is also reported that their division of frequency bands is six instead of five to be more exact in the specification of frequency levels (Neufeld and Billuroğlu (2005: 11-2).

Leech et al. (2001) created the BNC word frequency list based on the frequency, range and dispersion of the words. Here frequency corresponds to the number of occurrences of a word, range in how many texts the word occurs and dispersion how evenly the word occurs across texts (Nation 2004: 4). Nation (2004) created a BNC frequency list from the most frequent 6,500 lemmas which occur more than 10,000 times or more in the whole BNC – 100,000,000 words. He expanded the most frequent 1,000 lemmas into word families, including it the days of the week, months, letters of the alphabet and numbers although not based on frequency, range or dispersion criteria (p. 5). He also established the second and third 1,000 frequency bands similarly. Then, the list expanded to 14 (Nation 2006) and 25 (Nation and Webb 2011) 1,000 frequency bands. The first 20 word family frequency bands are based on frequency and range values – not dispersion (Nation and Webb 2011: 147-8). The five remainder bands are proper names (K21), non-words and marginal words (e.g., *ah*, *aargh*, *um*, *whew*, etc.) (K22), transparent compounds (e.g., *schoolhouse*, *long-term*, *shoemaker*, etc.) (K23), remainder of existing word families (K24) and unchecked items (K25) (Nation and Webb 2011: 138-41).

The BNC is widely used in various studies such as, the creation of PHRASE (Phrasal Expressions List) (Martinez and Schmitt 2012), evaluation of corpus homogeneity (Kilgarriff 1997), comparison of corpora (Rayson and Garside 2000), text genre detection (Stamatatos et al. 2000), adjusting the frequency lists based on chunks (O'Donnell 2011), investigating the teachers' word frequency intuitions (McCrostie 2007), frequency of "core idioms" (Grant 2005), transitive and intransitive uses of verbs (Newman and Rice 2006), investigating the frequency characteristics of American Sign Language (Morford and MacFarlane 2003), investigating the frequency models in SLA (Brown 2012), lexical coverage of movies (Webb and Rodgers 2009a), frequency analysis of subtitles (Baayen et al. 2016) and synonymous adjective-noun collocations (Sonbul 2014).

Gilner (2011) compared the BNC to the GSL and found that the first 2,000 word families in both lists correspond to each other to a great extent on an item per item basis. Nation (2004) compared GSL+AWL with BNC 3,000 – the first three thousand words in the British National Corpus – and found that both lists are very close to each other despite the fact that the former (especially the GSL) was introduced long before the BNC. He observed as the main difference between the two groups that the distribution of the most frequent 3,000 words across one thousand levels is different as the first 2,000 words in the BNC contains many words from the AWL (p. 12).

COCA frequency lists are more sophisticated as they are relatively newer lists and more reliably calculated (and even compared to other lists) as a result of a web-based technological infrastructure. The full lists can be obtained through https://www.wordfrequency.info. There are two kinds of lists available – in tokens and lemmas. The token list reaches up to 100,000 words whereas the lemma list goes up to 60,000 (https://www.wordfrequency.info/100k_compare_to_60k_etc.asp). The lists were created based on frequency, range and dispersion, and queries based on genres, collocates and n-grams can be made.

Schmitt and Schmitt (2014: 494) claim that it is the best corpus of general English with regard to size, balance and currency. The COCA was used in the creation of "A Frequency Dictionary of Contemporary American English" by Davies and Gardner (2010) and the Academic Vocabulary List (AVL) (Gardner and Davies 2013), reassessment of L2 vocabulary frequencies (Schmitt and Schmitt 2014), developing a list of phrasal words (Garnier and Schmitt 2015), POS-tagging SUBTLEX-US (subtitle corpus) (Brysbaert et al. 2012), examining chunking ability (McCauley and Christiansen 2015), individual differences in perception of certain structures compared against Google corpus (Caldwell-Harris et al. 2012), studying multi-word expressions (Martinez 2013), syntactic simplification (Medero and Ostendorf 2011), etc.

As mentioned in *Corpus Linguistics*, the texts in the BNC were created between 1960 and 1993 whereas the oldest texts in the COCA dates back to 1990 and the whole content extends up to 2017 as of April 2019. Thus, in terms of the year spans they cover, "COCA starts […] where the BNC ends" (Liu 2011: 669). Nation and his colleagues integrated the BNC and COCA lists in 2012 so that they can be compatible with the *Antconc* software and provide an as comprehensive as possible list in terms of high, mid and low frequency words. The list is based on word families and consists of 25 1,000-word frequency levels complemented by four more levels (yielding a total of 29 bands). The additional four levels contain proper names, marginal words, transparent compounds and acronyms (Nation and Anthony 2013: 9). The number of the frequency bands was increased to 32 in Nation (2016: 132). It is interesting to note that the first 3,000 words include the majority of AWL words (Masrai and Milton 2018: 44)

The BNC and the COCA are commonly used together in frequency analyses. They have been used in examining the vocabulary of the English-medium traditional Chinese medicine books (Hsu 2018), creating the academic article word list for social sciences (Kwary and Artha 2017), identifying the "semantic structure of the lexical unit *educational management* in the modern English language" (Glukhova 2014), creating a word list for ELT (Kwary and Jurianto 2017), introducing a new word sampling method (Ehara et al. 2014), in the creation of the NGSL (Browne 2014), creating medical corpora (Quero and Coxhead 2018), comparing native speakers' judgements against the frequency lists (Okamoto 2015), creating a list of the most frequent formulaic structures (Hsu 2014), testing aural vocabulary knowledge (Matthews 2017), developing a new vocabulary levels test (NVLT) (McLean and Kramer 2015) and developing a word list on plumbing (Coxhead and Demecheleer 2018). The BNC and the COCA also have been used separately for comparative purposes, as in Liu's (2011) research on the most frequently used phrasal verbs. Other studies combined the BNC/COCA with other lists such as the GSL (Quero and Coxhead 2018, Kwary and Artha 2017), the AWL (McLean and Kramer 2015), NGSL and the new-GSL (Kwary and Jurianto 2017 and Browne 2014).

2.3.3.2.1.3. LFP Software

There is a good deal of corpus analysis software which processes otherwise extremely difficult or impossible analyses easily and quickly (Koester 2012: 49). This kind of software basically runs specific corpora. The frequency lists are usually embedded in online tools or software developed for measuring lexical frequency levels of texts. Different text analysis tools[2] have been used for

[2] Some other analysis tools are Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2015), Wordle (Feinberg 2008), WordSmith Tools (Scott 2017), UAM Corpus Tool (O'Donnell 2016), etc.

the analysis of song lyrics (see Petrie et al. 2008 and Lieb 2011) and various other texts.

Perhaps, the most crucial feature of LFP tools is their comparability as they can run on the same word frequency lists. LFP has been measured mostly through a couple of noteworthy software. Foremost among them is *Vocabprofile*, which Neufeld and Billuroğlu (2005), in their study on the GSL and the AWL, mention as "excellent" software. It is an online text analysis tool which is available at http://www.lextutor.ca/vp/eng/. It is the online version of the software called *Range* which was developed by Heatley, Nation and Coxhead (2002). The web version was adapted by T. Cobb. *Vocabprofile* could be regarded as a 'lite' version of *Range* as the former lacks some of the features, which are available in the latter. Unlike *Range*, *Vocabprofile* basically works with relatively small texts and does not parse multiple texts at the same time. Despite the benefits and ease of use it offers, *Vocabprofile* is not as practical when it comes to profiling multiple texts. For such a purpose, software like *Range* or *AntConc* could be preferred. *AntConc,* in particular, is highly practical as it offers numerous options to fine-tune the results depending on the research questions.

AntWordProfiler, developed by Anthony (2014), is one of the useful tools used for the calculation of lexical sophistication of texts. The software is freely available for download at https://www.laurenceanthony.net/software/antwordprofiler/. The software comes with GSL+AWL word frequency lists; however, the BNC/COCA and the NGSL word lists are also available. The word frequency lists are provided in separate files split in thousand words. This feature enables the users to choose whichever frequency level they would like to analyze. For example, if the aim is to see the words that fall in the Beyond 2,000 range, only the first two frequency files, each containing the most frequent thousand words, could be chosen so that the Beyond 2,000 words could be grouped as Off-List rather than being split into numerous frequency levels. The software analyzes the input text files based on the chosen frequency list and generates a report detailing the frequency distribution of words. A screenshot of the software, taken from the official website, is provided in Figure 2.

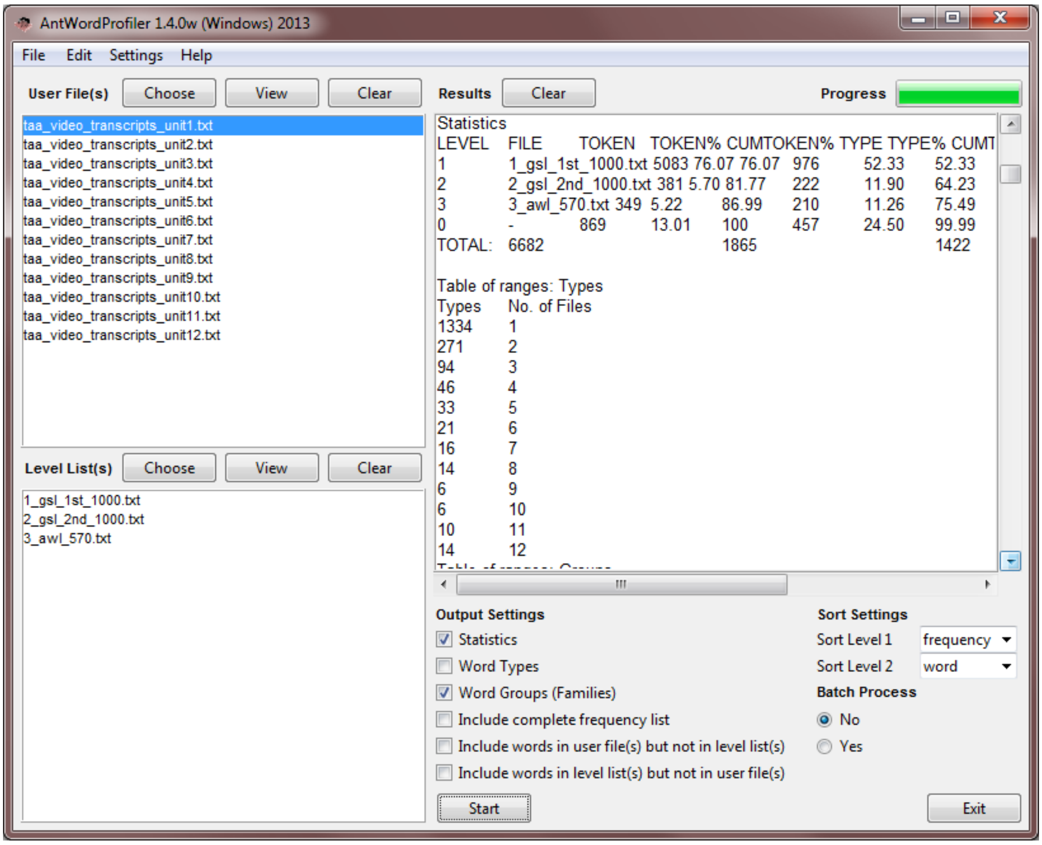

Figure 2: *AntWordProfiler* screenshot downloaded from https://www.laurenceanthony.net/software/antwordprofiler/ on 31.07.2019, 12:40 (Anthony 2014)

The User File(s) pane on top left shows the input files that are to be analyzed and the Level List(s) pane on bottom left shows the frequency list files. The analyses results are displayed in the number and percentages of tokens, types and word families. A sample analysis is shown in Table 14 below:

Table 14: Sample Output from *AntWordProfiler*

File name:										
Number of lines: 28										
Number of types: 88										
Number of tokens: 164										
Level list 1: 1_gsl_1st_1000.txt										
Number of types: 4114										
Number of groups (families): 998										
Level list 2: 2_gsl_2nd_1000.txt										
Number of types: 3708										
Number of groups (families): 988										
Statistics										
LEVEL	FILE	TOKEN	TOKEN%	CUMTOKEN%	TYPE	TYPE%	CUMTYPE%	GROUP	GROUP%	CUMGROUP%
1	1_gsl_1st_1000.txt	146	89.02	89.02	73	82.95	82.95	63	80.77	80.77
2	2_gsl_2nd_1000.txt	9	5.49	94.51	6	6.82	89.77	6	7.69	88.46
0	-	9	5.49	100	9	10.23	100	9	11.54	100
TOTAL:		164			88			78		

The details in the top left of the table are the line, type and token numbers of the input text, which are followed by the level lists loaded by the user. In this case, only the first two thousand word lists of the GSL are loaded and the type and word family numbers for each level list are provided before the results. The statistics are provided in tokens, types and families for each frequency level. Out of the total of 164 tokens, 146 (89.02%) are in the K1 list. The number of total types is 88 and 73 (82.95%) are in the K1 band. Finally, 63 of the 78 word families (80.77%) are in the first thousand list. Level 0, also indicated as "-" refers to the words that not included in the above lists; i.e., Off-List.

Off-List words are those with lower frequency and thus an indicator of higher lexical richness. Thus, the analysis could be read backwards as well – see "Beyond 2,000" [B2K] in *Lexical Frequency Profile (LFP)* – focusing on the ratio of Off-List words rather than that of K1 and K2 words. Their numbers/ratios could also be used as a criterion to assess the lexical richness of a given text. As described in 2.3. Lexical Richness, this notion is referred to as lexical sophistication because it measures the ratio of "advanced words" against frequent words (Laufer and Nation 1995: 309 and Muncie 2002: 227). The 'ideal' output results are 70-10-10-10 percent (K1, K2, AWL and Off-List respectively) based on NS performance for GSL+AWL (Research Uses of *Vocabprofile*). The interpretation of the results may vary depending on the research questions through laying more emphasis on the selected frequency levels.

The difference between the BNC/COCA and GSL+AWL results lie in the number of frequency bands. The former has 25 bands whereas the latter has 3. Hence, the BNC/COCA provides a more detailed division and leaves fewer words in the Off-List. If the aim is a finer distinction of word levels and there is no academic vocabulary concerned, the BNC/COCA can be the list of choice. Another advantage is that it is a newer list as opposed to over 60 year-old GSL. However, GSL+AWL still stand their ground by offering coverage almost as high as the other lists and keep on being inspirations for the development of newer ones.

2.3.3.2.2. Other LS Measures

P_Lex is a measurement method of lexical richness developed by Meara and Bell (2001) in an attempt to improve Laufer and Nation's (1995) LFP. The motivation behind this method is to create a tool that will work with shorter texts (<200 words) unlike LFP. "P-Lex is based on a definition of rare words and therefore needs a word list" (Daller and Xue 2007: 153). The input text is divided into 10-word segments. Then, the numbers of infrequent words in each segment are counted and lambda values for these words are calculated through a series of formulas. Meara and Bell (2001) found the scores obtained by P_Lex comparable to those of LFP, which justifies the validity of the LFP measure as well.

Although P_Lex and LFP provide similar results, as stated by the developer of the former, the two advantages of P_Lex over LFP are 1) its ability to work with short texts and 2) providing results anchored on zero instead of giving percentages as LFP, which makes P_Lex results easier to be statistically tested. A higher lambda value is indicative of a wider range of vocabulary incorporating the use of less frequent words (Skehan 2009: 515). Malvern et al. (2004: 159-60) opted for the use of P_Lex instead of LFP on grounds that the former is sensitive to text length. Skehan and Foster (2012: 207) compared native and nonnative speaker performances based on lambda scores and found that native speakers naturally achieved higher lambda scores.

As for the disadvantages, the reliability of P_Lex on short texts has not been confirmed and it is still unclear how long a text is necessary for a reliable score (Kojima and Yamashita 2014: 26). Another point is that the resulting values are not easy to interpret and compare (ibid.). The fact that P_Lex calculates the easy words based on the first one thousand words in the GSL (Crossley et al. 2013: 970) could be considered another disadvantage of the method as it sets the threshold lower than the other methods.

Kojima (2011 qtd. in Kojima and Yamashita 2014: 24) proposes another lexical sophistication measurement method, **S**, which claims to be less affected by text length. Her motive in developing the measure was to form a calculation method, which is based on relative frequency scales rather than clear-cut ones as in LFP. The method is based on sampling 50 words from the beginning of the text and find “the most frequent 500, 1000, 1500, 2000, 2500, and 3000 words [based on the BNC list]” The procedure is repeated for another 50 words, this time covering the range between the 2nd and the 51st word. The calculation is thus repeated until the final word in the text is reached. Finally, an S value is created based on the best fit between the empirical and theoretical curves (Kojima and Yamashita 2014: 28). In this respect, the model resembles MATTR, where a window size is operationalized to collect continuous samples from a text. The resulting values are not percentages as in LFP, but absolute values. This method is hardly comparable to other methods and studies, as is the case with P_Lex, owing to the fact that they are not used as commonly as LFP.

Another measure of lexical sophistication by Hayes (1996) is **LEX**. Hayes calls it a measure of lexical difficulty but essentially what it measures is the lexical sophistication of texts as it is based on the frequency list of The American Heritage Word Frequency Book (Carroll et al. 1971 qtd. in Lu et al. 2014). It runs on a lognormal distribution of occurrences of words in the corpus of newspapers published since 1665. The LEX values of the newspapers fluctuate so low that it is almost close to zero, i.e., the lexical richness of newspapers remain mostly the same. Therefore, the newspaper lexicon has a value of 0.0 LEX. If a text scores positive values, it means that the language is more complex – and vice versa (Hayes 1996, Lu et al. 2014). Using LEX, Lu et al. (2014) compared the lexicon of books published between 1910 and 2000 – spanning ten decades. They found that 1940s had the lowest values (indicating low lexical sophistication) whereas 2000 had the highest. As is the case with P_Lex and S, LEX is not a common method of measuring lexical sophistication.

Advanced TTR and ***Guiraud Advanced*** could be used in the assessment of LS (Daller et al. 2003: 202-3). Advanced TTR is calculated as the ratio of the number of advanced types to the total number of tokens in a given text which is formulized as follows:

$$Advanced\ TTR = \frac{Number\ of\ advanced\ types}{Total\ number\ of\ tokens}$$

Guiraud Advanced is a similar calculation method to Advanced TTR. It can be considered a combination of Advanced TTR and Guiraud (Root TTR) which is calculated as the ratio of the number of advanced types to the square root of the total number of tokens.

$$Guiraud\ Advanced = \frac{Number\ of\ advanced\ types}{\sqrt{Total}\ number\ of\ tokens}$$

They are both frequency based methods capable of providing qualitative insights and they have clear advantages to the traditional TTR and Guiraud as mentioned by Daller et al. (2003: 217-8). They propose Advanced TTR and Guiraud Advanced as more advantageous measures – especially the latter (p. 197). They carried out their research in Turkish and German. As there are no comparable frequency lists for these languages, they used Oehler's (1983 qtd. in Daller et al. 2003: 208) frequency list for German and seven teachers as judges for Turkish in the specification of advanced words. They found that these measures are more reliable than classical TTR and Guiraud methods. Malvern et al. (2004), too, report the superiority of these methods over the classical variations of TTR.

2.3.3.3. Measures of Lexical Density

There are various measures of lexical density. Since they are all based on ratios of content words, some details are to be provided. In English, words are divided into two categories: function words and content words (Miller et al. 1958: 377). Content words (aka lexical words, major word classes [Aarts and Aarts 1982: 22]) are the words that refer to people, places, things (in the broadest sense), actions, states and properties (Miller 2002: 35), entities (persons and things), names of processes (actions, events, etc.) and names of qualities (Halliday and Matthiessen 2004: 37). Content words consist of nouns, verbs, adjectives and adverbs and carry meaning on their own. These words are also called open-class words (Quirk et al. 1985: 72) since there is no limit to their emergence and expansion.

Function words are the grammatical items such as pronouns, determiners, articles, conjunctions, numerals, prepositions, quantifiers, interjections, auxiliaries and some irregular forms (Aarts and Aarts 1982: 22, Miller et al. 1958: 377). Halliday (1989: 61) also includes some classes of adverbs, and finite verbs. This set of words is called closed-class words as their numbers are limited (Quirk et al.

1985: 71-2). They are “much more resistant to innovations” (Miller et al. 1958: 377); they do not allow the creation of new members, and therefore, can be easily listed (Aarts and Aarts 1982: 22). These words occur so frequently that 45% of the words in academic and 50% in informal NS texts consist of function words (Morris and Cobb 2004: 77). They, in a way, function as glue to hold the sentences together (Miller et al. 1958: 384). Function words have higher frequency ratios than content words. In general corpora, content words dominate the frequency lists after the most frequent 150 words, which are mostly made up of function words (Kennedy 1998: 102)

Although these definitions seem clear-cut, there are many controversies on some word classes with regard to which category they belong to. Aarts and Aarts (1982: 22) keep the borders solid by allowing nouns, verbs, adjectives and adverbs in the open-class category and everything else in the closed-class. The only exception is the acknowledgement of auxiliary verbs as closed-class words. They propose another class, though, called semi-auxiliaries, which are followed by *to*+infinitives. Some examples of these are *tend to, happen to, fail to, seem to, appear to, turn out to, be to, have to* and *be going to*. Their division leaves these semi-auxiliaries on the content side.

Quirk et al. (1985: 73) assert that word classes are heterogeneous and especially adverbs, which are “notoriously heterogeneous”, could fall into either category. They state that adverbs created from adjectives with the *–ly* suffix could be considered open class while *here*, *there* and *now* are closed class. Miller (2002: 35) argues that some modal verbs such as *may*, *must* and *could* and some prepositions (*with*, *from*, *by*, etc.) seem to be on the borderline although they are widely accepted as grammatical items, i.e., function words. He considers these words as content words in his study. Halliday (1989: 63) stands a similar ground, arguing that prepositions and some modal adverbs such as *always* and *perhaps* are on the borderline. He refers to those items as “intermediate cases”. Halliday offers one the most controversial divisions on the subject by considering some finite verbs as function words.

The only real accident that I’ve ever <u>had</u> was in fog and ice.

He argues, in this example, that the verb *had* is a function word although it is not an auxiliary (Halliday 1989: 61). On the other hand, Halliday and Matthiessen (2004: 37) treat *is* and *are* as content words. In their example below, taken from a speech by Nelson Mandela, they point out that the auxiliaries (*is* and *are*) describe names of processes and therefore considered content words.

> *To my compatriots I have no hesitation in saying that each of us* <u>*is*</u> *as intimately attached to the soil of this beautiful country as* <u>*are*</u> *the famous jacaranda trees of Pretoria and the mimosa trees of the bushveld.*

Arnaud (1984: 20) used automatic processing in his analysis and treated some problematic words such as *have* and *do* as grammatical items and did not manually evaluate them. Despite proposing different views on the division between content and function words, Halliday (1989: 63) remarks that it is not important where we draw the line as long as consistency is maintained.

Following the distinction between function and content words, the measurement methods will be introduced. The first and the most commonly used one is the proportion of content words to the number of total words as proposed by Ure (1971). Ure (1971 qtd. in Johansson 2008: 65) proposed that spoken texts have a proportion of content words below 40% whereas written ones have above that value. Laufer and Nation (1995) formulize the method as follows:

$$LD = \frac{Number\ of\ lexical\ tokens\ \times 100}{Total\ number\ of\ tokens}$$

The second method is a tweaked version of the first one. Halliday (1989: 64-5) proposes that, even though lexical density is calculated based on the ratio of lexical items to the total number of words, some highly frequent lexical items, such as "*thing*, *people*, *way*, *do*, *make*, *get*, *have*, *go*, *good*, *many*", should have lower values. This can be called weighted lexical density as the words are scored by their frequencies of occurrence. O'Loughlin (1995) applied this method and the classical lexical density measurement in his study. He set the frequency threshold as the 700 most frequently used words in the COBUILD Dictionary Project (1987) listed in Willis and Willis (1988 qtd. in O'Loughlin 1995: 227). He found that the results were similar if not identical. Johansson (2008), for instance, uses this version of lexical density measurement.

Another method is the calculation of the ratio of content words to function words. This is not as common as the first and the second method. Ishikawa (2007) applied this method together with the first method and the results were mostly parallel to each other. There other variations of lexical density such as noun density, which calculates the ratio of nouns in the text. The same approach could be followed with adjectives or verbs (Johansson 2008: 65).

As seen in the above-mentioned methods, a text is usually considered as a whole and the analyses are carried out disregarding sub-elements of the texts (i.e., sentence, clause, phrase, etc.). From this viewpoint, it is a measurement method for lexical richness, yet it might indicate more sophisticated syntax at the same time. LD also depends on the syntactic properties of texts (Laufer and Nation 1995: 309) as the words' being packed into larger units is what determines the density of a text (Halliday 1989: 66). Halliday's indication of the consideration of "larger units" refers to clauses. Given the fact that Halliday's (1989) study focused on spoken language, where distinguishing sentences is a problematic issue, he proposed a clause-level analysis. He suggests the lexical density analysis be carried out on the basis of clauses by dividing the number of lexical items by the number of clauses. His example analysis is as follows:

> ||| The basic 'stuff' of living organisms is protoplasm. ||| There is no set composition of this || and it varies between one individual and the next. |||
>
> … [T]he three clauses in the above text contain, respectively, five (*basic, stuff, living, organisms, protoplasm*), two (*set, composition*), and two (*varies, individual*) lexical items; a total of nine, giving an average of three per clause. We will therefore say that this text has a mean lexical density of 3.0 (Halliday 1989: 66).

Thus, the formula for the lexical density based on Halliday (1989: 67) is:

$$LD = \frac{CW_C}{C}$$

In this formula, CW_C refers to the number of content words (lexical items) in clauses and C to the number of the clauses. Following Halliday's (1989) approach, in order to measure lexical density at clausal level, it is obligatory to identify both the content words in the text (as discussed above) and the clauses.

2.3.4. Studies on Lexical Richness

Many studies have focused on lexical richness and in doing so, they more often than not incorporate various methods. This section will provide an account of such studies under separate headings. However, it is not possible to clearly separate all the studies since they usually employ a combination of several methods. Therefore, more comprehensive studies have also been mentioned under headings specific to a particular field. An account of comparative studies – comparing materials or methods – will also be provided in a separate section. Lexical analyses have also been carried out in authorship attribution studies as well, in order to assign a text to a particular writer (see Holmes 1992 and 1994). This objective entails a bottom-up approach whereas comparing the texts of authors so as to distinguish their styles is of top-down nature. Hence, such studies are not taken into consideration in the present study. Another reason for their exclusion is that those studies are focused on particular texts or particular authors, unlike a corpus study which aims at uncovering the differences between genres at large.

2.3.4.1. Lexical Variation

Lexical variation analysis methods have extensively been used in task complexity studies in SLA, in particular. Johnson (2017) carried out a meta-analysis of such studies and notes that "studies examining lexical diversity have indicated a positive relationship between lexical diversity – regardless of how it is calculated […] – and L2 writing performance" (p. 16). He also notes that the majority of the studies he has analyzed employ lexical diversity methods whereas only a few of them include lexical density. This finding is confirmed by the author of the present study according to the papers that have been reviewed. Some studies focusing on task-based methods from a lexical viewpoint is given below.

Yi (2000) compared the lexical richness of student essays written in four different tasks (narrative, informative, persuasive and expressive) in two different dimensions: lexical diversity and lexical density. Lexical diversity was found to be more a more distinctive feature than lexical density. Kormos and Dörnyei (2004) studied the effect of motivational factors on second language task performance. One of their methods was lexical richness, operationalized by Uber index due to the fact that it is not affected by text length. They found negative correlations between lexical richness and task attitudes and L2 use anxiety. They interpret the finding as motivated students being willing to talk but not having enough capacity to use different words and less anxious students being likely to use richer vocabulary (p. 9). Kuiken et al. (2005), on the other hand, found no effect of the task complexity on lexical diversity. Gilabert (2007: 59) reports that oral production in planned conditions are lexically richer than in unplanned ones as measured by Guiraud's index. Ong and Zhang (2010) found that complex tasks yielded lower lexical complexity in their analysis, which was carried out with a different measurement method – V2/N. They claim that this method takes into account the text length factor yet they do not explain how it does that and where else it has been used. Kuiken and Vedder (2008) report that measures of TTR and V/√2N (Wolfe-Quintero et al. 1998 qtd. in Kuiken and Vedder 2008: 53) showed no significant difference between the texts produced in complex and less complex tasks. In terms of pre-task planning, Johnson's (2011) study showed that it does not have any effect on lexical variation as measured by MTLD. Meylan and Gahl's (2014) study report that older speakers use more diverse vocabulary than younger speakers as measured by the Uber index (p. 1009). Frear and Bitchener (2015) also compared essays produced for tasks of varying complexity using MSTTR and found that the lexical variation values were different although not statistically significant. Abrams and Byrd (2016) also used MSTTR to test the effect of pre-task activities on writing and found that they result in higher levels of lexical diversity.

2.3.4.2. Lexical Sophistication

Frequency results depend on text, styles, authors and other properties (Hlaváčová 2006: 373). For example, spoken language requires less vocabulary than written language (Nation 2001: 202). Differences can be observed even across various works of the same author. Simonton (1990: 262) analyzed 154 sonnets written by Shakespeare and concluded that "the best" sonnets contained fewer rare words compared to "the worst" ones, indicating a relationship between the popularity of the sonnets and the use of frequent words. In a similar study, Forsyth (2000: 57-8) compared popular poems with "obscure" ones and found that the latter contained more common (more frequent) words that the former.

There are even interesting findings such as that the word *red* is four times more frequent than *brown* or *pink* in the in the spoken British English segment of the Cambridge English Corpus (CEC) or that only four days of the week are listed among the most frequent 1,000 words in in the North American English conversational segment of the CEC (McCarthy and McCarten 2012: 228).

Criado and Sanchez's (2012: 87-8) comparison of two ELT textbooks revealed that both books have a similar score (ca. 91%) of the most frequent two thousand words. Meara (1993) analyzed the frequency levels of radio transcripts, series, comic strip books and songs taken from BBC. He analyzed the corpus through a computer program that runs on Nation's (1986) frequency list (qtd. in Meara 1993: 3) which Meara (1993) developed with his students. His song analysis included the album *Into the Light* by Chris de Burgh and the results indicated more than 90% coverage of high frequency vocabulary (the most frequent 1,500 words in Nation's [1986] frequency list [qtd. in Meara 1993]). Meara's (1993) analyses placed song lyrics at the highest level of frequency and comic strips at the lowest. It must be noted that Meara (1993: 3) chose word types as the unit of analysis.

LFP has been used in studies focusing on comprehension coverage levels in L2. As Laufer (1989: 319) argues, knowledge of 95% of the words is necessary for proper comprehension of a text. Hsueh-Chao and Nation (2000: 422) state that the coverage should be 98% for proper comprehension even though 90% and 95% could work for some learners. Webb and Rodgers (2009a) compared movies of different genres to detect their vocabulary demand in order to reach 95% and 98% of comprehension. They found that "knowledge of the most frequent 3,000 word families plus proper nouns and marginal words provided 95.76% coverage" (p. 407). The demand went up to 6,000 word families in order to reach 98.15% coverage. Horror movies were found to be the least demanding movie genre among the 11 genres by reaching 96.37% and 98.17% coverage with knowledge of 3,000 and 5,000 word families respectively. They also looked at the difference between British and American movies and found that there was little difference at 3,000 word-level, which resulted in 95.35% and 95.76% coverage respectively. However, the coverage of 98% was attained by 6,000 word families for American movies and 7,000 for British. They comment on this finding by referring to the

BNC asserting that the result might stem from the fact that the BNC is more representative of the language in British movies. An interesting finding from their study is that children's movies contained higher low frequency vocabulary, surpassing genres such as comedy, crime and romance. They attribute to this finding to the fact that words such as *penguin*, *zebra*, *squirrel*, etc. are well-known by children although they are in the K6 band.

In their follow-up study, Webb and Rodgers (2009b) carried out the same analyses on British and American TV programs. The findings indicated that children's programs were the least demanding (as opposed to children's movies [cf. Webb and Rodgers 2009a]) ones whereas news and science fiction were the most. They re-confirmed their finding that British TV was less demanding than American due to the calculation based on the BNC as explained above. Van Zeeland and Schmitt (2013) tested the coverage in listening comprehension and found that 98% coverage is necessary if only high comprehension is necessary and that 95% and 90% can also be sufficient, the former apparently being the optimum one (p. 474).

LFP has not been without any objections, though. Meara (2005), in particular, asserts that LFP could be simulated to achieve similar texts to what the subjects produce. His simulator produces texts of random words at desired frequencies and sizes and Meara (2005) remarks that it is a convenient way of testing the reliability of the LFP. Among his long list of items of criticism, he states that texts actually do not differ as much as they are supposed to in terms of their lexical frequency levels and that percentage of the most frequent one thousand words accumulate around 80%. Another argument put forward by Meara (2005: 40) is that it is very difficult to distinguish texts in similar sizes.

Laufer (2005) responds to Meara's (2005) criticism arguing that simulated data would not reveal real scores and claims that this method is a "convenient escape from the real world" (p. 587). The refutation Laufer (2005) provides is the implication used by Meara (2005) himself that the assumption that it is plausible to model his method with further study, which Laufer (2005: 586) comments on as "real data collected from real learners". Laufer (2005: 583) also states that Meara's (2005) argument being based on vocabulary size is wrong in the first place as the former's study looked at vocabulary "use" – not "size". Different aspects of vocabulary proficiency develop in different ways and the fact that a person being possessive of the knowledge of a word does not necessarily mean that s/he is supposed to choose to use it (Laufer 2005: 583). This is also the response to the argument of little difference across text of similar vocabulary size.

Brezina and Gablasova (2013) used *Vocabprofile* in their research where they compiled a new word list called new-GSL to compare their list with the GSL and the AWL. *Vocabprofile* has been used in many studies on lexical sophistication and received positive feedback most of the time. Cobb and Horst (2004) tested seven 2,000+ word segments from the Brown corpus in disciplines of linguistics, sociology, history, social psychology, development, medicine (anatomy) and zoology and 17 non-academic articles using *Vocabprofile* to check the coverage of the AWL. They found that the approximate distribution pattern of average values is 73-5-11-11 percent (K1-K2-AWL-Off-List) for academic texts, and 75-6-6-13 percent for non-academic ones (pp. 22-3). Their findings support that there is difference in the percentage of the AWL words between academic (11%) and non-academic texts (6%) by half and they conclude that the texts have distinct lexical properties as specified by *Vocabprofile*.

Morris and Cobb (2004) used *Vocabprofile* to assess the academic performance of TESL students after they collected essays from 300 students and entered them into computer. Their study showed that there were differences between NS and NNS performances. They maintain that vocabulary profiles are easy to use, cost-effective as well as they are able to sort NS and NNS performances into different levels. However, it must be noted that using vocabulary profilers as an addition to other methods might help obtaining better results (Morris and Cobb 2004: 84-5).

Kirkness and Neill (2009) used *Vocabprofile* and interviews to check the difference in frequency and comprehension levels between different types of texts. They compared a textbook chapter and a journal article to find out the difference in language demands between the two texts. They also asked a group of students to read the texts to check the levels of comprehension. It was found that the comprehension and frequency results were parallel to each other. The textbook contained fewer academic words and higher K1+K2 words. In terms of comprehension, it took less time for the students to read the textbook than the academic text. The textbook was regarded as easier to read due to the lower content of academic words and the clarification it provided for the new vocabulary. The results of their analysis are listed in Table 15:

Table 15: The results of Kirkness and Neill's (2009: 10) comparison of a journal article to a textbook chapter using *Vocabprofile*.

	Journal Article	**Textbook Chapter**
First 1000 words	65.82%	70.90%
Second 1000 words	9.93%	9.36%
Academic word list	12.31%	6.64%
Off-List words	11.94%	13.03%
Words in text	7,018	12,971

It can be noted that the approximate band distribution of the academic text is 66-10-12-12 percent whereas that of the textbook chapter is 71-9-7-13. Kirkness and Neill (2009: 14) conclude their paper stating that teachers can assess the linguistic demands of the texts and make better decisions as to which texts to use in classroom based on the levels of the students.

The software has been used in the analysis of song lyrics as well. Öztürk (2017) compared 177 songs from fourteen albums by four different artists to find out about their vocabulary loads (see corpus details at Lyrics Corpora and Studies). She selected the artists *Adele*, *Bon Jovi*, *One Direction* and *Taylor Swift* and used *Vocabprofile* to detect the frequency levels of the song lyrics based on the BNC and the COCA (Davies 2008). The results showed a very high level of low frequency words (K1+K2) ranging from 96.23% (*Bon Jovi*) to 98.34% (*One Direction*) (pp. 61-2).

Although the software is designed for the English language, a French version is also available. Sundberg (2015) conducted a similar study to Öztürk (2017) in French using *Bande à Part,* a web-based music tool which is based on the frequency levels of song lyrics. The software and the songs it contains are in French and the frequency levels of the songs were found using the online tool, lextutor.ca, French v.5 1-25k corpus – the French version of *Vocabprofile*. Sundberg (2015: 44) remarks that it is possible to determine the vocabulary content of the songs based on their frequencies through *Vocabprofile* tool on lextutor.ca. He chose 23 French songs and graded them into three groups as “easy”, “medium” and “hard”, where easy meant 90-94% of words in the K1 band, medium 85-89% and hard under 85%. The songs in the corpus were then ranked based on their difficulty levels from “easy” to “hard”. The frequencies were calculated based on lemmas since the French version of *Vocabprofile* runs at the said level, and, as regards the tokens, the total K1+K2 value for the songs was 92.14% (pp. 36-7).

2.3.4.3. Lexical Density

Various studies have been conducted on lexical density most of which feature Ure’s (1971) method. It has been used to compare different genres to each other. Fiction texts, for example, have been found to have a lexical density level of 40-54% while non-fiction has 40-65% (Stubbs 1996 qtd. in Camiciottoli 2007: 73). Stubb’s (1986 qtd. in O’Loughlin 1995: 222-3) study reports lexical density levels of 44% for business calls, 45% for calls between friends, 54% for a radio commentary on cricket and 56% for a radio commentary on a state funeral. Zora and Johns-Lewis (1989) compared interviews with conversations between undergraduate and graduate students and found that the overall lexical density level for interviews was 48% whereas for conversations it was 46.96%. The highest level was achieved in the conversation of graduate students with a value of 49.8%. They considered phrasal verbs as one lexical and one grammatical word. Despite the fact that the sample size was rather small and therefore the results were not found to be statistically significant, they have the potential to address to a difference in a more thorough methodology. It needs to be noted that both Stubb’s (1986) and Zora and Johns-Lewis’s (1989) studies were carried out on educated native speakers of English (O’Loughlin 1995: 223).

Ishikawa (2007) found that the task complexity had no significant effect on the lexical density levels of the produced texts. Camiciottoli (2007) compared the lexical density levels of university lectures as they show characteristics of both written and spoken language and found that lectures on business studies are lexically denser than those on multidisciplinary studies (~44% and ~35% respectively). She concludes that the differences are not wide enough to infer any effect of mode or field on the lexical density levels (p. 77). Reid (1990) compared the LD levels of comparison/contrast essays to essays on describing a graph and found that the former type displayed more use of content words, hence a higher level of LD. Hasan (1988) analyzed the lexical density levels of spoken texts based on T-Units as his focus was on syntactic analysis of the complexity of utterances. His results range between 37.97% (informal classroom discussion) and 47.02% (formal interview) for native speakers (p. 127). Nguyen and Nguyen (2016) analyzed the essays of Vietnamese mathematics freshmen students with "Lexical Complexity Analyzer" developed by Ai and Lu (2010) and found that the lexical density levels varied between 52.9% and 68.5% (p. 11). Vidaković and Barker's (2009) study found no correlation between the proficiency levels and lexical density.

To et al. (2013) compared English text books at four levels (elementary, pre-intermediate, intermediate and upper-intermediate) in terms of lexical density in an attempt to find out whether the lexical density level increased with the proficiency level. Interestingly, the lexical density levels for elementary and pre-intermediate books were relatively very high (53.2% and 53.8% respectively) whereas intermediate and upper-intermediate books scored lower (46.3% and 45.5% respectively). These scores were obtained from the calculation based on Ure's (1971) formula. They also tested the same texts with Halliday's (1989) clausal approach and the results were found to be highly correlated (p. 67). The results indicate that the lexical density level does not necessarily increase parallel to the text difficulty. Nevertheless, they only analyzed 160-173 words from each book and the total size of the corpus was 660 words. Their sample size was apparently too small for such a generalization.

Similarly, Hanafiah and Yusuf (2016) tested the LD levels of undergraduate thesis abstracts in an attempt to find out whether they could be categorized into written or spoken language. They found that the LD level of the abstracts was 57%; hence, they concluded that the abstracts could be classified into written language. It is difficult to take their study as a reliable base, though, since a) thesis abstracts are inevitably written and b) their corpus consisted of mere seven abstracts (1,797 words in total). However, the results could be indicative of a tendency of written language to feature an LD level of 40% and above as proposed by Ure (1971). Engber's (1995: 148) study showed that LD has little relationship to writing quality.

2.3.4.4. Comparison

Various studies have been conducted in order to distinguish between different proficiency levels or L1 and L2 speakers using lexical richness dimensions and methods. Lexical variation, in particular, is investigated in a comparative fashion (Koizumi and In'nami 2012: 555) as there is no single ideal measure (Malvern and Richards 2012: 4). Van Hout and Vermeer (2007), for example, tested the vocabulary levels of Dutch L1 and L2 speakers using TTR, Guiraud and *Vocd* and found that Guiraud was the measure that performed the best and that *Vocd* could not differentiate between the two groups. Tidball and Treffers-Daller (2007), similarly, compared the lexical variation performances of three groups who have different proficiency levels in French using *D*, Guiraud, Guiraud Advanced and LRD (a method based on the ratio of noun to verbs). The results showed that all these measures were able to differentiate between the groups except for LRD.

Hyltenstam (1988) compared the lexical traits of native and near-native speakers of English using a combination of lexical density (content words/all tokens), lexical variation (TTR and Hultman and Westman's [1977] formula) and lexical sophistication. The measurement method for LS was based on a Swedish word list and the threshold was set as beyond the most frequent 7,000 words. He also applied a quantitative method, lexical errors; however, it is beyond the scope of the present paper. The findings for the LD were comparable to Ure (1971) in that the writing scores were above 40% whereas those of speaking were below that value. He reports that the three variables fail to yield significant results, though. He also mentions that high values obtained in the measures may be misleading. For example, a high LV value may be attributed to a lack of cohesion in a text; a telegraphic style may yield a high level of LD due to less use of pronouns, or a high LS value may be the result of idiosyncratic use (pp. 74-5).

Daller and Xue (2007) used two sets of lexical richness measurement methods as word-list-free and word-list-based to compare the lexical richness of Chinese EFL students studying in China and the UK. The former set consisted of

TTR, Guiraud and *D* and the latter LFP, P_Lex and Guiraud Advanced. The results indicated that all measures – except TTR – were able to distinguish between both groups significantly. A further effect size analysis also showed that the best methods were Guiraud, *D*, Guiraud Advanced and LFP, in order, for their study, which pictures a domination of word-list-free methods. They interpret the superiority of word-list-free methods as being applicable to different contacts since they are not dependent on word lists.

On the other hand, Daller and Phelan (2007) assessed students' essays based on *D*, Advanced Types, P_Lex, Guiraud, Guiraud Advanced and TTR and correlated the results with teacher's ratings of the essays. Their findings suggested that word-list-based (lexical sophistication) measures, as well as Guiraud, were found to be highly correlated to the ratings. McCarthy and Jarvis (2007) compared 14 different LV measures to test the efficacy of *Vocd* and found that K is best measure of LV followed by *Vocd* in terms of stability across the breadth of text length ranges. They report that *Vocd* offers good stability; however, it is not purely immune to text length. It is important to note that the most ideal text length ranges for the optimum stability in results are different for each method. The ideal ranges for the best performing five measures in McCarthy and Jarvis (2007) are shown in Table 16:

Table 16: Best ranges for the OG category provided from the results of a Bonferroni test on the five best performing LD measures (Adapted from McCarthy and Jarvis [2007: 482])

	Range 1	**Range 2**	**Range 3**	**Range 4**
K	100–500	154–666	250–1,000	400–2,000
D (*Vocd*)	100–400	200–500	250–666	400–1,000
U	154-250	200–500	254–1,000	286–2,000
Maas	100–154	154–333	200–666	250–2,000
D (orig.)	100–200	154–286	200–233	250–400

McCarthy and Jarvis (2010) compared MTLD, *Vocd*, HD-D, K, and Maas and found that MTLD has no correlation to text length unlike the others. They also point out that *Vocd* and HD-D measure the same trait despite some fluctuations in *Vocd* as a result of random sampling (p. 390). They suggest the researchers to use MTLD, *Vocd* (or HD-D) and Maas in their studies. In another similar study, Koizumi and In'nami (2012) compared TTR, Guiraud, Maas, MTLD, *D*, and HD-D using texts differing in size of 50 to 200 words. The results showed that MTLD was the measure to be least affected by text length. Their most surprising finding was that Guiraud was affected more than TTR by the text length. In the wake of their study, Fergadiotis et al. (2015) compared MTLD, MATTR, *D* and Maas and concluded that the former two are stronger predictors of lexical variation.

Lu et al. (2014) used LEX, CTTR and *D* and reported that the books published in 2000s had the highest lexical richness values in all indices in a 100-year period. Verspoor et al. (2012: 252) analyzed Dutch EFL learners' essays and reported that mean word length and CLFP were not good discriminators of proficiency levels but Guiraud was. The CLFP was a method they devised as an alternative to LFP, which calculates the relative frequency of the words in the corpus instead of benchmarking against an external frequency list. In a way, this could be regarded as an intrinsic measure of lexical sophistication to a minimal extent.

Some studies have addressed the differences between different tasks or genres through the use of lexical richness methods. Gregori-Signes and Clavel-Arroitia (2015) assessed university students' written essays. They found that their combination of LD and LFP works stable in the assessment of lexical richness (p. 555). Olinghouse and Wilson (2013) compared the lexical richness of narrative, persuasive and informative essays using MTLD for lexical diversity and LFP for lexical sophistication (with the GSL) and lexical density. The narrative and persuasive essays had higher levels of lexical diversity than informative; yet, the informative texts had lexical density levels approximately three times higher than the narratives and persuasives. Yang's (2014) study resembles the present one in that it employs all three dimensions of lexical richness. She compared rhetorical tasks with topic-familiarity tasks in an L2 setting in terms of lexical diversity, sophistication and density. She measured these aspects using *Vocd*, first 2,000 words in ANC, and the ratio of lexical words to the total number of tokens in the text, respectively. Rhetorical tasks were found to affect lexical density but not diversity or sophistication. Yoon and Polio (2017) compared narrative and argumentative essays produced by EFL students both against each other and also against essays written by native speakers. They used average word length, word frequency and *Vocd* as measurement methods and found that argumentative essays were more lexically sophisticated than narratives; yet, the narratives had higher lexical variation values. In terms of NS, they produced more lexically diverse language in argumentative essays than narratives (p. 288). However, one of the methods they used to test lexical sophistication was word length, which is not commonly used for this purpose.

Perhaps, the most similar study to the present one in terms of lexical analysis coverage is the one carried out by Šišková (2012) where she used measurement methods of lexical variation, lexical sophistication and lexical density to compare narratives by Czech EFL students varying between text lengths of 200 and 500 words. She used types and tokens as the counting unit as "English language contains only a minimum of inflections compared to other languages" (p. 27). She measured lexical variation using, TTR, Guiraud, Herdan's Index, Uber Index, Mass, *Vocd* and HD-D; lexical sophistication using advanced types: >K1 (GSL), >K2 (GSL), > K1 (BNC), > K2 (BNC), Guiraud Advanced; and lexical density as content words/total words. The results suggested that the correlations within the groups (between the measurement methods of the same aspect) are higher than those across the groups, indicating that the methods measure similar constructs (p. 33). She has also found correlation between lexical diversity and sophistication methods, if not strong, between Guiraud and >K1 (both GSL and BNC), in particular. Lexical density measure was found to correlate weakly with both lexical sophistication and lexical diversity; however, the correlation with diversity was significant.

In another similar study, Zheng (2016) used B2K; Uber index and content word ratio to tap into all three sub-aspects of lexical richness. The study, however, differs from others in that it is a longitudinal one, which aims at assessing EFL learners' lexicon throughout an academic year. She found that lexical variation and sophistication tended to increase over time while lexical density remained mostly stable. She comments on the correlations between B2,000 and CW and U and CW as "strong and significant at the beginning, moderate but non-significant in the middle, and strong and significant in the end [of the academic year]" (p. 47).

Torruella and Capsada (2013) compared seven different lexical diversity measures (TTR, RTTR, CTTR, Maas, MSTTR, MTLD and HD-D) and their finding was that different indices produced similar results across texts. TTR, RTTR and CTTR have been found to be highly text-length dependent. In terms of the other measures, Maas was the most stable one and MTLD the least. Their analyses were carried out on texts from different genres and the measures they used were capable of identifying the differences. For example, poetry was the lexically richest genre while academic texts were the least rich. However, it must be noted that the analyses were implemented to test lexical variation and therefore cannot be generalized in terms of lexical richness as it does not employ lexical sophistication and density measures. An analysis of lexical sophistication is likely to increase the lexical richness of academic texts.

In an attempt to compare the existing lexical sophistication measurement methods and introduce a new one – S – Kojima and Yamashita (2014) analyzed B2K, Advanced Guiraud, P_Lex and S to investigate the extent to which they are dependent on text length. They found that S is the least text length dependent measure, which is capable of producing results at a reliability level of .70 with token counts as low as 210, whereas P_Lex required 260, Advanced Guiraud 520 and LFP 1,060. As interesting as the results might be, it is still not a widely adopted approach and the possibility of a comparative study based on existing results in the literature seem quite low.

2.3.5. The Approach in the Current Study

Following the literature review on lexical richness, its levels and measurement methods, a division and classification is proposed, mostly based on Read's (2000) approach. According to this classification, which is illustrated in Figure 3, lexical richness could be identified mainly in four dimensions: lexical sophistication (LS), lexical variation (LV), lexical density (LD) and lexical originality (LO). Lexical rareness/rarity has been subsumed under LS, lexical diversity/variety under LV, and lexical individuality under LO.

In terms of measurement methods, LS is divided into two categories as *intrinsic* and *extrinsic* measures. Intrinsic measures consist of hapax legomena/dislegomena and mean word length. Extrinsic measures are the ratio of advanced words to total number of words, LFP, P_Lex, S, LEX, Advanced TTR and Guiraud Advanced. LV measures are classified based on Torruella and Capsada's (2013: 448-9) division. The indices are grouped as TTR-related, logarithmic and complex ones for each of which some examples are given. LD can be calculated in at least five methods: 1) by dividing the number of content words to the total number of words, 2) by dividing the number of content words to the number of function words, 3) weighted approach, 4) ratio of nouns/verbs/adjectives to the total number of words and 5) the ratio of content words to the number of clauses. LO will not be taken into consideration in the present study as it is dependent on the number of words used by one writer. This approach may not be applicable in a corpus analysis, which aims to distinguish between genres – not individual producers of texts. It needs to be noted that this classification is not fully definitive. There are (and may be) various other ways of classification and measurement. The present one has been drawn up based on the most commonly used ones in the literature.

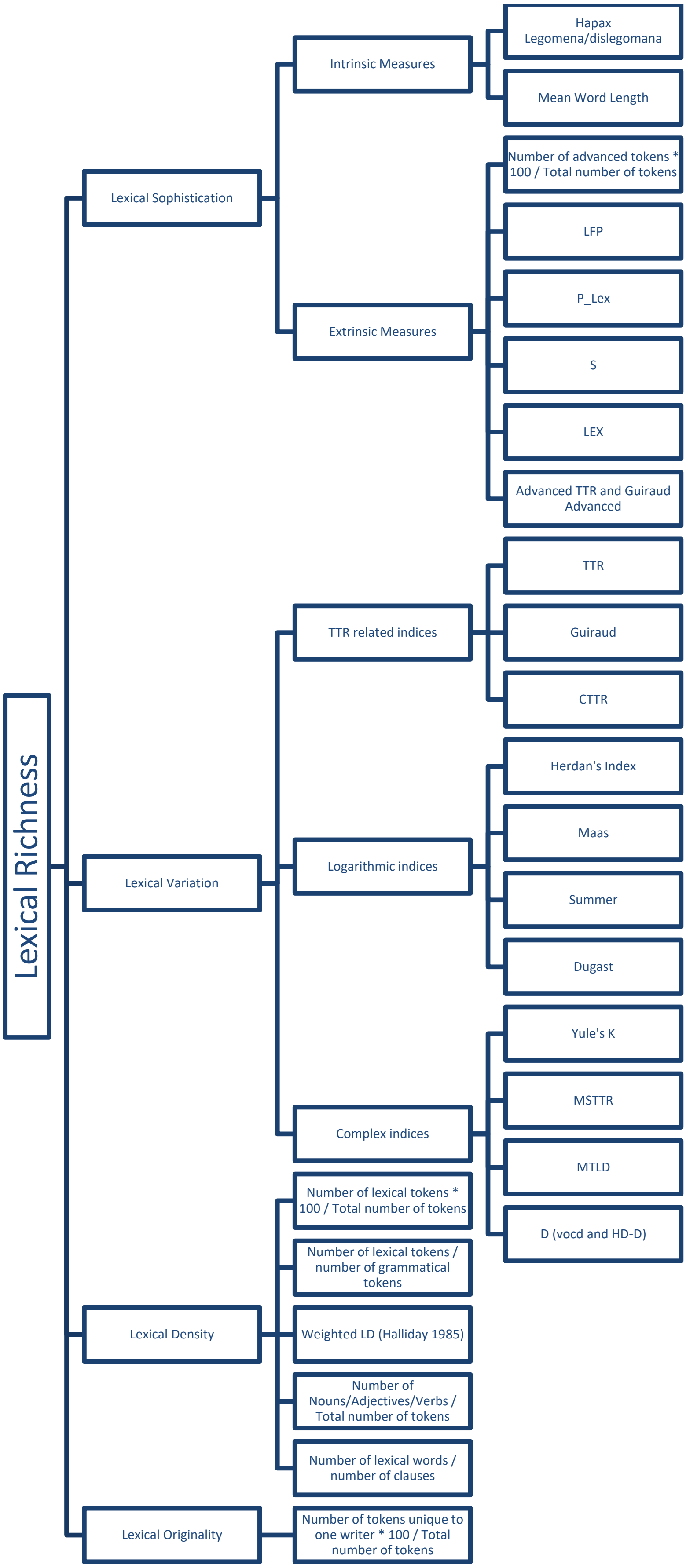

Figure 3: Levels and measurement methods of lexical richness

The following example by Daller et al. (2007: 13) gives an overview of different results that could be reached from different viewpoints.

The cat sat on the mat.

The lexical richness values for the sentence are 0.833 (TTR), 1 (LD) and 0.167 (P_Lex). It is evident that all dimensions tap into different aspects of vocabulary and one does not (or cannot) substitute the other. One measure is not better than the other, either, as each one provides information on a specific dimension (Tidball and Treffers-Daller 2007: 134). Therefore, it makes sense to incorporate these dimensions through the use of measurement methods from each dimension to complement each other. Richards and Malvern (2007: 84) advocate a multiple measure approach and consider the methods "interrelated but complementary sources of information".

The detailed account of lexical richness has shown that the studies, more often than not, focus on L2 or its comparison to L1, yet hardly L1 alone. L1 lexical richness studies are mostly confined to authorship attribution and task-based studies. The present paper does not aim to have any concern with authorship attribution or task performances; it solely aims to pinpoint the lexical differences between distinct genres. In this respect, it could be argued that it is mostly in the same vein as the comparison studies mentioned above where the already established different groups are further compared based on their lexical traits. However, the divergence point is that, contrary to the above-mentioned approach (or studies), the differences between the groups do not (or may not – as it will be the exact aspect to be discovered) lie in their linguistic properties but in musical ones. Hence, the measurement methods are not expected to give ideal results that will prove the existing distinction, but to indicate the differences (if any) between the subgenres from the viewpoint of lexis. Although there is a continuum of musical heaviness (and/or sophistication) among the subgenres, the results might not fall into the same continuum to show a parallel lexical richness level – or might even show an inverse relationship.

Another point is that the results that will be obtained from the present study probably will not reflect the true lexical competency of the artists. Daller and Xue (2007: 150) note that,

> A person’s language proficiency is closely related to the size and depth of their vocabulary, and this is true of both first and foreign languages. The lexical richness displayed in an oral or written text is a result of this underlying vocabulary knowledge.

Within the framework of the current study, such an inference cannot be made as the reason for the difference between the subgenres. A heavy metal band can have good knowledge of vocabulary but prefer to write according to the codes of their subgenre (see Swales 1990, Laufer 2005). Hence, the lexical results of the genres will be of generic nature rather than academic.

2.4. Keyness

Keyness is what a text is about in terms of lexis. It is calculated based on how frequently words occur compared to a reference corpus (Baker 2012: 107). These words are called keywords. “Keywords are not necessarily the most frequent words in [a] corpus; rather, they are the most unusually frequent words” (O’Keeffe 2012: 119). They are the “statistically significant word[s] characterizing a document, text, or corpus” (Rayson 2012: 1).

Keyword analysis, in the statistical sense, was introduced by Scott (1999) with the *WordSmith* software he developed (Culpeper 2014: 12). Keywords give insights about genres and styles in a non-biased way (Culpeper 2014 and Groom 2010). As Culpeper (2014: 10) puts it, “style […] is a matter of ‘frequencies’, ‘probabilities’, and ‘norms’”. Therefore, keywords can provide information on stylistic orientation of a text free from human factor as it draws “attention to otherwise difficult to detect stylistic moves of the author” (Gerbig 2010: 150).

“The value of the statistic or ‘keyness’ is proportional to the difference in relative frequencies. In other words, the larger the difference in relative frequencies, the larger the value of the statistic or ‘keyness’” (Rayson 2012: 2). Keyness differs from raw frequency counts in that it is a relative phenomenon depending on the reference corpus used. Different results could be obtained as a result of implementation of different reference corpora (O’Keeffe 2012: 119). In other words, keyword analyses do not provide absolute results but relative ones (Hunston 2012: 244). It is interesting in this regard that a very low frequency word in a language could be a keyword in a particular text (Baker 2004: 352).

Gabrielatos (2018) is a very comprehensive study, detailing the notion of keyness in many aspects. He notes that although comparison of raw corpora to each other may have its deficiencies such as multi-word units, polysemy, part of speech, it is still useful (p. 2). Keyness analyses mainly aim at finding out the differences between texts but recently similarity has become a focal point as well. Also, keyword analyses could be carried out in two angles to list positive and negative key words. Positive keywords are those which are "overused" in the text in question as compared to a reference corpus whereas negative ones are those which are "underused". (Rayson 2012: 2).

> [A] keyness analysis is essentially a comparison of frequencies. As it is currently practised, it usually aims to identify large differences between the frequency of word-forms in two corpora (usually referred to as the *study* and *reference* corpus) (Gabrielatos 2018: 3).

The term *reference corpus* means a large corpus of the language in question – usually a general corpus – capable of reflecting various aspects and genres of that language. However, the size of the reference corpus is a controversial issue and Culpeper (2014: 14) claims that "there is no magic formula" for it. Scott and Tribble (2006: 64-5) mention Berber's (2004: 101-3) proposal that the reference corpus should be approximately five times larger than the specialized corpus but underline at the same time that it is hard to pin down a specific rule about it. Xiao and McEnery (2005: 70-1) compared the BNC with FLOB and found that they both yielded similar keywords. Milizia (2010) has a similar finding in her study where she compared Tony Blair's speeches against George Bush's. She compared a 1 million word corpus of Blair's speech against a 10 million word corpus of Bush and a 2.5 million one. The results were more or less similar. If the reference corpus is similar to the study corpus in content, the key word will be quite few and concentrated. On the other hand, a comparison with a totally different one will yield very different and diverse key words. Finally, if the study corpus is compared to a general reference corpus, such as the BNC, etc., there will still be many keywords but they will not be as diverse as the previous option (O'Keeffe 2012: 121). Baker (2006: 37) highlights two advantages of using a large reference corpus as 1) they can be used to undercover particular features of the languages since they are representative enough and 2) they can serve as a benchmark against our own data to decide on what is normal in a language and what is not.

According to Scott (2010: 51), even an “absurd” reference corpus will provide similar key words, which indicates that the system itself is robust. The term “absurd” here refers to the domain differences between the study and reference corpora, such as measuring the keyness of a nanotechnology corpus with a reference corpus of 19th century poems. Gabrielatos (2018: 29) takes a more flexible approach and claim that the reference corpus neither has to be a general one and larger than the study corpus as long as they are able to handle the research questions and/or hypotheses. He also finds assigning the “reference” and “study” labels to corpora misleading. Scott (2009) carried out a very interesting study, “in search for a bad reference corpus”, and compared his corpus consisting of commerce books and transcriptions of doctor-patient dialogues to 22 different reference corpora taken from the BNC varying between 10 and 4,000 texts. The generated keywords were very similar. To take a more radical approach, he compared his corpus to a totally irrelevant reference corpus which was a collection of Shakespeare’s plays. The result was that when the Shakespeare’s plays were used as a reference corpus more keywords were detected but they were not irrelevant or – as Scott (2009: 89) puts it – “absurd”. In contrast to O’Keeffe’s (2012: 119) claim that employing different corpora yields different results, Scott concludes the study stating that there is no bad reference corpus and the keyword measure is robust given that it has stable results over extreme variations in size and content.

Keyness is calculated through the use of lexical software which compares the words in a text with a reference corpus (O’Keeffe 2012: 119). *Antconc* (Anthony 2018) is one of the freely available and widely used tools for keyness. The tools produce a list of words which are ranked based on their levels of “aboutness” in the text in comparison to a reference corpus. Although it is based on repetition it does not mean all kinds of repetition but those which statistically stand out with reference to another corpus. Thus, it “is a matter of being statistically unusual relative to some norm” (Culpeper 2014: 12). Scott and Tribble (2006) expand this by asserting that the word “elephant” could be found as key in a text on Serengeti; however, this is not the case when the English language is taken as a whole (p. 56).

Gabrielatos (2018) primarily deals with the statistics in keyness analyses. There are two common ways of assessing the results for significance: Log-likelihood and Chi-squared tests. Although the former is reported to be more reliable (Rayson 2012), Gabrielatos tested both measures and found that the ideal measure is to apply an effect-size metric which is backed up by statistical analysis. He further notes that the BIC score ("the degree of evidence against the null hypothesis" [Wilson 2013: 5-6 qtd. in Gabrielatos 2018: 18]) could be the most ideal way of assessing keyness values; however, this metric is not included in any keyness measurement tool as of 2018 (p. 29).

Keyword lists provide a clear distinction of lexis between the genres (Koester 2012: 50). McCarthy and McCarten (2012: 228) remark keyword analysis can show a good picture of a particular data set. As Philip (2010: 185) notes, "a discourse cannot hinge upon hapax legomena. For this reason, the calculation of key words is dependent on frequency measures and repetition". There may be words in text which occur only once (hapax legomena) which amount to approximately 40% of a frequency list based on the BNC. Scott and Tribble (2006: 26) describe them as mostly proper nouns, typos, non-English words, etc. They also mention that a usual threshold level of at least two or three occurrences of a word is taken into consideration for that word to be deemed key (p. 59). There are some limitations of the keyness analysis as well. For example, the results are affected by the reference corpus "in ways which are still not fully understood" and a definitive set of keywords cannot be established as they change depending on the reference corpus and cut-off points (Scott 2010: 52). These issues naturally hinder full comparability. Comparability across studies could be maintained only if the same reference corpus and same statistical measures are used, though (Culpeper 2014).

Keyness analyses could be carried out in two ways: focused and exploratory (Gabrielatos 2018: 2-3) where focused means targeting specific linguistic features before the analysis and exploratory seeks to view a broader picture of the results without any hypotheses. These approaches may be combined depending on the purposes of the study. Keyword analyses could be used to compare styles and genres (Baker 2006: 146). It could be viewed as a more practical and fully lexis-oriented version of Biber's MDA, which looks at various dimensions of language. Both methods are in fact about frequency of occurrences of specific items in an attempt to differentiate genres/registers/styles. Following Tribble's (1999 qtd. in Xiao and McEnery 2005) claim that keyword analysis could prove as effective as MDA, Xiao and McEnery (2005), in fact, compared MDA to keyness analysis. Their motivation was to find a "low effort" alternative to the MDA. Interestingly, they found similar results with the MDA. This obviously does not mean that keyword analysis can substitute MDA as the authors acknowledge as well; however, the former could be more practical and comparable than the latter.

Scott and Tribble (2006) studied the keyness of Shakespeare's Romeo and Juliet. They first compared Romeo and Juliet to Shakespeare's all other plays, and then to a) the tragedies alone, b) the Complete Works including poetry, c) the BNC. The results suggested that the KWs are mostly similar (p. 64). Baker (2012) compared a corpus of 355 newspaper articles containing a total of around 250,000 words to BE06 Corpus, which is a general corpus of English featuring one million words (p. 107). His aim was to study the articles, which were about metrosexuality, and therefore, his selection of newspaper articles was based on the results of the keyword search for "metrosexual". He, then, was able to group his keywords in categories such as, politics, sports, fashion, etc. (pp. 107-8).

León (2015) carried out a keyness analysis of maritime logistics texts on the subcorpora of LogisTRANS Corpus (Losey 2015 qtd. in León 2015). In this methodology, both corpora were the subcorpora of the macro corpus LogisTRANS. The specialized corpus, GreenLog (GL), contained 537,774 words from 151 text files. The reference corpus, MLogistics (ML), contained 1.5 million words. Both corpora are comprised of texts from a single domain but GL has two different text types as research articles and technical reports. She defines the corpora as "monolingual, untagged, textual, chronological, specialized, closed, pyramidal, simple, modular and stem from the LogisTRANS macrocorpus, which comprises further conceptual subdomains" (León 2015: 527-8). After finding out the keywords in GL, she ran the same procedure separately for research articles and technical reports and found that the results were mostly similar. However, the technical reports contained more specialized vocabulary than research articles (León 2015: 530-1).

In another study, Pojanapunya and Todd (2015) compiled two specialized corpora of research articles on humanities/social sciences and applied sciences. They sought to find out the keywords, which are typical to each corpus and also shared keywords which commonly occur in both corpora by comparing them to the BNC. They found that the differences between the disciplines are greater than similarities. Fraysse-Kim (2010) compared the Korean history books that are being taught in North Korea, South Korea, Korean residents in China and Korean residents in Japan. The results showed that while the most key pronoun was *I* in three Korean groups, *we* was the most key in North Korean books, which indicative of a collectivist and communist stance. It is also interesting that the word *great-marshal* is more key than *I* in North Korean books.

3. METHODOLOGICAL APPROACH

Following the detailed account on genres, metal, corpus linguistics, lexical richness and keyness, the methodological chapter will provide the complete steps taken for the formation of the research methods. It starts with the building of the corpus, including the procedures for the selection of the songs. It then delves into the treatment of song lyrics as a preparatory phase for analysis. Finally, the lexical richness and keyness analyses, which have been carried out on song lyrics, will be expanded on separately.

3.1. The Corpus

Various methods have been applied in establishing song corpora for different kinds of analyses. One common method of choosing songs is referring to well-known song charts. Examples of chart-based corpora have been provided in Table 6. In addition to the studies mentioned earlier, there are others, which are based on corpora built on the researchers' taste, preference and knowledge.

For instance, Walser (1993) selected bands based on their popularity, media coverage and fans' opinions. He states that "I have elected to concentrate primarily on music of the 1980s, since that is the music both my informants and I know best" (p. XIV). His remark gives space for some personal expertise if subjective. Arnett (1996) selected six bands in his study on the basis of being "the most popular" ones (p. 44). His corpus was comprised of three mainstream metal bands (*Judas Priest*, *Iron Maiden* and *Ozzy Osbourne*) and three speed/thrash metal bands (*Metallica*, *Megadeth* and *Slayer*). He reflects on his selection by proposing that the corpus could have been set up in other ways. He justifies his selection by claiming to have chosen what most metalheads are listening to (p. 175). Öztürk (2017: 59), too, compiled her corpus of pop and rock songs according to her personal music taste.

Weinstein (2009: 23) followed a slightly diverse method in selecting the bands and asked for "expert" help in order to avoid "cherry-picking" bands and songs. She asked ten people including a performer, writer, band manager and filmmaker who were all metal fans to name bands that they think fall under a certain genre and conducted her research accordingly. Tsatsishvili (2011) selected seven metal subgenres which he found popular among the metal community and those having “more or less clear distinctive characteristics” (p. 32). As the final example, Sundberg (2015: 34) selected the songs based on their word frequency levels, which is different from the other methods mentioned so far in that it focuses on a linguistic feature of songs rather than musical ones.

It is understood from these examples that various methods could be applied in establishing a corpus. Arnett (1996: 41) acknowledges that there is debate even among the metal fans on which bands qualify as heavy metal and which do not. It is important, though, to draw a line in the study in order to avoid confusion and transgression between the genres and therefore results.

Based on the above-mentioned pretext, this study aims to analyze three metal subgenres (heavy metal, thrash metal and death metal). These subgenres have been selected as they are three main subgenres of metal while at the same time they possess overt differences in terms of content and form. The subgenre selection was based on a claim that ‘the harder the music genre, the lexically richer it is’ as indicated in 1.1. Research Questions. In order to test this claim, a scientific and clear-cut division among the genres was necessary.

The first step towards the analyses was to establish a corpus of songs written by ‘prototypical’ representatives of their respective genres. This was not as easy to a researcher as it is to a metal fan. Any fan could name any band into some genres; however, when it comes to academic context, concrete references are necessary. In this respect, a literature review on metal music was carried out, which not only included books and articles but also documentaries and dedicated websites.

As indicated under the section Classification of Metal Subgenres, the perception of some bands’ genre by the fans and authorities may be a) the same with the bands’ perception, b) totally the opposite c) interwoven to reach a common labeling. Such occurrences make it more difficult for the researcher to carry out analyses under specific subgenres. The way to establish a corpus of such bands is to find the ones that are widely accepted as ‘prototypes’ by the academics and reliable mediators.

There was another key criterion, which could not have been overlooked: band origins. Care has been taken to maintain equal number of songs from British and American bands in all subgenres. This was another major concern in the establishing of the corpus so as to eliminate the L2 factor. The corpus shows similarity in this sense to Werner's (2012) corpus of pop songs, which involves only British and American acts. It could be argued that British English and American English factors might have had an effect on the results and this will be discussed in detail in 6.1. Limitations.

3.1.1. MC, HMSC, TMSC and DMSC

The present study has followed a more comprehensive approach in the establishment of the corpus taking into account various aspects of metal music as described in The Three Metal Subgenres Used in the Current Study. After a comprehensive literature review, which involved not only academic papers but also metal sources, the corpus was formed with 600 songs from 105 albums in heavy metal, thrash metal and death metal subgenres. This resulted in three subcorpora as Heavy Metal Subcorpus (HMSC), Thrash Metal Subcorpus (TMSC) and Death Metal Subcorpus (DMSC), all of which are nested under the Metal Corpus (MC).

Only full-length studio albums by all bands have been selected, which means the exclusion of singles, demos, EPs (Extended Plays, releases which contain at least three but no more than five different songs [About The Awards – RIAA]), live albums, videos, boxed-sets, splits and compilations. The reasons to include only full-lengths were to have,

a) Similar numbers of songs per album: Most full-length albums contain around ten songs whereas EPs and singles contain around three or four.

b) Bands' own compositions: EPs, and live recordings usually contain tracks which are covers of other bands and their live songs.

c) Exact lyrics: Sometimes the lyrics are distorted in live versions. They could be shortened, repeated more, altered or sung in another language depending on the audience.

Only the original pressings of the full-length albums have been taken into consideration. All songs in the albums have been included in the corpus with the following exceptions:

- Bonus tracks: Those included in re-mastered versions or country-specific editions (Japanese, etc.), such as covers, live recordings, remixes, re-mastered versions, demo versions, etc., were removed.
- Covers: Songs originally composed by other artists have been removed from the albums even when they are not bonus tracks.
- Instrumental songs: Songs with no lyrics have been omitted.

The release years of the albums have not been considered so as to maintain a broader reserve of eligible albums. The albums in MC have been released between 1970 and 2019. Only bands which have released a minimum of three full-length albums have been selected. This was done for two reasons: 1) to opt for the bands which have a career that can be measured by their discographies. Having released a number of albums, bands get to be known within the "discourse community" of metal and can be considered to be a part of a particular genre. However, it must be noted that it does not necessarily mean that a band, which have released eight albums, is superior to another band with three releases or that it is a more typical representative of a particular genre. 2) as only songs with 100 to 400 words were used (see MC in numbers) more than one or two albums could have been necessary in order to find enough songs from full-length albums to meet the criterion.

Another point of consideration was to have one band of an artist. That is, the side projects of the members of some bands have not been chosen to maintain ultimate diversity. For example, *Bruce Dickinson*, the singer for *Iron Maiden* (heavy metal), has his own self-titled solo project where he plays heavy metal as well. His albums have not been included in the MC due to the concern for maintaining diversity.

In terms of genre distinction, academic resources and Metal-Archives have been referred to. Bands playing in hybrid genres have been avoided if they combine two of the subgenres in the present study (e.g., thrash/death) but were included if they combine other bordering genres, such as *Virgin Steele* (Hard

Rock/Heavy Metal). For HMSC, the bands listed or mentioned as "Heavy Metal", "NWOBHM" and "Classic Metal" have been included. Bands shifting genres in their careers have been represented with their most typical albums of the related genre, e.g., *Carcass - Necroticism - Descanting the Insalubrious* for death metal.

The songs were coded in a way to reflect their genre, country or origin, artist and album. For example "heavy_uk_06-01-05" stands for the 5th song in the 1st album of the 6th heavy metal band from the UK. The complete list of the songs is available in Appendices D through F.

Care has been taken to choose the bands that have been mentioned in academic publications and metal media. Most bands in the MC have been mentioned in various academic papers and books. Nevertheless, unsurprisingly, not all bands can be featured in academia. This does not mean that they cannot be studied or that it is barrier against building a comprehensive corpus. Hence, some bands in MC have only been mentioned in Metal-Archives.

It is interesting that the amount of references for British bands in heavy metal is far more than American bands whereas it is the opposite in thrash and death metal. The obvious reason for this could be the fact that heavy metal was formed in the UK while thrash and death metal in the USA. The list of the bands, albums and songs in MC can be found in detail in Appendices D to F and complete list of references to each band in the corpus is available in Appendix G. Bands with no references have been indicated as "none" and this corresponds to the fact that no references to them have been found in the literature that has been examined. Naturally, they might have been mentioned in publications, which have not been available to the author, or in those that have been published later than the date of the building of the corpus.

3.1.2. Compiling and Processing the Data

All the lyrics have been downloaded from Metal-Archives; thus, only the albums whose lyrics are featured on the website have been selected. Some songs lyrics have been manually corrected as their lyrics did not specify the repeats clearly. This was done by listening to the respective songs thoroughly and correcting the lyrics where necessary. Out of the six methods in lyrics processing proposed in Lyrics Corpora and Studies, the present study fully employed methods 1 and 3 by removing all markups as they are not part of the original artwork and by correcting the typos. Method 4 was also applied but in a limited fashion. Due to the nature of the song lyrics, not all repeats are exactly repeated even though they are called ‘repeat’. In this regard, the repeats were manually examined and divided into two categories as 1) exact repeats and 2) partial repeats. Only one exact repeat in a song was kept throughout the corpus. Partial repeats, on the other hand, usually display a different word or line within a verse which make them unique; thus, they were not removed. As a result, the processing of the lyrics had three major courses of action: a) correcting the typos, b) removing the exact repeats, and c) retaining the partial repeats. Method 2, removing markups and inserting corresponding text, was not applied as it would affect the results of the frequency analysis. Method 5, removing a part of the main text, was used in a few songs to remove some lines – not words – which were in languages other than English, such as *Mauer*, *Weg*, etc. Method 6, no modification, was not used either as there are numerous typos and markups in song lyrics.

TTR and other lexical diversity measures are highly susceptible to the repeats as they affect the absolute number of tokens and therefore the ratio. For example, one song in HMSC (heavy_uk_01-01-01) has a TTR value of 0.36 with the repeats whereas it rises up to 0.48 without the repeats. Although not as much as TTR, token-based frequency analysis is also affected by the inclusion or exclusion of repeats. The same song, for instance, has GSL B2K token ratio of 15.18% with repeats included. The value changes to 6.38% when the repeats are removed. LD value also changes – although minimally – from 0.39 with repeats to 0.38 without repeats. As seen in the examples, more repeats could cause fluctuations in the values, whether significant or insignificant. The fact that not all songs have the exact number of repeats makes it difficult to ‘equalize’ the texts while the repeats remain. Hence, it seems that the optimum method to achieve a balanced structure of lyrics is to omit the repeats and simplify them to the unique occurrences of verses.

Additionally, the lyrics are printed and/or published in different formats with regard to the repeats. Some sources prefer to write all the repeats the way they are sung while others simply insert metadata such as ‘repeat’, ‘×2’, etc. Some sources even neither indicate the repeats nor write them as many times as they are sung. All facts considered, texts were processed identically in both analyses for the sake of consistency and the repeats were treated as ‘repeats’ immediately at their second occurrence.

Another point of consideration in the removal of the repeats was the way they were written. It was easy to delete a whole verse as long as they were repeated exactly. Such repeats were identified as ‘exact repeats’ and removed completely from the songs. An example could be,

Running silent, Running deep, we are your final prayer,
Warriors in secret sleep, a merchantman’s nightmare,
A silent death lies waiting, for all of you below,
Running silent, Running deep, sink into your final sleep
(*heavy_uk_06-01-07*)

This verse is completely repeated three times throughout the song. At the end of the song, the first two lines of the verse are repeated again and this was disregarded as well since those two lines are the exact repeats from the original verse.

Managing exact repeats is relatively straightforward, but where it gets more complicated are the cases in which only one or two words out of a whole verse are different. Removing the whole lines or verses would have an effect on the frequency results; therefore, such repeats – identified as ‘partial repeats’ – were retained as a whole. The following verse could provide an insight on partial repeats:

Sport the war, war support
The sport is war, total war
When victory’s really <u>*massacre*</u>
The final swing is not a drill
It’s how many people I can kill

Sport the war, war support
The sport is war, total war
When victory’s really <u>*survival*</u>
The final swing is not a drill
It’s how many people I can kill
(*thrash_us_09-01-01*)

The verses provided above follow each other in the song and the mere difference between the two are the two underlined words – *massacre* and *survival*. The parts preceding and following the underlined words were kept for the sake of not distorting the song structure. The repeat part in *heavy_uk_06-01-09* has a partial repeat which is again different from the one in *thrash_us_09-01-01:*

> *Bring your daughter, bring your daughter to the slaughter*
> *Let her go, let her go, let her go*
> *[…]*
> *Bring me your daughter, bring me your daughter, bring me your daughter*
> *To the slaughter*
> *Fetch your daughter, bring me your daughter, fetch your daughter*
> *To the slaughter*
> (*heavy_uk_06-01-09*)

A look at the underlined words shows that the second repeat involves an extra pronoun (*me*) and a synonymous word (*fetch*). Such occurrences also were retained in their originality so as to keep the interferences at a minimum.

Typos in the lyrics are very frequent even at times when they are retrieved from album sleeves or downloaded from the official websites. With this fact in mind, all the lyrics were manually checked for errors through scanning and with the help of Microsoft Office 2010 Word spellcheck feature for US English. Words written in British spelling were retained. Petrie et al. (2008) changed all words in their study to US spelling but such an approach was not employed in the present paper for the sake of fidelity to authenticity. Some words marked by MS Word to be misspelt were kept originally since the artists chose to use those words in that manner. Some examples are: *bloodthirst, arse* [marked as British English], *ballotation, spasmic, metamorphial, devourance, cozing, strickened, wisemen*, and *transmutated*. Proper words, such as *Gein,* were not taken into consideration.

Auxiliary contractions such as *'s, 'll, 've, 'm* were spelled out in full in order for the analysis tools to group them together and treat them as the same word. Colloquial words and non-standard spellings were standardized so that they could be recognized by the analysis tools. Hence, words like *wanna, gonna, ya*, *outta* etc. were standardized as *want to, going to, you*, *out of*, etc. Webb and Rodgers (2009a and 2009b) followed the same course of action and converted the contractions, connected speech and hyphenations into their respective forms in the BNC word frequency list. Occurrences of *g*-dropping such as *goin'* were changed to their full versions (*going*) (see Öztürk 2017). The only exception to this is *ain't*. No change has been made to *ain't* as it may not necessarily correspond to a specific auxiliary. Moreover, it is a different word which can be used as an alternative to other auxiliaries and this makes it a choice rather than non-standard spelling.

Other instances of non-standard spelling (e.g., *fukin'*), which were different from their original forms only in term of orthography, were corrected. Although Kreyer and Mukherjee (2007) retained such spellings, the reason was that their research design made it necessary. They established a link between non-standard spellings and the way they are sung stating that those words were sung with more emphasis. As the present study has no such aim, these spellings were normalized. This would only make a difference in the frequency and keyness analyses; however, a word cannot be deemed low-frequent just because it is written with a /k/ and instead of /c/. Such an approach would place the word *kat* into K12 frequency band whereas *cat* is in K1. As the focus in the present study is on the lyricists' choice of words, orthographical non-standard spellings were reverted to their original forms. On the other hand, some words, which are deliberately formed by the lyricists (e.g., *observaillance*, *suicideology*, *illegitimeat*, *funereality*, etc.), were retained as they are puns formed with compound words which are mostly based on homonymy – not only orthography.

The corpus was untagged as is the case with other similar studies, such as Meara (1993), Falk (2012), León (2015), Öztürk (2017), etc. The analyses in the study do not require a tagged corpus as the methods are type and token related and grammatical aspect is not included in the study. The only exception is the LD analysis, where the identification of content words was necessary. This was automatically carried out using the related software (see *The Analyses*)

3.1.3. MC in numbers

All subcorpora were formed to contain 200 songs each, equally divided between British and American bands. Equality has been maintained in the number of songs per band by allocating 10 songs each. The numbers of albums are 33 for HMSC, 37 for TMSC and 35 for DMSC. The MC contains 600 songs in total, which are highly above the average number of songs (280) in corpus-based lyrics studies. In other words, one subcorpus of the present study almost equals to a full corpus of a similar one. This allows the present corpus an advantage in size.

Following the processing of the lyrics as mentioned above, some songs were found to drop below 100 tokens, which was a critical threshold for some analyses. Petrie et al. (2008) analyzed the complete *The Beatles* discography but removed the songs which contained fewer than 50 words. *Vocd* is reported to run well between 100 and 400 tokens, Yule's K 100 and 500, Maas 100 and 154, HD-D 100-200 (McCarthy and Jarvis 2007), MTLD 100 and 200 (Koizumi and In'nami 2012), LFP 200 and 400 (Laufer 1995), etc. The scales proposed in these findings vary to a considerable extent and finding songs of specific lengths after they are processed is more demanding than, say, getting students write essays of a certain length. Thus, the lower and upper limits have been set as 100 and 400 tokens to encompass all the methods mentioned. It was not plausible to set the lower limit to 200 as it would render the compilation of the corpus almost impossible. Out of 200 songs in all subcorpora, only 46 have 200 or more tokens in HMSC, 56 in TMSC and 43 in DMSC. This shows they only make up approximately a quarter of the MC. As regards the upper limit, although it is set at 400, only one song reaches 397 tokens and the remainder are lower. The above-mentioned band/release/song selection criteria poses a dire strait through which very few of them can be allowed. A total of 600 songs have been compiled and processed in line with the procedures described above and the resulting corpus can be detailed as shown in Table 17 below:

Table 17: Word distribution across the MC.

Subcorpus	**# of songs**	**# of words**	**Average # of words per song**
HMSC	200	33,158	165.8
TMSC	200	34,799	174
DMSC	200	33,859	169.3
TOTAL (MC)	600	101,816	169.7

It needs to be noted that the values presented in Table 17 are those obtained following the processing of lyrics; i.e., removal of repeats, metadata, etc. In this respect, the MC has a total of 101,816 words which is almost equally distributed across three subgenres. TMSC has the highest number of words

(34,799) and is followed by DMSC by 33,859 words. The values converge to each other by 165.8, 174 and 169.3 words per song, respectively. As mentioned earlier in Lyrics Corpora and Studies, the average number of words per song was found to differ between 161 and 443. Nevertheless, these values were obtained from corpora which did not exclude the repeats. Falk's (2012) methodology is comparable as it followed the same approach and achieved an average word number of 176 per song. The process of removing the repeats seems to have normalized the song lengths on an equal basis.

3.2. The Analyses

The study has been designed to focus on lyrics in a lexical approach and based on counts and rates of lexical items as proposed by Biber et al. (1998: 269-73). The aim was to use multiple methods of a single dimension of lexical richness so that a more comparable and broad set of results could be attained. This is in line with Malvern and Richards's (2012: 4) claim that there is no single ideal measure. In addition to lexical richness, a keyness analysis was also conducted. The rationale behind the use of these two methods is to employ a mixture of both quantitative and qualitative measures at the level of lexis while maintaining objectivity and comparability at the same time.

Four different measures of lexical variation were chosen for analyzing the corpus. In line with Torruella and Capsada's (2013) division of lexical variation measurement methods into three classes, one measure from each class was chosen. Guiraud was chosen from the first class as it was reported to be a better transformation of the TTR and performing better (van Hout and Vermeer 1988 and Vermeer 2000 qtd. in Daller et al. 2003: 7; Park 2013: 135). However, TTR was also added to investigate whether it performed as inadequately on songs as reported for other text types and lengths.

From the second class of measures, which are based on logarithmic transformations, Uber index was selected. Zheng (2016) used Uber index together with B2K measure. Uber index was also found by Jarvis (2002) to be one of the best formulas in curve fitting. Šišková (2012: 32) found that Uber results were correlated to a very high extent to those of *Vocd*.

Finally, from the complex measures, HD-D was chosen. Ellis and Juan (2004 and 2005) found the measure stable and able to differentiate between registers. HD-D was chosen as it is a more direct measure of the "industry standard" *Vocd* (McCarthy and Jarvis 2010: 387) which has been used in numerous studies (e.g., Daller and Phelan 2007, Šišková 2012, Koizumi and In'nami 2012, Torruella and Capsada 2013, Lissón and Ballier 2018, etc.).

Similar to the lexical variation measures, three different lexical sophistication measures have been used in the present study – LFP and P_Lex. LFP has been operationalized as B2K, which is a condensed and more comparable version of the original LFP (Laufer 1995: 267). B2K is further split into two according to the frequency lists it operates: the GSL and the BNC/COCA. The BNC/COCA offers a very good coverage as it is the combination of the frequency lists of two very large general corpora. Another advantage of the BNC/COCA is that it merges British and American English words, thus providing an optimal list for the analysis of texts produced by British and American artists. The GSL was also used both because it is still a widely used combination and in order to examine its coverage against the BNC/COCA. As Brezina and Gablasova's (2013) new-GSL and Browne's (2013) NGSL are new and have almost no reference in the literature they were not used. The advantages and disadvantages of these lists have been discussed in the literature review and both lists have been used in order to find out the one that differentiates better. The most frequent words in all subcorpora have been identified by means of the *AntWordProfiler* (Anthony 2014) software which automatically generates a list of the most frequent words in a given text.

P_Lex, on the other hand, is the second most frequently used extrinsic lexical sophistication method. It is claimed by its creators, Meara and Bell (2001), to be effective with short texts, which makes it a reasonable choice for songs as their lengths range between 100 and 400 words in the present study. They also advocate that the scores obtained from P_Lex are comparable to those from LFP; therefore, the present study will test to what extent this claim holds with song lyrics.

The present study employed the tokens and types as the unit of measurement as 1) the text coverage is based on the ratio of tokens and 2) related research has followed the same approach (see: Meara 1993, Katagiri and Kawai 2008, Sundberg 2015 and Öztürk 2017) and 3) lemmas or word families would lead to distorted results as the analysis methods are mostly dependent on type and token ratios. The removal of the repeats in song lyrics serves this method well as only the unique parts of the texts will be retained, keeping the numbers of repeated tokens at a minimum, if not eliminating completely – for further details see Compiling and Processing the Data.

Intrinsic measures such as hapax legomena and dislegomena have not been used as the scores obtained will not be comparable and supported by well-established frequency lists. Other extrinsic measures such as S and LEX have not been used, either, as they have not been used much in the literature and a comparison will be quite difficult.

As regards the lexical density, the ratio of content words to total number of token was applied. The method of calculating the ratio of content words to function words was not considered, as it is not a common method and gives parallel results according to Ishikawa (2007). Clause-based approaches were not considered, either, due to the fact that not all words occur in clauses in song lyrics. Furthermore, punctuation is an important marker in the distinction of clauses. However, it fails to be so for song lyrics due to typos and variations of texts from different sources. Therefore, an approach to determine full sentences on the basis of counting the full-stops and other punctuation marks is bound to provide different results for the same lyrics when they are acquired from different sources. In song lyrics, it is rarely the case that all lines correspond to a full sentence – or a clause. The lines are formed by clauses and phrases most of the time. A look at the excerpt from thrash_us_09-01-01 could support this claim. The lines are numbered for better reference.

1. *Infiltration push reserves*
2. *Encircle the front lines*
3. *Supreme art of strategy*
4. *Playing on the minds*

5. *Bombard till submission*
6. *Take all to their graves*

7. *Indication of triumph*
8. *The number's that are dead*

A search for full sentences will give us only three lines (2, 5 and 6) all of which are imperatives. Furthermore, punctuation – let alone spelling – on lyrics sheets (album sleeves) are not the best examples of grammatical pinnacle. For instance, the apostrophe in Line 8 (*number's*) is excessive. Even though identifying the clauses in song lyrics could be a more preferable way than identifying sentences, not all lines correspond to a clause as can be seen in lines 3 and 7 in the example above. *Supreme art of strategy* and *Indication of triumph* are noun phrases constituting a line each. Halliday's (1989) method did not take phrases into consideration. Hence, this method may not be fully applicable to song lyrics analysis.

Lexical originality (LO) was not selected because it is dependent on the members (in this case the songs) in a group. The obtained results could not have been comparable to other bands or songs in a new corpus which in turn would deem the method incapable of replication. Furthermore, a standard by which to sort unique words might be necessary for an LO analysis. Proper nouns, for instance, could be uniquely used but whether it could be regarded as a marker of lexical richness or not is debatable.

Lexical richness analyses were carried out using four different tools and software. Lexical diversity measures – Uber and HD-D – were calculated with koRpus, which is an R plugin developed by Michalke (2018). The free plugin can be downloaded from https://reaktanz.de/?c=hacking&s=koRpus. Lexical sophistication was calculated with *AntWordProfiler*. The software is developed by Anthony (2014) and freely available at http://www.laurenceanthony.net/software/antwordprofiler/. The percentages of B2K tokens and types were calculated based on the GSL and the BNC/COCA frequency lists. The last measure of lexical sophistication, P_Lex, was calculated using an online tool residing at http://www.lognostics.co.uk/tools/P_Lex/P_Lex.htm (Meara 2018). Unlike the previous software mentioned, P_Lex calculator does not batch calculate multiple files; therefore, all files were manually entered and results were recorded. Finally, lexical density was calculated using the online version of Lexical Complexity

Analyzer (https://aihaiyang.com/software/lca/batch/) developed by Ai and Lu (2010) and Lu (2012). The tool automatically POS-tags the input files and generates the results as .csv files. The songs by British and American bands were calculated separately so that they can be tagged accordingly.

The measurement methods for lexical richness were introduced in the previous chapter (see Lexical Richness). The selected corpus of songs will be analyzed for their lexical richness using data-driven and quantitative methods. The variety of the number of analysis is expected to provide an opportunity for triangulation of the results. Gregori-Signes and Clavel-Arroitia (2015), similarly, used two methods (LD and LFP) in the assessment of the lexical richness of L2 students' essays, which is quite similar to the present study in terms of methodology. Johansson (2008: 62) emphasizes the usage of different methods is the assessment of text complexity, namely lexical diversity (ratio of different words used in a text) and lexical density, as follows,

> [N]either lexical diversity nor lexical density is the one and only measure. However, both measures are easily accessible and easy to apply to corpora of different kinds. No doubt they also provide important insights into the texts, and as long as the measures are not used as the only way to judge a text qualitatively, they are very useful.

The chosen statistical measures were Kolmogorov-Smirnov test for normality of distribution. The chosen lexical values were compared to each other based on the subgenres using an ANOVA test, as proposed by Biber et al. (1998: 273) and Daller and Xue (2007: 160), where values are normally distributed. Where not applicable, the non-parametric variant of ANOVA, Kruskal-Wallis H tests were applied. When differences are observed between the subgenres, Mann-Whitney U tests were applied to check which groups differed from which and to what extent. Following the ANOVA test, upon detection that the variances are not equal by means of a Levene test, a Welch ANOVA test was applied. The test was complemented with a Dunnett's C post hoc test to find out the differences between the groups. After all comparisons, the values were tested for their correlation levels as applied by Daller and Xue (2007) and Zheng (2016). A Spearman's correlation test was administered due to the fact that the data is not normally distributed. Finally, a quadratic discriminant analysis was carried out to observe how much the lexical values corresponded to the existing subgenres. As this analysis required normally distributed data, the values were normalized on SPSS 22 and then the analysis was run.

Keyness analysis was carried out using *AntConc* v 3.5.7 (Anthony 2018). It is one of the major keyness analysis tools used in numerous studies. It has been cited 1861 times according to Google Scholar as of 14.02.2020 (https://scholar.google.co.jp/citations?view_op=view_citation&hl=en&user=DS8jIxUAAAAJ&citation_for_view=DS8jIxUAAAAJ:oPLKW5k6eA4C). The lyrics, which had been saved in .txt format in separate files, were loaded into the software one subcorpus at a time. The keyness levels for the subcorpora were calculated based on the BNC owing to the fact that 1) The BNC/COCA lists are not available for the software and 2) it is a large enough and reliable corpus of English language and many studies have used the same corpus in their analyses (e.g., Xiao and McEnery 2005, Scott 2009, Stubbs 2010, Warren 2010, Yang 2012, etc.). Xiao and McEnery (2005) even tested an American English corpus against the BNC and FLOB and his results were very similar to those obtained from the MDA. Their justification was that the genres they studied contained similar amount of Americanisms so that they would all be equally affected by any difference. Adapting this methodology here, it could be claimed that all subgenres in metal, regardless of the origin of the bands and composers, are expected to contain similar linguistic items (NNS factor will be discussed in the limitations). Therefore, the use of the BNC in the present study is considered appropriate.

Only the content words were taken into consideration as it is the open-class words that reveal more about the lexical choices of the authors. They "shed light on what is important in a text" (Leone 2010: 244). This approach has been followed by Baker (2006) and Yang (2012). Closed-class words could also be used in analysis, as Groom (2010), Burns et al. (2014) or Fraysse-Kim (2010) did, if the focus is on grammatical structures and style (Baker 2006 and Groom 2010). Groom (2010: 73) advocates this as he thinks that closed-class words reveal underlying factors when concordanced which could not be uncovered merely with open-class words. Culpeper (2014: 20) remark that open-class words are related with aboutness, whereas closed-class ones are with stylistics.

Baker's (2004) and Culpeper's (2014) approach has been adopted in the present study in that the words were not lemmatized as they may not be fully accurate. Utka (2004 qtd. in Baker 2004: 354-5), on the other hand, chose to lemmatize the words in Orwell's 1984 in his research on keywords. Baker (2004: 355) opposes this approach claiming that different senses of the same word could be collocated with different words and lemmatization may obscure these differences.

In terms of statistics, the log-likelihood measure has been chosen. Rayson (2003), Ignat et al. (2006), Baker (2004, 2006), Archer et al. (2009), Scott (2009), Culpeper (2014) used the same measure in their studies. It is proposed that in comparing genres against a large reference corpus log-likelihood method gives better keyness estimates (Scott 2013: 209 qtd. in Culpeper 2014: 12). In fact, Culpeper (2014: 15-6) also used the Chi-squared test on his analyses and found only minor differences in the results. The p value has been set as $p < 0.000001$ in the same vein as Baker (2004, 2006) and Scott (2009) in order to get as few and manageable keywords as possible (Scott 1999 qtd. in Baker 2004: 352).

The cut-off point is a very controversial issue in keyword analyses. As Baker (2006: 173) notes, humans like to round the numbers up and look at the first 10, 20, 50, etc. words. However, the next word after the cut-off point could be statistically equally key as the preceding one. Still, the cut-off points are up to the research design. Baker (2004), for instance, set the cut-off point at the significance level of $p= .000001$. Since the present study seeks to compare subgenres on an equal level, a specific cut-off point has been set at 30 words. The first 30 keywords have then been simplified by removing the function words.

Finally, the keywords have been concordanced to the context they occur in. Concordancing is the investigation of a particular word based on the words that precede or follow them. It is "a collection of the occurrences of a word-form, each in its own textual environment" Sinclair (1991: 32). All the occurrences of a word to be looked for can be listed by a concordancer (Leech 1992: 114). Concordancers have been used as a major tool for analyzing corpora to identify polysemy, word-class ambiguity, etc., which is not possible with wordlists (Kennedy 1998: 247). They allow searching for words in an untagged text (Kübler and Zinsmeister 2015: 197). The words are generally displayed in Key Word in Context (KWIC) format (Kennedy 1998: 251 and Kübler and Zinsmeister 2015: 197). Baker (2012: 108-9), for instance, after finding out the keywords in the articles on metrosexuality, looked at the concordances of the keywords. This procedure allowed him to see and comment on the individual contexts within which the keywords are mostly used. Similarly, the concordances for the most key words will be provided in the discussion section.

4. RESULTS

The results of lexical variation, lexical sophistication and lexical density analyses will be presented separately. The relevant data and tables detailing the results are available under each analysis heading. Firstly, a Kolmogorov-Smirnov Test was conducted to find out the distribution of normality for each test.

Table 18: Kolmogorov-Smirnov test results for each variable.

		Lexical Density	GSL B2K Tokens	GSL B2K Types	BNC/COCA B2K Tokens	BNC/COCA B2K Types	P_Lex (Lambda)	TTR	R	U	HDD
N		600	600	600	600	600	600	600	600	600	600
Normal Parameters	Mean	0.54	13.45	18.46	13.05	17.89	2.33	0.6	7.66	23.78	33.72
	Std. Deviation	0.07	9.08	11.29	9.11	11.42	1.12	0.1	1.38	7.44	2.47
Most Extreme Differences	Absolute	0.09	0.097	0.091	0.085	0.080	0.066	0.052	0.025	0.049	0.092
	Positive	0.09	0.097	0.091	0.085	0.080	0.066	0.035	0.025	0.049	0.05
	Negative	-0.045	-0.082	-0.067	-0.081	-0.063	-0.036	-0.052	-0.02	-0.039	-0.092
Test Statistic		0.09	0.097	0.091	0.085	0.080	0.066	0.052	0.025	0.049	0.092
Asymp. Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.200	.002	.000

As shown in Table 18, only Guiraud (R) values are normally distributed (p=.200) for each parameter at p=0.01. Therefore, these results were tested with ANOVA and the remaining ones with Kruskal-Wallis H test.

4.1. Lexical Richness

The results of all dimensions of lexical richness will be reported in separate sections. Later, the results of a correlation analysis of the results will be provided.

4.2. Lexical Variation

The selected analysis methods for LV were TTR, Guiraud, Uber and HD-D. The results of the Kolmogorov-Smirnov normality test given in Table 18 showed that Guiraud values were normally distributed whereas TTR, Uber and HD-D were not. Therefore, they were grouped accordingly and Guiraud results were subjected to an Analysis of Variance (ANOVA) test, as suggested by Biber et al. (1998: 273), and the others to a Kruskal-Wallis H test. The descriptive values for the Kruskal-Wallis H test are provided in Table 19.

Table 19: Descriptive statistics for TTR, Uber (U) and HD-D values.

Subcorpus	LV Method	N	Mean	Std. Deviation	Minimum	Maximum
HMSC	TTR	200	.54	.11	.24	.75
	U	200	19.05	5.41	7.97	36.88
	HD-D	200	32.07	2.54	23.04	37.33
TMSC	TTR	200	.61	.08	.32	.79
	U	200	23.93	5.41	8.34	40.97
	HD-D	200	34.18	1.83	24.16	38.06
DMSC	TTR	200	.66	.09	.31	.85
	U	200	28.35	8.04	10.87	62.89
	HD-D	200	34.92	2.04	25.83	39.03

The TTR values for HMSC, TMSC and DMSC are .54 (±.11), .61 (±.08) and .66 (±.09) respectively. U values are 19.05 (±5.41) (HMSC), 23.93 (±5.41) (TMSC) and 28.35 (±8.04) (DMSC). HD-D scores also display and increasing pattern as 32.07 (±2.54) for HMSC, 34.18 (±1.83) for TMSC and 34.92 (±2.04) for DMSC.

Table 20: Results of the Kruskal-Wallis H test for TTR, Uber (U) and HD-D.

	Subcorpus	N	Mean Rank	df	$\chi 2$	p
TTR	HMSC	200	197.36	2	132.669	.000
	TMSC	200	307.49			
	DMSC	200	396.66			
U	HMSC	200	181.97	2	166.937	.000
	TMSC	200	315.01			
	DMSC	200	404.53			
HD-D	HMSC	200	183.52	2	151.905	.000
	TMSC	200	325.09			
	DMSC	200	392.89			

The results of the Kruskal-Wallis H test shown in Table 20, pointed to significant differences across subcorpora for TTR ($\chi 2(2) = 132.669$, $p= 0.000$), U ($\chi 2(2) = 166{,}937$, $p= 0.000$) and HD-D ($\chi 2(2) = 151.905$, $p= 0.000$). In order to find out the differences between the subcorpora a Mann-Whitney U test was conducted.

Table 21: Mann-Whitney U test results for TTR, Uber (U) and HD-D.

	Subcorpus	N	Mean Rank	Sum of Ranks	U	p
TTR	HMSC	200	159.97	31994.50	11894.500	.000
	TMSC	200	241.03	48205.50		
U	HMSC	200	150.24	30047.00	9947.000	.000
	TMSC	200	250.77	50153.00		
HD-D	HMSC	200	148.71	29742.00	9642.000	.000
	TMSC	200	252.29	50458.00		
TTR	HMSC	200	137.89	27577.00	7477.000	.000
	DMSC	200	263.12	52623.00		
U	HMSC	200	132.23	26446.00	6346.000	.000
	DMSC	200	268.77	53754.00		
HD-D	HMSC	200	135.31	27062.50	6962.500	.000
	DMSC	200	265.69	53137.50		
TTR	TMSC	200	166.96	33391.50	13291.500	.000
	DMSC	200	234.04	46808.50		
U	TMSC	200	164.75	32949.00	12849.000	.000
	DMSC	200	236.26	47251.00		
HD-D	TMSC	200	173.30	34660.50	14560.500	.000
	DMSC	200	227.70	45539.50		

The results of the Mann-Whitney U test, as shown in Table 21, revealed that there are significant differences between all three subcorpora both in terms of TTR, Uber (U) and HD-D. The TTR values for HMSC (mean=159.97) are significantly lower than TMSC (241.03) (U = 11894.500, p= .000). The difference between TMSC (mean=166.96) and DMSC (mean=234.04) results are also significant as well (U = 13291.500, p= .000). The difference between HMSC (mean=137.89) and DMSC (mean=263.12) is the largest (U = 7477.000, p= .000).

The differences between the genres are also significant according to Uber results. HMSC (150.24) scores are higher than TMSC (250.77) (U = 9947.000, p= .000) and TMSC (164.75) are higher DMSC (236.26) (U = 12849.000, p= .000). The difference is the largest between HMSC (132.23) and DMSC (268.77) (U = 6346.000, p= .000).

HD-D values were similar to TTR and Uber. TMSC values (mean=252.29) are significantly higher than HMSC (mean=148.71) (U = 9642.000, p= .000) but significantly lower than DMSC (TMSC mean=173.30, DMSC mean= 227.70) (U = 14560.500, p= .000). Finally, the difference between HMSC (mean=135.31) and DMSC (mean=265.69) is the largest (U = 6962.500, p= .000).

Following the non-parametric statistical analyses of TTR, Uber (U) and HD-D values, the Guiraud values were tested for their statistical significance.

Table 22: Descriptive statistics for Guiraud values.

	Subcorpus	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Guiraud	HMSC	200	6.75	1.11	0.08	6.60	6.91	3.13	9.57
	TMSC	200	7.88	1.12	0.08	7.73	8.04	3.32	11.43
	DMSC	200	8.35	1.39	0.10	8.15	8.54	4.64	12.64

The Guiraud (R) mean values for HMSC, TMSC and DMSC, as provided in Table 22, are 6.75 (±1.11), 7.88 (±1.12) and 8.35 (±1.39) respectively. A Levene test was conducted to find out whether variances are equal. The results given in Table 23 showed that the variances are not equal F(2,597) = 5.847, *p*= 0.003.

Table 23: Levene test results for Guiraud.

	Levene Statistic	df1	df2	Sig.
Guiraud	5.847	2	597	.003

Table 24: Welch ANOVA test results for Guiraud values

	Statistic	df1	df2	Sig.
Welch	93,650	2	394.255	.000

Since the homogeneity of variance was not met, A Welch ANOVA test was conducted (Table 24). The results showed significant results between at least two groups (*Welch's F*(2,394.255) = 93,650, *p*=.000). In order to find out the differences between the groups, a Dunnett's C post hoc test was carried out.

Table 25: Dunnett's C post hoc test results for Guiraud values

Subcorpus		Mean Difference	Std. Error	99% Confidence Interval	
				Lower Bound	Upper Bound
HMSC	TMSC	-1.13*	.11	-1.46	-.80
	DMSC	-1.59*	.13	-1.96	-1.22
TMSC	HMSC	1.13*	.11	0.80	1.46
	DMSC	-.46*	.13	-0.83	-.09
DMSC	HMSC	1.59*	.13	1.22	1.96
	TMSC	.46*	.13	0.09	.83

*. The mean difference is significant at the 0.01 level.

The results of the Dunnett's C, given in Table 25, showed significant differences between each pair (*p*<0.01). The largest difference was observed between HMSC and DMSC (1.59 ±.13) and the smallest one is between TMSC and DMSC (.46 ±.13). The range of difference between HMSC and TMSC was 1.13 ±.11.

4.3. Lexical Sophistication

LS analyses were based on two different methods: LFP and P_Lex. LFP analyses were carried out based on both tokens and types as both methods are used in the literature. The first analysis was carried out using the GSL and the descriptive values are shown Table 26.

Table 26: Descriptive statistics for GSL B2K tokens and types in percentages.

	Subcorpus	N	Mean	Std. Deviation	Minimum	Maximum
GSL B2K Tokens	HMSC	200	7.19	4.57	0	24.09
	TMSC	200	12.71	6.19	1.23	30.94
	DMSC	200	20.46	9.98	2.68	67.29
GSL B2K Types	HMSC	200	10.12	5.91	0	30.48
	TMSC	200	18.02	8.24	2.15	45
	DMSC	200	27.22	11.55	3.28	70.89

The mean values of the tokens outside the first two thousand words in the GSL for HMSC is 7.19%, for TMSC 12.71% and for DMSC 20.46%. The distribution for the same frequency band based on types is 10.12% for HMSC, 18.02% for TMSC and 27.22% for DMSC.

Table 27: Results of the Kruskal-Wallis H test for GSL B2K tokens and types.

	Subcorpus	N	Mean Rank	df	$\chi 2$	p
GSL B2K Tokens	HMSC	200	164.22	2	239.894	.000
	TMSC	200	304.67			
	DMSC	200	432.61			
GSL B2K Types	HMSC	200	157.69	2	254.190	.000
	TMSC	200	310.27			
	DMSC	200	433.55			

A Kruskal-Wallis H test (Table 27) indicated that the three subcorpora differed from each other significantly both in terms of tokens ($\chi 2(2) = 239.894$, $p= 0.000$) and types ($\chi 2(2) = 254.190$, $p= 0.000$), a Mann-Whitney U test was conducted to find out the differences between the subgenres.

Table 28: Mann-Whitney U test results for GSL B2K token and type values.

	Subcorpus	N	Mean Rank	Sum of Ranks	U	p
GSL B2K Tokens	HMSC	200	146.53	29305.00	9205.000	.000
	TMSC	200	254.48	50895.00		
GSL B2K Types	HMSC	200	142.25	28450.00	8350.500	.000
	TMSC	200	258.75	51750.00		
GSL B2K Tokens	HMSC	200	118.19	23638.50	3538.500	.000
	DMSC	200	282.81	56561.50		
GSL B2K Types	HMSC	200	115.94	23187.50	3087.500	.000
	DMSC	200	285.06	57012.50		
GSL B2K Tokens	TMSC	200	150.70	30139.00	10039.000	.000
	DMSC	200	250.31	50061.00		
GSL B2K Types	TMSC	200	152.02	30403.50	10303.500	.000
	DMSC	200	248.98	49796.50		

Mann-Whitney U test results, shown in Table 28, indicate significant differences between all three subcorpora both in terms of GSL B2K tokens and types. The token values for HMSC (mean=146.53) are significantly lower than TMSC (254.48) (U = 9205.000, p= .000). The difference between TMSC (mean=150.70) and DMSC (mean=250.31) results are also significant as well (U = 10039.000, p= .000). The difference between HMSC (mean=118.19) and DMSC (mean=282.81) is the largest (U = 3538.500, p= .000).

A similar pattern is observable in type values across three subcorpora. TMSC values (mean=258.75) are significantly higher than HMSC (mean=142.25) (U = 8350.500, p= .000) but significantly lower than DMSC (TMSC mean=152.02, DMSC mean= 248.98) (U = 10303.500, p= .000). The difference between HMSC (mean=115.94) and DMSC (mean=285.06) is again the largest (U = 3087.500, p= .000).

Following the GSL B2K token and type comparison across three subgenres, the same procedure was repeated using the BNC/COCA wordlist and the B2K results are provided in the tables Table 29 and Table 30.

Table 29: Descriptive statistics for BNC/COCA B2K tokens and types in percentages.

	Subcorpus	N	Mean	Std. Deviation	Minimum	Maximum
BNC/COCA B2K Tokens	HMSC	200	6.83	4.75	0	26.05
	TMSC	200	12.11	6.24	1.23	30.15
	DMSC	200	20.21	9.85	3.23	70.09
BNC/COCA B2K Types	HMSC	200	9.50	6.24	0	34.29
	TMSC	200	17.18	8.43	2.15	47.50
	DMSC	200	27.00	11.33	5.88	74.68

BNC/COCA B2K token percentage values for HMSC is 6.83%, for TMSC 12.11% and for DMSC 20.21%. The distribution for the same frequency band based on types is 9.50% for HMSC, 17.18% for TMSC and 27.00% for DMSC.

Table 30: Results of the Kruskal-Wallis H test for BNC/COCA B2K tokens and types.

	Subcorpus	N	Mean Rank	df	Chi-Square	p
BNC/COCA B2K Tokens	HMSC	200	166.34	2.00	241.386	.000
	TMSC	200	299.51			
	DMSC	200	435.66			
BNC/COCA B2K Types	HMSC	200	160.12	2.00	256.502	.000
	TMSC	200	303.69			
	DMSC	200	437.69			

Table 30 shows the Kruskal-Wallis H test results for BNC/COCA B2K tokens and types. As is the case with the GSL B2K results, all subcorpora show significant differences between each other both in terms of tokens ($\chi 2(2)$ = 241.386, p= 0.000) and types ($\chi 2(2)$ = 256.502, p= 0.000). In order to find out if there are differences between each group, a Mann-Whitney U test was conducted.

Table 31: Mann-Whitney U test results for BNC/COCA B2K token and type values.

	Subcorpus	N	Mean Rank	Sum of Ranks	U	p
BNC/COCA B2K Tokens	HMSC	200	149.45	29890.50	9790.500	.000
	TMSC	200	251.55	50309.50		
BNC/COCA B2K Types	HMSC	200	145.66	29132.00	9032.000	.000
	TMSC	200	255.34	51068.00		
BNC/COCA B2K Tokens	HMSC	200	117.38	23476.50	3376.500	.000
	DMSC	200	283.62	56723.50		
BNC/COCA B2K Types	HMSC	200	114.96	22991.50	2891.500	.000
	DMSC	200	286.04	57208.50		
BNC/COCA B2K Tokens	TMSC	200	148.46	29692.50	9592.500	.000
	DMSC	200	252.54	50507.50		
BNC/COCA B2K Types	TMSC	200	148.85	29770.00	9670.000	.000
	DMSC	200	252.15	50430.00		

As described in Table 31, a Mann-Whitney U test showed significant differences between all three subcorpora both in terms of BNC/COCA B2K tokens and types. The token values for HMSC (mean=149.45) are significantly lower than TMSC (251.55) (U = 9790.500, p= .000). The difference between TMSC (mean=148.46) and DMSC (mean=252.54) results are significant as well (U = 9592.500, p= .000). The difference between HMSC (mean=117.38) and DMSC (mean=273.62) is the largest (U = 3376.500, p= .000).

In terms of type values, the results were similar. TMSC values (mean=255.34) are significantly higher than HMSC (mean=145.66) (U = 9032.000, p= .000) but significantly lower than DMSC (TMSC mean=148.85, DMSC mean= 252.15) (U = 9670.000, p= .000). The difference between HMSC (mean=114.96) and DMSC (mean=286.04) is again the largest (U = 2891.500, p= .000).

As the second method of LS, P_Lex has been used and the results of the Kruskal-Wallis H test are provided in Table 32.

Table 32: Kruskal-Wallis H descriptive statistics for P_Lex values.

Subcorpus	N	Mean	Std. Deviation	Minimum	Maximum
HMSC	200	1.53	0.75	0.08	4.38
TMSC	200	2.23	0.84	0.62	4.76
DMSC	200	3.2	1.04	1	7.8

The subgenres display difference between each other in terms of P_Lex values ($\chi 2(2) = 237.313$, $p= 0.000$) with mean values of 1.53 for HMSC, 2.23 for DMSC and 3.2 for DMSC (see Table 33).

Table 33: Kruskal-Wallis H test results for P_Lex values.

Subcorpus	N	Mean Rank	df	Chi-Square	p
HMSC	200	169.8575	2	237.3134	.000
TMSC	200	294.92			
DMSC	200	436.7225			

The results indicated significant differences between subcorpora ($\chi 2(2) = 237.3134$, $p= 0.000$) with mean values of 169.8575for HMSC, 294.92 for TMSC and 436.7225 for DMSC. Therefore, a Mann-Whitney U test was conducted to find out the differences between the subcorpora of which results are provided in Table 34.

Table 34: Mann-Whitney U test results for P_Lex values.

Subcorpus	N	Mean Rank	Sum of Ranks	U	p
HMSC	200	152.72	30543.5	10443.500	.000
TMSC	200	248.28	49656.5		
HMSC	200	117.64	23528.00	3428.000	.000
DMSC	200	283.36	56672.00		
TMSC	200	147.14	29427.50	9327.50	.000
DMSC	200	253.86	50772.50		

According to the Mann-Whitney U test results as shown in Table 34, there are differences between the three subgenres. The P_Lex values for HMSC (mean=152.72) are significantly lower than TMSC (mean=248.28) ($U = 10443.500$, $p= .000$). The difference between TMSC (mean=147.14) and DMSC (mean=253.86) results are also significant ($U = 9327.500$, $p= .000$). The difference between HMSC (mean=117.64) and DMSC (mean=283.36) is the largest (U = 3428.000, p= .000). The gradual increase from HMSC to DMSC is observable in P_Lex values as well.

4.4. Lexical Density

The LD calculations of the songs in MC are shown in Table 35. The maximum value for DMSC is 0.83 while it is 0.71 for the other subcorpora. The values of DMSC (mean=0.57805) are greater than those of TMSC (mean=0.5342). Similarly, the values of TMSC are higher than those of HMSC (mean=0.4931). This pattern shows a gradual increase from HMSC to DMSC, placing TMSC in between.

Table 35: Descriptive statistics for LD values.

Subcorpus	N	Mean	Std. Deviation	Minimum	Maximum
HMSC	200	0.4931	0.054297	0.36	0.71
TMSC	200	0.5342	0.058278	0.42	0.71
DMSC	200	0.57805	0.076017	0.37	0.83

The results of the LD levels of the songs in the MC do not show normal distribution, either. Therefore, a Kruskal-Wallis H test was conducted and the results are shown in Table 36.

Table 36: Results of the Kruskal-Wallis H test for LD values.

Subcorpus	N	Mean Rank	df	$\chi 2$	p
HMSC	200	195.08	2	143.049	.000
TMSC	200	304.37			
DMSC	200	402.06			

As the results show difference between each other ($\chi 2(2)$ = 143.049, p= 0.000) with mean values of 195.08 for HMSC, 304.37 for TMSC and 402.06 for DMSC, a Mann-Whitney U test was conducted to find out the differences between the subgenres.

Table 37: Mann-Whitney U test results for LD values.

Subcorpus	N	Mean Rank	Sum of Ranks	U	p
HMSC	200	161.22	32244.50	12144.500	.000
TMSC	200	239.78	47955.50		
HMSC	200	134.35	26870.50	6770.500	.000
DMSC	200	266.65	53329.50		
TMSC	200	165.09	33018.00	12918.000	.000
DMSC	200	235.91	47182.00		

Mann-Whitney U test results, shown in Table 37, indicate significant differences between all three subgenres. The LD values for HMSC (mean=161.22) are significantly lower than TMSC (mean=239.78) (U = 12144.500, p= .000). The difference between TMSC (mean=165.09) and DMSC (mean=235.91) results are significant as well (U = 12918.000, p= .000). The difference between HMSC (mean=134.35) and DMSC (mean=266.65) is the largest (U = 6770.500, p= .000).

4.5. Correlation of Lexical Richness Measures

Following all the lexical richness analyses, a correlation test was conducted to identify which measures correlate with each other and to what extent.

Table 38: Spearman's Correlation Analysis results for lexical richness values.

	Lexical Density	GSL B2K Tokens	GSL B2K Types	BNC/COCA B2K Tokens	BNC/COCA B2K Types	P_Lex	TTR	R	U	HD-D
Lexical Density	-									
GSL B2K Tokens	.701**	-								
GSL B2K Types	.657**	.963**	-							
BNC/COCA B2K Tokens	.684**	.949**	.927**	-						
BNC/COCA B2K Types	.643**	.919**	.957**	.966**	-					
P_Lex (Lambda)	.740**	.899**	.881**	.918**	.900**	-				
TTR	.554**	.538**	.489**	.520**	.475**	.552**	-			
R	.449**	.526**	.558**	.522**	.555**	.512**	.617**	-		
U	.574**	.597**	.576**	.583**	.564**	.597**	.934**	.844**	-	
HD-D	.543**	.533**	.509**	.505**	.484**	.521**	.832**	.828**	.926**	-

**. Correlation is significant at the 0.01 level (2-tailed).

The results of the correlation analysis, shown in Table 38, indicate significant correlations between all values (p<0.01) ranging between r=.449 and r=.966. The correlation between BNC/COCA B2K types and tokens is the highest (r=.966). A very close value is observed between GSL B2K types and tokens (r=.963). BNC/COCA B2K types and GSL B2K types have the third highest correlation (r=.957), which is followed by BNC/COCA B2K tokens and GSL B2K tokens (r=.949).

The other LS measure, P_Lex, has been found to correlate highly with the B2K results. The correlation values are r=.899, r=.881, r=.918 and r=.900 for GSL B2K tokens, GSL B2K types, BNC/COCA B2K tokens and BNC/COCA B2K types respectively. The highest correlation for LD was with P_Lex (r=.740). It was also highly correlated with B2K values – r=.701 for GSL B2K tokens, r=.684 for BNC/COCA B2K tokens, r=.657 for GSL B2K types, r=.643 for BNC/COCA B2K types.

In terms of LV the measures, the highest correlation has been observed between TTR and Uber (r=.934) which is followed by Uber and HD-D (r=.926). Uber is also very highly correlated with Guiraud (r=.844). HD-D has been found to correlate highly with both TTR and Guiraud (r=.832 and r=.828). The weakest, although significant, correlation has been observed between TTR and Guiraud (r=.617).

Correlation levels decrease between different dimensions of lexical richness although they are all statistically significant. LD correlates with LV results in a range between r=.574 (Uber) and r=.449 (Guiraud). LV and LS measures correlate with each other in varying degrees from r=.597 (Uber and GSL B2K tokens) to r=.475 (TTR and BNC/COCA types).

4.6. Quadratic Discriminant Analysis

All the analyses so far showed significant differences between subcorpora. As the final investigation of lexical richness, a quadratic discriminant analysis was carried out in order to find out to what extent the existing subgenre divisions overlap with the findings of the lexical richness analysis. Since the data is not normally distributed, the values were normalized using SPSS as the first step.

Then the quadratic discriminant analysis was run of which results are shown in Table 39.

Table 39: Results of the quadratic discriminant analysis for lexical richness

Subcorpus			Predicted Group Membership			Total
			HMSC	TMSC	DMSC	
Original	Count	HMSC	**141**	47	12	200
		TMSC	36	**118**	46	200
		DMSC	14	58	**128**	200
	%	HMSC	**70.5**	23.5	6.0	100.0
		TMSC	18.0	**59.0**	23.0	100.0
		DMSC	7.0	29.0	**64.0**	100.0

The results show that 141 out of 200 songs in HMSC (70.5%) are predicted as HMSC, marking it as the highest ratio. 47 (23.5%) HMSC songs are predicted as TMSC and 12 (6%) as DMSC. Secondly, 118 songs out of 200 in TMSC (59%) are predicted as TMSC, 36 (18%) as HMSC and 46 (23%) as DMSC. Finally, 128 songs out of 200 in DMSC (64%) are predicted as DMSC, 58 (29%) as TMSC and 14 (7%) as HMSC. The overall prediction match rate has been found as 64.5%.

4.7. Keyness Analysis

The results of the keyness analysis are shown in the tables below as obtained from *AntConc* (Anthony 2018). Although the focus is on content words, function words will be briefed to a limited extent. As the corpus is not POS-tagged, some of the keywords such as *will, going, do, got*, etc. may actually be content words. In order to check whether or not this is the case, these function words were manually counted using the concordancing feature of *AntConc*.

Table 40: Top 30 keywords in HMSC.

#Keyword Types: 230				
#Keyword Tokens: 14,900				
Rank	**Frequency**	**Keyness**	**Effect**	**Keyword**
1	1143	+2303.38	0.0043	You
2	1181	+1867.22	0.0034	I
3	445	+1321.04	0.0061	Me
4	260	+1205.44	0.0087	Am
5	410	+1130.96	0.0055	Your
6	411	+1044.01	0.005	My
7	135	+520.39	0.005	Cannot
8	352	+457.01	0.0026	Will
9	167	+426.13	0.0039	Got
10	117	+423.9	0.0044	Love
11	258	+419.71	0.003	Do
12	52	+402.99	0.003	Hey
13	103	+331.3	0.0037	Oh
14	128	+287.72	0.0032	Never
15	102	+281.72	0.0033	Let
16	49	+276.15	0.0027	Soul
17	674	+271.3	0.0014	Is
18	102	+262.73	0.0031	Night
19	376	+254.47	0.0018	Not
20	53	+251.3	0.0028	Die
21	20	+231.34	0.0012	Mazes
22	36	+225.26	0.0021	Wheels
23	220	+210.97	0.002	No
24	138	+208.91	0.0024	Know
25	43	+207.77	0.0023	Fly
26	39	+203.19	0.0022	Cry
27	106	+198.52	0.0026	Going
28	151	+198.46	0.0023	Just
29	33	+198.09	0.0019	Forever
30	32	+197.16	0.0018	Ain

Out of a total of 33,158 total tokens in HMSC, 14,900 have been found as key. At the type level, the keyword value is 230. Table 40 shows the top 30 keywords in HMSC, which consist of 17 function words and 13 content words – as is the case with many other texts (see: Baker 2006: 122). The function words are mostly determiners, pronouns and auxiliaries. The first six content words have significantly higher keyness values ranging between +1044.01 and +2303.38. Content words, on the other hand, are verbs and nouns to a great extent and their keyness values start from +423.9 and go down to +198.09. The first key content word is interestingly *love*.

TMSC contains 11,264 keyword tokens and 220 types out of a total of 34,799 tokens. In this respect, the results look similar for HMSC and TMSC. Compared to HMSC, the number of function words is lower in TMSC. There are 12 function words and 18 content words in the top 30 keyword list as can be seen in Table 41. The keyness values of *you* and *your* are significantly higher than the rest of the keywords – +1823.83 and +1798.68 respectively. The first key content word in TMSC is *blood*, which is, in a way, indicative of a difference from HMSC.

Table 41: Top 30 keywords in TMSC.

#Keyword Types: 220				
#Keyword Tokens: 11,624				
Rank	**Frequency**	**Keyness**	**Effect**	**Keyword**
1	1034	+1823.83	0.0039	You
2	557	+1798.68	0.0074	Your
3	350	+760.79	0.0042	My
4	707	+534.9	0.002	I
5	253	+478.85	0.0035	Me
6	136	+451.39	0.0044	Am
7	97	+435.91	0.0043	Blood
8	351	+428.46	0.0026	Will
9	117	+422.75	0.0043	Death
10	78	+420.6	0.0039	Die
11	72	+406.6	0.0037	Kill
12	169	+405.07	0.0037	Life
13	295	+387.72	0.0027	No
14	67	+369.14	0.0034	Hell
15	109	+366.63	0.0039	Cannot
16	57	+332.52	0.003	Soul
17	218	+280.77	0.0025	Do
18	40	+252.28	0.0022	Burn
19	60	+232.34	0.0028	Fear
20	666	+227.83	0.0014	Is
21	277	+222.91	0.0019	All
22	126	+212.15	0.0026	Take
23	61	+207.85	0.0026	Dead
24	39	+206.57	0.0021	Evil
25	49	+193.15	0.0023	Pain
26	23	+188.65	0.0013	Insane
27	47	+188.05	0.0023	Fight
28	28	+187.57	0.0016	Fucking
29	76	+172.8	0.0025	Eyes
30	41	+171.94	0.0021	Lies

Despite the fact that HMSC and TMSC are somewhat similar in terms of function/content word distribution, DMSC differs markedly. It has 8,390 keyword tokens and 288 keyword types out of a total of 33,859 tokens. Although it has fewer keyword tokens than the other two subcorpora, it has more keyword types. As shown in Table 42, there are only 7 function words in the top 30 keywords in DMSC. Furthermore, the second keyword is a content word (*death*). The word *your* has the highest keyness value (+1242.18) which is followed by *death* (+621.19).

Table 42: Top 30 keywords in DMSC.

#Keyword Types: 288				
#Keyword Tokens: 8,390				
Rank	**Frequency**	**Keyness**	**Effect**	**Keyword**
1	437	+1242.18	0.0059	Your
2	150	+621.19	0.0056	Death
3	308	+618.96	0.0037	My
4	72	+489.25	0.004	Flesh
5	104	+487.58	0.0047	Blood
6	85	+477.75	0.0044	Die
7	71	+463.1	0.0039	Evil
8	47	+448.74	0.0027	Satan
9	171	+422.89	0.0038	Life
10	95	+408.53	0.0042	Dead
11	65	+399.82	0.0035	Soul
12	80	+395.04	0.0039	Pain
13	67	+373.15	0.0035	Kill
14	64	+350.87	0.0034	Hell
15	100	+281.56	0.0033	Am
16	67	+277.44	0.0031	Fear
17	285	+275.42	0.0021	Will
18	82	+245.08	0.0031	God
19	509	+208.86	0.0015	I
20	30	+202.02	0.0017	Decay
21	409	+201.66	0.0015	You
22	32	+189.09	0.0018	Forever
23	164	+186.3	0.0021	Now
24	75	+185.03	0.0026	Mind
25	156	+184.19	0.0022	Me
26	23	+177.88	0.0013	Mortal
27	32	+176.13	0.0018	Grave
28	21	+174.73	0.0012	Rotting
29	35	+169.71	0.0019	Darkness
30	22	+168.99	0.0013	Lust

The combined list of the keywords for all subcorpora is presented in Table 43. The content words are shown in bold.

Table 43: Keywords in all subcorpora.

	HMSC	**TMSC**	**DMSC**
1	You	You	Your
2	I	Your	**Death**
3	Me	My	My
4	Am	I	**Flesh**
5	Your	Me	**Blood**
6	My	Am	**Die**
7	Cannot	**Blood**	**Evil**
8	Will	Will	**Satan**
9	Got	**Death**	**Life**
10	**Love**	**Die**	**Dead**
11	Do	**Kill**	**Soul**
12	Hey	**Life**	**Pain**
13	Oh	No	**Kill**
14	**Never**	**Hell**	**Hell**
15	**Let**	Cannot	Am
16	**Soul**	**Soul**	**Fear**
17	Is	Do	Will
18	**Night**	**Burn**	**God**
19	Not	**Fear**	I
20	**Die**	Is	**Decay**
21	**Mazes**	All	You
22	**Wheels**	**Take**	**Forever**
23	No	**Dead**	**Now**
24	**Know**	**Evil**	**Mind**
25	**Fly**	**Pain**	Me
26	**Cry**	**Insane**	**Mortal**
27	Going	**Fight**	**Grave**
28	**Just**	**Fucking**	**Rotting**
29	**Forever**	**Eyes**	**Darkness**
30	Ain	**Lies**	**Lust**

Only 11 occurrences of *will* in HMSC are as nouns or verbs whereas 341 are future auxiliary. Similarly, *got* occurs 167 times in HMSC and 36 of them are in *got to*, 35 *have got*, 16 *has got*, 35 dropped versions of *I have got* (*I got*), and 20 dropped versions of *you have got* (*you got*). *Do* occurs 258 times in HMSC and only 46 occurrences are lexical verbs. Finally, out of the 103 occurrences of *going* in HMSC, 83 appear in the future tense auxiliary *going to*.

Will occurs as lexical verbs or nouns 11 times out of a total of 351 tokens in TMSC. 59 occurrences of *do* are lexical verbs whereas the remaining 159 are auxiliary. The top-30 keywords in DMSC only contain *will* as a potential intermediate case and it occurs 31 times as noun/lexical verb and 254 times as future auxiliary. Given these few occurrences of the said words as content words, it is considered rational to exclude them as key content words.

Table 44: Key content words in top-30 keywords by subcorpus.

	HMSC	**TMSC**	**DMSC**
1	Love	Blood	Death
2	Never	Death	Flesh
3	Let	**Die**	Blood
4	**Soul**	Kill	**Die**
5	Night	Life	Evil
6	**Die**	Hell	Satan
7	Mazes	**Soul**	Life
8	Wheels	Burn	Dead
9	Know	Fear	**Soul**
10	Fly	Take	Pain
11	Cry	Dead	Kill
12	Just	Evil	Hell
13	Forever	Pain	Fear
14		Insane	God
15		Fight	Decay
16		Fucking	Forever
17		Eyes	Now
18		Lies	Mind
19			Mortal
20			Grave
21			Rotting
22			Darkness
23			Lust

Table 44 shows the key content words in the top-30 keywords in each subcorpus. As noted earlier, HMSC has 13 content words while TMSC has 18 and DMSC 23. Only two keywords are shared across all corpora: *soul* and *die*. *Soul* is more key than *die* in HMSC; however, it is vice-versa with TMSC and DMSC. A detailed investigation of these keywords will be provided in the discussion through concordancing.

5. DISCUSSION

The aim of the study was to compare three metal subgenres, namely, heavy metal, thrash metal and death metal in terms of their lexical richness and keyness properties. The methodology was selected to achieve quantitative results with the lexical richness analyses and qualitative ones with keyness. The research stands between linguistics and music studies and is a pioneering one in the sense that it features a methodology that has never been applied to metal music – or any other music genre for that matter. There were no established theories or fully comparable existing results for the music genres covered in the study to base the hypotheses, either. There were four research questions for the present study:

1. Are there any differences between the lexical richness levels of metal subgenres?
    a. Are there any differences between lexical density levels of metal subgenres?
    b. Are there any differences between lexical sophistication levels of metal subgenres?
    c. Are there any differences between lexical variation levels of metal subgenres?
2. Are there any differences between the keywords of heavy metal, thrash metal and death metal and if any, how do they relate to the respective subgenre?
3. Can lexical richness and keyness analyses be used in music genre comparison?
4. To what extent do the existing categorizations overlap with the lexical categorization?

The hypotheses for these questions were that in both of the analyses heavy metal would be the least lexically rich subgenre and death metal the richest one. Thrash metal was expected to be placed in the middle. Results show that the hypotheses hold for all analyses. The immediate answers to the questions are given below which are followed by in-depth interpretation.

Answer 1: All subgenres have been found to be statistically significantly different from each other in all aspects of lexical richness as measured by the selected methods. The richness levels were the lowest for heavy metal and highest for death metal.

Answer 2: All subgenres have been found to contain different sets of keywords. The keywords in heavy metal are more different from the other two genres in a way that suggests a support for Weinstein's (2000) theory of Dionysian themes in heavy metal. Unlike heavy metal, thrash metal and death metal displayed darker, more violent and more religion-related keywords which can be deemed Chaotic.

Answer 3: Both analyses were capable of drawing the borders between the subgenres at a lexical level. Hence, the selected methods could be used in (sub)genre comparison. The fact that all the lexical richness measures were positively correlated with each other provides support to the utilization of the methods together or interchangeably.

Answer 4: From a musical point of view, death metal is the heaviest and most brutal subgenre among those that have been studied here. Thrash metal is the second heaviest and heavy metal is the least heavy subgenre despite its name. The lexical analyses portrayed the very same picture in terms of grading the subgenres by placing death metal at the highest level of lexical richness and thrash metal the second. Although a qualitative analysis, keyness results also showed a gradual increase of heaviness and violence from heavy metal to death metal. Analyzed in a bottom-up manner, the divisions obtained from lexical analysis do not fully overlap with the existing classification.

Methodology-wise, the MC has enough songs to draw significant conclusions and the diversity of bands adds to the representativeness of the corpus. As indicated earlier in Table 6, the present study features a higher number of songs than many others conducted in similar topics. The treatment of lyrics was already discussed in Compiling and Processing the Data. It needs to be highlighted that removing the repeats was helpful in not only in equalizing song lengths but also achieving more reliable results in LV analyses which depend on the number of different words.

Biber et al. (1998: 136-7) claim that register studies which focus on few linguistic elements are hardly likely to generate comprehensive descriptions. In this respect, the study sought to include as many dimensions of lexical analysis as possible. LS was operationalized in five ways (GSL B2K tokens, GSL B2K types, BNC/COCA B2K tokens, BNC/COCA B2K types and P_Lex), LV in four (TTR, Guiraud, Uber and HD-D) and LD as content words/total running words. The number and variety of methods applied show similarities to other lexical richness studies (Daller and Xue 2007, Daller and Phelan 2007, Van Hout and Vermeer 2007, Šišková 2012, Zheng 2016, etc.).

Keyness analysis was added to the study in order to have a qualitative side and compare the results with each other. This was done to maintain a balance between the analyses as proposed by Meyer (2002) and Seidlhofer (2012). In fact, LS measures are of qualitative nature (Daller et al. 2003: 203) in addition to the data and ratios it provides as they can also show the distribution of the most frequent words. In this respect, LS measures are good examples of a combination of qualitative and quantitative methodology.

Of course, more analyses could be carried out at phrase or clause levels which include lexico-grammatical features. This, in turn, leads to an MDA-like analysis which is quite complex and time consuming. This study showed that a comprehensive lexical richness and a keyness analysis are able to generate tangible and significant results in the distinguishing of music genres.

5.1. Lexical Variation

The three subcorpora in the study have been found to be statistically different from each other with regard to their LV values. A summary of the results obtained from four different measures of LV is provided in Table 45.

Table 45: Summary of the LV analyses.

	HMSC	TMSC	DMSC
TTR	.54	.61	.66
U	19.05	23.93	28.35
HD-D	32.07	34.18	34.92
Guiraud	6.75	7.88	8.35

The ranking of the subgenres from the least to most lexically diverse is HMSC, TMSC and DMSC, regardless of which measurement method have been used. Although severely criticized for its sensitivity to word length, TTR was capable of yielding significant results on a par with more advanced methods of calculation. Murphey's (1992: 771) study reported an average TTR value of 0.29. This is far too low compared to metal songs as the lowest TTR value observed in the present study is 0.54. This is parallel to the hypothesis of the study that the heavier the music, the more lexically rich. Kuiken et al. (2005) report that the tasks produced in Italian as a second language had TTR levels of 0.5 and 0.51 for more and less complex tasks, which are close to but still lower than HMSC values. Ishikawa's (2007: 125) and Kuiken and Vedder's (2008) findings are similar for second language learners of French and Italian. Broeder et al. (1987 qtd. in Daller et al. 2007: 126) report TTR levels varying between .24 and .35 for Swedish L2 speakers of English in film retelling and free conversation modes. Metal songs seem to surpass these values with a range of .54 and .66. The same study employed Guiraud as well and the results were between 5.89 and 7.19. In this respect, only HMSC seems comparable to the above mentioned modes; however, it is observed that TMSC and DMSC are more lexically diverse than film retelling and free conversation. Of course, the comparison cannot be completely meaningful as the reference study contains only 20 texts and carried out in an L2 setting. Verspoor et al.'s (2012: 252) study on Dutch EFL learners' essays at various levels of competence reports Guiraud values of 6.0 for the highest level pupils. This value falls short of HMSC; nevertheless, the L2 factor needs to be taken into consideration.

Daller and Xue's (2007: 159) findings indicate *D* scores of 28.59 for Chinese students living in China and 36.22 for those living in the UK. The UK scores are higher than all the subgenres in present study. This is not likely to be the real case, though. The measurement methods seem to have an effect on the *D* results in a way that produces more difference than the other methods. For example, the Guiraud scores from the same study are 5.03 and 6.18, which are lower than all the subcorpora in the present study. In terms of TTR, they found 0.35 and 0.39, lower than the present findings by large. This comparison is parallel to the others in the literature.

Lu et al. (2014) compared American elementary school textbooks and one of the measures they used was *D*. They compared the results across ten decades and the *D* results were between 64.3 and 74.8. These results are comparable to the present study as they are in L1 and *D* method is claimed not be sensitive to text length. Yang's (2014: 84) study reports similar values ranging between 66.64 and 73.5. Yoon and Polio's (2017: 139) findings indicate *D* values of 78.75 and 79.83 for narrative and argumentative essays by the EFL learners at the highest level. Similarly, Révész et al.'s (2019: 230) research yielded *D* values of 68.74 for a simple task and 67.97 for a complex one. As evident in the examples, there is massive difference between the values as the ones in the present study range between 32.07 and 34.92. The reason for this is the fact that the *D* values as measured by HD-D are lower than those by *Vocd*. This difference is noted in Fergadiotis et al. (2013: 401) where the same texts yielded the HD-D and *Vocd* scores of -7.78 and 31.55 respectively. Although most studies are carried out with *Vocd* and results cannot be fully compared to the present study, studies show positive correlations between HD-D and *Vocd* scores (McCarthy and Jarvis 2010, Šišková 2012, Fergadiotis et al. 2013). Fergadiotis et al. (2013: 405) report the relationship between HD-D and *D* as "close-to-perfect". However, the fact that *D* results vary across different studies and as a result of different measurement methods and tools make it a less reliable method for comparison.

In terms of Uber results, Kormos and Dörnyei (2004) analyzed the argumentative essays written by Hungarian EFL students at intermediate level and reported a mean Uber value of 14.79 (p. 6). The Uber values for the MC were 19.05 for HMSC, 23.93 for TMSC and 28.35 for DMSC. In this regard, even HMSC score is higher than that which was achieved by EFL learners. The Uber scores for academic texts produced by Chinese university students varied between 22.71 and 27.16 (Zheng 2016: 45). These scores are higher than HMSC and TMSC and close to DMSC. Zheng's corpus consisted of 58,645 words, around 60% of the present study and it was conducted in an EFL setting. Still, it can be claimed that academic texts have high Uber values and TMSC and DMSC are no less than them in terms of lexical variation. American students at fifth, seventh and ninth grades scored Uber values of 13.95, 12.94 and 14.24 respectively (Jarvis 2002: 68) which are very low compared to the MC. Lissón and Ballier's (2018)

Uber scores of the texts written by Spanish L3 learners of French ranged between 20.06 and 23.08 (p. 25). These scores are slightly higher than that of HMSC but lower than the other two subcorpora.

5.2. Lexical Sophistication

The results of the LFP analysis showed significant differences across all subcorpora as well as the LV analyses. Pairwise comparisons also indicated significant differences between each one of the subgenres. The discussion of the results will start with the interpretation of Table 46, a simplified version of the results table for the LFP analyses.

Table 46: Summary of LFP B2K token and type ratio distribution in percentages based on the GSL and the BNC/COCA.

		HMSC	TMSC	DMSC
Tokens	GSL	7.19	12.71	20.46
	BNC/COCA	6.83	12.11	20.21
Types	GSL	10.12	18.02	27.22
	BNC/COCA	9.50	17.18	27.00

The initial intuitive hypothesis that heavy metal is the least lexically rich of the three subgenres proved true in this analysis as HMSC achieved the lowest scores in LS. In other words, HMSC contains the most high frequency words across all subcorpora, containing only 7.19% of B2K tokens in the GSL and 6.83% in the BNC/COCA. In terms of types, HMSC has a coverage of 10.12% GSL and 9.50% BNC/COCA B2K types. These results are indicative of relative scarcity of low frequency words in heavy metal lyrics.

TMSC, as anticipated, came as the second subcorpus after HMSC with regard to lexical frequency. It contains ~12% of B2K tokens in both frequency lists (the GSL: 12.71% and the BNC/COCA: 12.11%). The results for types are very close to each other as well (the GSL: 18.02% and the BNC/COCA: 17.18%). These scores show a higher ratio of low frequency words compared to HMSC. These findings might indicate that heavy metal is less lexically rich than thrash metal as far as lexical frequency is concerned. In other words, thrash metal songs feature more low frequency words compared to heavy metal.

DMSC, the heaviest and most brutal subcorpus in the MC in terms of musicality and semantic value of its lyrics, was found, indeed, to have the highest lexical richness level album in the corpus with regard to LFP. Its total proportion of B2K tokens is 20.46% in the GSL and 20.21% in the BNC/COCA, which are the highest values in the MC. DMSC token scores are almost three times higher than HMSC and two times higher than TMSC. The results are similar in types with values of 27.22% in the GSL and 27% in the BNC/COCA. These values are strong indicators of the use of a very low frequency vocabulary in death metal lyrics.

The results of the lexical frequency analysis showed that HMSC was less lexically rich in terms of lexical frequency than the other two subcorpora. It could be argued that, frequency-wise thrash metal sits in the middle between heavy metal and death metal with a little inclination towards the former. HMSC has the highest ratio of high frequency words whereas DMSC the lowest. DMSC contains many words in the lower frequency bands. It appears that thrash metal is more or less the average subgenre in the present study in terms of lexical frequency.

These results neither support nor refute the theory that “death metal was created out of thrash metal” (2000: 17). It must be noted, however, that these genre definitions are mostly based on musical traits. Lyrically, heavy metal does not necessarily deal with a narrow set of themes unlike thrash and death metal which use the themes of violence, death, madness, injury, death, anger dominance, etc. (Walser 1993, Buts and Buelens 2008, Weinstein 2000, Arnett 1996, Stelzner, Frandsen 2011). The fact that these themes and associated words are not frequent seems to be a key factor in the obtained results. The ranking of the three genres in the present study could be regarded as a quantitative support for Moore’s (2013) Death Metal English theory.

Zheng’s (2016: 45) results indicate B2K values of 10.05-12.03% for EFL learners. The study was based on the GSL and in this respect it could be claimed that the EFL students’ lexical sophistication levels are parallel to those reflected in HMSC. Kojima and Yamashita’s (2014) study indicated a B2K level of 6.77 for ESL essays. They used the BNC as a reference list and the results are quite similar to HMSC; however, lower than TMSC and DMSC. Daller and Xue (2007: 159) found GSL B2K levels of 10.96 and 16.38 for China and the UK groups, respectively; however, it is not stated in the study whether these are token or type results. Therefore, a comparison is not possible to the present study. Nevertheless, in any case, all values are lower than DMSC.

The lexical frequency analysis method has not been applied to metal songs before but there are a few examples related with pop and rock songs. The results will be discussed mainly in comparison to Meara (1993), Sundberg (2015) and Öztürk (2017) on the basis of corpora and obtained values. Taina's (2017) study is, in fact, quite similar to the present one as he looked at the lexical frequency of metal lyrics. Nonetheless, his frequency analysis is limited to the identification of the most frequent words and does not contain the distribution of the words based on a frequency list.

Excluding Taina (2014), corpus-wise, the present study differs from these three in that it takes a comparative approach across music genres rather than comparing song lyrics to other texts or establishing a corpus of randomly selected songs based on their frequencies. Meara (1993) compared only one album to other types of texts. His analysis was a comparative one, yet not among the songs but in a fashion comparing different broadcast types. Sundberg (2015) used songs – not albums. This and the fact that the songs in the corpus were in French makes his study quite different, however, the measurement method is similar.

Öztürk (2017) did not clearly divide the genres in her comparison but rather made her selection based on her personal music taste. She selected four artists and included more than one album by each artist. In this respect, the present study could be considered more comprehensive as the corpus it has made use of contains 600 songs whereas Öztürk's (2017) study had 177. She also compared the albums by the same artist; yet, this is not comparable to the present research as each subcorpus has 10 songs by each band.

She included the first two albums by *Adele*, who released her third and latest album *25* in 2015 (it may be possible that this album had not been released by the time she conducted her analyses). *One Direction* was represented with their first four studio albums out of five, the latest of which was released in 2015. The time of analysis again might be the reason for this selection. *Taylor Swift* has six full-length albums and Öztürk (2017) analyzed only the first four releases between 2006 and 2012. Finally, *Bon Jovi*, the oldest band in her corpus, was represented with four albums released between 2005 and 2013. Bon Jovi has a full-length discography of 14 studio albums as of 2020 and his earlier works are more rock/metal than the later albums, which may be the reason for selection.

In terms of obtained values, Meara's (1993) analysis of *Chris de Burgh*'s *Into the Light* showed a K2 level higher than 90%. His scores were obtained based on Nation's (1986 qtd. in Meara 1993) word list and the findings suggest that the song lyrics represented the highest level of frequency (most simple vocabulary) across all BBC texts they studied. However, his analysis was based on word types. As the details of the study are not mentioned in the paper, *Chris de Burgh*'s *Into the Light* album was analyzed so that it could be compared to the findings of the present study and the results are given in Table 47. The album has 12 songs and the lyrics to each song were obtained from www.lyricsdepot.com. There were a total of 2,863 words prior to the processing of the lyrics in the same way as the lyrics in the MC. The final version features 2,335 words. These numbers, in fact, do not comply with Meara's (1993) analysis as he notes 1,500 words as an approximate number of tokens in the texts taken from "a set of songs" from the *Into the Light* album (p. 12). It shows that he may not have analyzed all the songs in the album.

Table 47: GSL and BNC/COCA frequency distribution for *Chris de Burgh - Into the Light*.

		TOKENS %	TYPES %
GSL	K1+K2	95.54	88.07
	B2K	4.46	11.93
BNC/COCA	K1+K2	95.54	89.51
	B2K	4.46	10.49

The re-run analysis of the album shows K2 type coverage of 89.51% in the BNC/COCA, corresponding to a B2K score of 10.49%. In terms of the GSL, the type scores for K2 frequency band is 88.07% (B2K=11.93%). These results are quite different from what Meara (1993) reported (over 90%) but it could be because not all songs were included and/or the lyrics were treated differently. In any case, *Chris de Burgh*'s *Into the Light* album is almost on a par with HMSC with respect to lexical frequency level.

However, this deduction may be far-fetched as it is based on a comparison of several albums against one. Another difference between the studies is that Meara (1993) looked at the ratio of types – not tokens. Analyzed in tokens, the scores for *Into the Light* amount to 95.54% in both the GSL and the BNC/COCA K2, corresponding to a B2K ratio of 4.46%. These results could be compared with the subcorpora in MC in terms of token percentages. In this regard, *Into the Light* has a lower score of B2K words against metal songs even as compared to HMSC (the GSL=7.19% and the BNC/COCA=6.83%). It can be concluded from the token results that the lyrics in heavy metal songs contain lower frequency words than pop songs. This finding could be interpreted as a support to the hypothesis of the present study that 'the heavier the music, the more lexically rich the lyrics'. Nevertheless, a pop song corpus of equal size is needed in order to prove or refute this claim.

The fact that different interpretations arise when approached from the viewpoints of types and tokens could lie in the possibility that the lyrics of *Into the Light* featured many in-line repetitions which could not have been removed without threatening the originality of the texts. Although this could be an effect, it must not be overlooked that, it is not likely to obtain tangible results through the comparison of a specialized corpus with only one album, regardless of whichever lyrics treatment method has been applied.

The results of the present study show a higher level of complexity compared to those of Sundberg (2015). His corpus of 23 French songs has an average K2 lemma coverage of 92%, corresponding to a B2K ratio of 8%. These scores could only be comparable to HMSC. However, this will not be a valid comparison as 1) the measurement parameters are different (lemmas vs. tokens and types), 2) the languages are different and 3) Sundberg's (2015) results are based on 23 songs while the present study is 600. Additionally, Sundberg (2015) deliberately chose easy songs since his aim was to teach French as second language. Therefore, the results cannot be fully comparable, but could be used as reference as to which songs qualify as 'easy' or 'hard' depending on their frequency levels. Taking these facts into consideration, these results could only hint at the closeness of the lexical frequency profile of French pop songs to heavy metal songs in general.

Öztürk's (2017) findings suggest a sizeable cumulative ratio of words at the BNC/COCA K2 level compared to what has been observed in metal songs. Her results were 96.67% for *Adele*, 96.23% for *Bon Jovi*, 98.34% for *One Direction* and 96.54% for *Taylor Swift* songs. These values correspond to B2K levels of 3.33%, 3.77%, 1.66% and 3.46% respectively. On the other hand, the lowest BNC/COCA B2K ratio achieved in the present study was 6.83% for HMSC. These results show an ostensible difference between pop/rock and metal songs. Although heavy metal songs seem to have relatively close results with pop songs, the difference increases with TMSC and DMSC. This finding is also indicative of a support to the hypothesis that heavier music posits lexically richer lyrics. The observable difference between the metal and the pop lyrics probably stem from the characteristics of the text types. Pop song lyrics feature simple and conversation-like language (Murphey 1992: 771) while metal lyrics deal with Dionysian and Chaotic themes (Weinstein 2000) which seems to result in higher level of rare words, i.e., higher frequency.

All three studies (Meara 1993, Sundberg 2015 and Öztürk 2017) mentioned here are the closest ones to the present study in terms of lexical frequency analysis and the overall results show that metal songs are lexically richer than pop songs. The comparison of the frequency results of all the three studies are provided in Table 48 below.

Table 48: B2K percentages of the MC as compared to the results of Meara (1993), Sundberg (2015) and Öztürk (2017)

Measurement Method	**Frequency List**	**Meara (1993)***	**Sundberg (2015)**	**Öztürk (2017)**				**The Present Study**		
		Chris de Burgh – *Into the Light*	**23 French songs**	**Adele**	**Bon Jovi**	**One Direction**	**Taylor Swift**	**HMSC**	**TMSC**	**DMSC**
Tokens	GSL	4.46%	N/A	N/A	N/A	N/A	N/A	7.19%	12.71%	20.46%
	BNC/COCA	4.46%	N/A	3.33%	3.77%	1.66%	3.46%	6.83%	12.11%	20.21%
Types	GSL	11.93%	N/A	N/A	N/A	N/A	N/A	10.12%	18.02%	27.22%
	BNC/COCA	10.49%	N/A	N/A	N/A	N/A	N/A	9.50%	17.18%	27.00%
Lemmas	N/A	N/A	8	N/A	N/A	N/A	N/A	N/A	N/A	N/A

* The results of the re-run analyses.

Firstly, it needs to be noted that the differences between scores in the mentioned studies and the present one have not been tested for their statistical significance as the data is not available, and, where applicable, not enough for analysis. Therefore, the discussion will be limited to observations. As noted, the results of the metal subgenres display a difference from the other artists with the exception of HMSC, which still has higher scores than other artists. It is interesting to note that there is considerable difference between TMSC and DMSC and the others. TMSC and DMSC have the highest ratios of B2K tokens and types. The comparison to French songs may not be reliable as they are obtained from another language based on another criterion. The results of the GSL in the present study are not comparable to other studies mentioned as it has not been used.

Simonton (1990) and Forsyth's (2000) findings that more popular or "better" sonnets/poems contain fewer rare words, on the other hand, cannot be comparable to the present study since the present study does not a) compare songs but albums, b) compare the works of the same artist, c) observe a method of ranking the bands/albums as best/worst or more/less popular, but as 'less heavy' or 'heavier'. For the sake of an attempt to find parallelism to their studies, it could be stated that the relatively more popular subgenre (heavy metal) uses fewer words, and the relatively less popular subgenre (death metal) features the most. It must be noted that this is a far-fetched argument which is not among the aims or the procedures of the present study and it could only be verified through appropriate methodology.

Finally, the lexical frequency profile (LFP) is intended for L2 use, yet the present research applied it to L1 speakers. Therefore, it does not test the language capabilities of the artists but compare their lexical choices. As Laufer (2005: 583) puts it, LFP takes into account the vocabulary use – not size. Swales' (1990) notion of genre fits here since artists usually produce "prototypical" lyrics in a music genre reflecting its characteristics, which both constitute and are shaped by their "discourse communities" in a mutual way. In other words, artist may choose to use high or low frequency words to fulfill the 'requirements' to belong to a particular genre, or, in fewer cases, one artist's usage of particular vocabulary may lead to the creation of a new genre, which, in turn, may pave the way for the following artists to write their lyrics in a similar fashion. Therefore, LFP might indeed be a useful method of measuring lexical richness of L1 speakers as well if the aim is to compare different genres. Meara (1993) and Öztürk (2017) also compared L1 bands. *Chris de Burgh*, *One Direction* and *Adele* are British, and *Taylor Swift* and *Bon Jovi* are American artists. Sundberg's (2015) collection of songs consists of artists who are French L1 speakers from France, Canada and Belgium as well as one English artist performing in French (L2). In this respect, contrary to the other papers and the present study, Sundberg's (2015) corpus does not exclusively contain L1 artists.

The BNC/COCA vs. The GSL

The present study used two different major word frequency lists for the comparison of metal subgenres – the BNC/COCA and the GSL. Although the former is more recent than the latter by approximately 50 years, their coverage is not so different from each other. Gilner (2011: 75) states that the BNC 2,000 and the GSL overlap to a great extent. Nation (2004) found similar results between the BNC 3,000 (K3) and the GSL+AWL. The present study puts forward a comparison of both lists using the same corpus.

As indicated in Table 46, the results are quite close to each other differing only by ~0.5%. This indicates that the GSL is still capable of offering sufficient coverage almost on a par with the BNC/COCA. Results obtained from both lists are highly correlated to each other. It could mean that both lists could be used to compare genres – song lyrics, in particular. If the aim is to find the distribution of words across specific frequency bands, the BNC/COCA could be preferred as it offers a detailed distribution of low frequency words with the broader spectrum of 25 bands as opposed to three of the GSL+AWL.

It is also possible to compare the MC frequency distribution results with the ideal outputs of the lists that have been used in the study. The typical output distribution for *Vocabprofile* is listed as 70-10-10-10 percent for K1, K2, AWL and Off-List tokens respectively (Research Uses of *Vocabprofile*). The results do not show much parallelism to the ideal output, nevertheless. As the study is based on B2K ratios, a detailed distribution of K1 and K2 bands have not been conducted. However, compared in B2K perspective, only DMSC (20.46% GSL

B2K tokens) seems to fit the distribution as the ideal output leaves a share of 20% for B2K tokens. As for the other two subcorpora, the B2K levels are remarkably lower than the typical output for tokens. The reason for the increased level of B2K words might be attributed to colloquial terms, contractions and proper nouns used in the songs as well as specialized vocabulary which are typical of death metal in themes such as, death, torture, gore, violence, etc. Some examples of Off-List words from DMSC are *abhorrent, cadaverous, disconsolance, fathomless, triturated*, etc.

Nation (2016: 135-7) checked the frequency band validity of the BNC/COCA in an analysis of the Wellington Written Corpus and the output of the first two thousand bands are 75.22% and 8.91%, amounting to coverage of 84.13% which corresponds to a B2K ratio of 15.87%. This is comparable to TMSC (12.11%) but lower than DMSC (20.21%). This indicates that the Wellington Written Corpus has fewer lower frequency words than DMSC. Compared to HMSC and TMSC, this could be regarded as an expected result as written language is typified as more formal and rich in low frequency words contrary to spoken language or song lyrics which are argued to have both written and spoken characteristics (Kreyer and Mukherjee 2007: 37, Falk 2012: 21). DMSC, however, contains more sophisticated vocabulary than written texts.

Criado and Sanchez (2012) compared two English textbooks using LFP with the BNC wordlist. Their K1+K2 results were 91.61% tokens and 62.66% types for the first book and 90.68% tokens and 60.75% types for the second one. They correspond to the respective B2K values of 8.39%, 37.34%, 9.32% and 39.25%. These results indicate big differences in themselves between token and type values. The reason for this is that course books involve repetitions to a high extent and this is also acknowledged by the authors as they state that "lexical repetitive practice is abundant in both textbooks" (p. 90). In terms of tokens, the course book results are only close to HMSC. However, a different picture is visible in terms of types as the values for the course books are higher than any subgenre in the MC.

As seen in this comparison, it may not be a good way to benchmark the performance of a frequency list on lyrics against results obtained from the analysis other text types and registers/genres. It is easy and even necessary to omit the repeats in a song whereas it cannot be applicable to books. Therefore, the comparison of results in types will be inevitable skewed. It could be a better method to compare different genres based on token values, especially when it is not certain whether any kind of repetition removal has been made or not.

As a final comparison, Webb and Rodgers (2009a) analyzed 318 movies from different genres and they found the following K1+K2 levels for each genre as shown in Table 49:

Table 49: BNC K1+K2 and B2K token percentages across movie genres (adapted from Webb and Rodgers (2009a: 419)

Movie Genre	BNC K1+K2	BNC B2K
Animated	91.70%	8.30%
War	91.85%	8.15%
Action	92.85%	7.15%
Science fiction	93.24%	6.76%
Comedy	93.99%	6.01%
Crime	94.26%	5.74%
Romance	94.36%	5.64%
Westerns	94.36%	5.64%
Classic	94.43%	5.57%
Drama	94.52%	5.48%
Horror	94.56%	5.44%

The movie genres in Table 49 is sorted by decreasing lexical richness and interestingly, animated movies are the ones which have fewer words in K1+K2 level, thus having the highest percentage in B2K. However, the differences across genres are very little, ranging only by 2.86% between the highest (8.3%) and lowest (5.44%) values. They did not report a statistical significance result and it is highly unlikely to reach it given these values. In their next study, they reported the following BNC B2K token values (subtracted from the original K1+K2 values) for American TV shows and movies: 9.3% (news), 7.9% (science fiction), 6.38% (drama), 5.58% (older), 5.22% (situation comedies) and 4.87% (children's) (Webb and Rodgers 2009b: 350). The difference here is 4.16%. In this respect, it is interesting to note that metal genres (even within the same supra-genre) can vary more than movies can. Animated, war and action movies seem to contain similar proportions of B2K words to HMSC and this could be the only overlapping point across movies and MC. TMSC and DMSC are far richer in B2K tokens than movies.

Although not as widely used as LFP, P_Lex is an alternative method of measuring lexical sophistication. It has been included in the study to have a second method, which could serve as another perspective apart from that of LFP results. The results, as well as the LFP, show that the three subgenres differ significantly from each other. TMSC is again in the middle of the spectrum with a lambda value of 2.36, HMSC on the lower end (1.53) and DMSC on the higher (3.2). The difference between HMSC and DMSC is twofold, suggesting that DMSC makes use of twice as many less frequent words as HMSC. This is again an indication of lexical sophistication increasing as the heaviness of the music genre increases.

Skehan and Foster (2012) used P_Lex for the comparison of different tasks across NSs and NNSs. Their findings with NSs showed that the lambda scores of planned and unplanned speech varied between 1.13 and 1.49 (p. 206). According to these results, HMSC, the lowest scoring metal subgenre, has a higher value than the planned narrative speech with the highest value. In this sense, the song lyrics display a more written trait – although they are written to be sung – than planned speech. In a NNS context, Daller and Xue (2007) measured texts written by two groups of Chinese students one of which lived in the UK and the other in China. The lambda results for the texts were 0.23 for the UK group and 0.16 for the China. Although not comparable to the findings of the present study, the notable difference might be indicative of the high lexical sophistication level of metal song lyrics.

Meara and Bell's (2001) results for two each discursive essays written by 49 NNS students were 1.3 and 1.46. Malvern et al. (2004) compared the lexical sophistication levels (among other parameters of lexical development) of the UK schoolchildren at the ages of 7, 11 and 14 using P_Lex. The study was carried out on narrative texts produced by the students and the results were – in the ascending order by age – 0.93, 1.18 and 1.29. Again, as is the case with Daller and Xue's (2007) and Meara and Bell's (2001) studies, this is not fully comparable to the present study as the groups are made up of very young children. This fact makes language development an important factor even though they are all NSs. All things considered, it could be inferred from the comparison of the findings of the present study with the ones mentioned above that metal song lyrics display a higher level of lexical sophistication than written essays and speech.

The Most Frequent Words across the Subcorpora

Following the detailed results and discussion of the frequency levels, a look at the most frequent words in each subcorpora might provide a qualitative insight regarding the similarities and the differences of the metal subgenres. Table 50 shows the list of the most frequent words in each subcorpus.

Table 50: The most frequent ten words in each subcorpus.

	HMSC	**TMSC**	**DMSC**
1	the	the	the
2	I	you	of
3	you	to	to
4	to	of	I
5	is	I	and
6	a	and	a
7	and	a	in
8	of	is	your
9	it	in	is
10	in	your	you

As expected from any text, all lists are dominated by function words. The most frequent word is *the. a* is either on the 6th or the 7th place. The rest are pronouns, auxiliaries, prepositions and the conjunction *and*. What is important here is that the most frequently used pronouns are *I* and *you*. This finding complies with Murphey's (1992: 771-2) that pop songs are conversation-like and full of *I* and *you* referents. Although the genres vastly differ from each other, the personal referencing seems to be the same. HMSC features more use of *I* than *you* but their occurrences are quite close to each other. This might be indicative of more personal and one-to-one relationships. On the other hand, TMSC has a higher use of *you* than *I*, which might explain the aggressive nature of the subgenre as this finding might tend to reflect the targeting characteristic of the genre. Thrash metal is known for its critical and anti-social lyrics; therefore, more use of *you* could be associated with accusation and blaming the evil on the society and desire to hurt the other. Different from both of these subcorpora, DMSC is heavily dominated with *I* and relatively fewer *you* and *your*. Given the fact that death metal lyrics are about death, killing, torture, etc., it could be deemed normal that the use of the first person referents points to an active mode of the evil deeds sung in a way as if they are done by the singers/bands themselves.

Table 51: The most frequent ten words across metal subgenres in Taina (2014).

	Traditional Heavy Metal	**Thrash Metal**	**Death Metal**
1	the	the	the
2	I	to	of
3	you	you	to
4	to	of	I
5	and	I	in
6	a	and	and
7	of	a	you
8	it	is	my
9	s	in	a
10	in	me	your

As seen in Table 51, these results correspond to Taina's (2014: 44) findings although his selection of bands is somewhat debatable. Looking at the content words might shed more light on the lyrics. Table 52 shows the most frequent ten content words in each subcorpus.

Table 52: The most frequent ten content words in each subcorpus.

	HMSC	**TMSC**	**DMSC**
1	know	life	death
2	see	death	life
3	time	take	blood
4	night	see	dead
5	love	blood	time
6	way	time	evil
7	get	like	die
8	like	world	pain
9	let	die	flesh
10	want	come	mind

A brief look at Table 52 shows that HMSC has relatively more positive words than TMSC and DMSC. It has more abstract nouns, expressions of positive emotions such as *love*, *like* and *want* and possible sexual indications judging by *love*, *want* and *night*. This outcome is compliant with Weinstein's division of metal lyrics into two as Dionysian and Chaotic. Dionysian themes, which are mostly associated with sex, drugs, alcohol, etc., are quite evident in HMSC. TMSC and DMSC, however, deal with the darker side of life through a wide use of words such as *death*, *blood*, *die*, etc., which places these two subgenres into the Chaotic division. The most frequent words in TMSC and DMSC are *life* and *death* and surprisingly *death* is more frequent in thrash metal than death metal – it is even the most frequent one. TMSC portrays a dark and destructive picture through preference of negative words such as *fear*, *die*, *death*, *blood* and DMSC takes it to the next level with the inclusion of words that recall infliction of pain, suffering and bodily abuse such as *blood*, *pain* and *flesh*. The frequent use of the word *evil* in DMSC can also be attributed to the more Chaotic nature of death metal.

Table 53: The most frequent ten content words across metal subgenres in Taina (2014).

	Traditional Heavy Metal	**Thrash Metal**	**Death Metal**
1	got	death	see
2	know	time	soul
3	life	bring	eyes
4	way	life	life
5	come	die	death
6	love	blood	feel
7	see	see	pain
8	go	fear	world
9	world	fire	come
10	take	world	god

Taina's (2014) corpus contains content words of similar nature as shown in Table 53. Heavy metal lyrics reflect a Dionysian perspective and thrash and death metal converge to the Chaotic themes. Words depicting physical and mental disturbance (*feel*, *pain*) are found in death metal as they have been found so in the present study.

As mentioned earlier, Taina's (2014) selection of bands is quite different from the approach in the current study. His list contains non-native speaking bands as well as those which are not necessarily assigned to the correct subgenre. This issue has been mentioned earlier in MC, HMSC, TMSC and DMSC. For example, *Mercenary* is in the traditional heavy metal subcorpus although the bands plays melodic death metal. The same subcorpus also features songs by *Black Sabbath* and *Ozzy Osbourne* as well as *Judas Priest* and *Halford*. *Ozzy Osbourne* is the singer for *Black Sabbath* and *Halford* for *Judas Priest*; hence, the diversity of the artists decreases. It may be argued that they write in different themes in their bands and solo projects and it might hold true. However, in order to prevent a potential stylistic overlap, side projects of the main bands have not been included in the MC.

Despite the relatively small size of Taina's (2014) corpus and his questionable methodology in building it, the frequency results have been found to be highly replicable and comparable to the present study. It must be noted that the methodology in the present study cannot be regarded as superior to Taina's (2014) as it has its own limitations which will be discussed at length in the *Limitations* section.

5.3. Lexical Density

The ranking of the subgenres after the LD analysis was realized as HMSC, TMSC and DMSC, from the least to most lexically dense. This finding is parallel with the hypothesis. The reason for this finding could lie in the theory that thrash metal and death metal use more Chaotic lyrics (Weinstein 2000) compared to heavy metal which is more inclined to opt for Dionysian themes. From this perspective, it could be asserted that song lyrics written in Chaotic themes are more lexically dense than those written in Dionysian. The results also showed that thrash metal is almost equally distant to the other two subgenres at the opposite ends of the spectrum of extremity – and lexical density as well. This is compliant with the theories that thrash metal is a "punked-up" version of heavy metal (Weinstein 2000: 288) and that death metal is created out of thrash metal (Kahn-Harris 2000: 17). It appears that the extremity in metal does not only manifest itself in music and imagery but also the lexical density of their lyrics.

Stubbs (1996 qtd. in Camiciottoli 2007: 73) found that the LD levels of fiction tests varied between 40-54% and non-fiction between 40-65%. In this regard, death metal lyrics could be on a par with non-fiction texts. Camiciottoli's (2007) findings indicate that business lectures has an LD level of 44% as compared to multidisciplinary studies (35%). The LD results of the present study in percentages are 49% for HMSC, 53% for TMSC and 57% for DMSC. From this point of view, it could be argued that the LD levels of metal song lyrics are higher than those of business or multidisciplinary lectures. Of course, this inference could be treated with caution since the corpus size has an important effect on the results.

Nguyen and Nguyen's (2016) findings show LD levels varying between 52.9% and 68.5% (p. 11) for the essays of Vietnamese mathematics freshmen students. These levels are quite high compared to the others mentioned earlier. They are higher than HMSC and TMSC altogether and could only be comparable to DMSC. However, a comparison between texts written by native and non-native speakers may not be appropriate. The tool used for the present study and Nguyen and Nguyen's (2016) are the same. To et al. (2013) showed that the LD levels of English textbooks across four levels (elementary, pre-intermediate, intermediate, upper-intermediate) were 53.2%, 53.8%, 46.3% and 45.5%, respectively. These results indicate that elementary and pre-intermediate text books are more lexically dense than HMSC and almost equal to TMSC. DMSC seems quite high in lexical density. It is also possible that different tools or principles applied in the specification of content words may provide different results which could deem comparisons to other studies utilizing other tools improper. All in all, whichever tool or principle is used, the LD analyses are capable of distinguishing different genres and/or proficiency levels significantly.

5.4. Overall Lexical Richness Discussion

The lexical richness analysis of the three subgenres showed that there was a significant difference between heavy metal and thrash metal and heavy metal and death metal. Thrash and death metal songs contained a higher proportion of less frequent words, and this could be interpreted as heavy metal being less lexically sophisticated than thrash and death metal, which are in fact subgenres of extreme metal. The same observation applies to differences between thrash metal and death metal. Although Purcell (2003) claims that death metal lyrics are often poorly written, the present study displayed counter results in terms of lexis, which is then in line with Phillipov's (2012: 89) argument that death metal lyricists try to outdo each other.

Lexical variation analyses were carried out using four difference indices of measurement (TTR, Guiraud, Uber and HD-D) and all of them yielded statistically significant results across all subcorpora. The fact that TTR performed similar to its more modern and complicated versions is remarkable. The reason for this could be the fact that the repeats were removed from each song and they were mostly equalized in terms of length, which probably covered for TTR's biggest weakness – text length. The other measures all yielded parallel results which suggest that each and any one of these measures could be used in genre comparisons as long as the text lengths are similar.

Owing to its automated nature overriding the risk of human error and saving time, LFP and P_Lex analyses could be made use of in the comparison of major subgenres. It might be argued that LFP and P_Lex analyses work well with song lyrics as well as they do with other texts and the GSL is still acknowledged as reliable list (see Sutarsyah et al. 1994, Reda 2003, Nation 2004, Gilner 2011, Brezina and Gablasova 2013). The BNC/COCA, on the other hand, also yielded similar results to the GSL. This indicates that both lists could be used interchangeably if a B2K analysis is to be conducted. The main difference between them is that the GSL consists of two one-thousand bands whereas the BNC/COCA 25. Therefore, the BNC/COCA could be preferred if the main concern is to identify the words in specific bands. The lexical density analysis also revealed statistically significant results across all three subgenres. It must be noted that the lexical density levels increased as the sophistication levels. This could be due to the fact that function words are mostly distributed within the higher frequency bands. When the frequency levels drop, it means that infrequent content words populate the texts thereby increasing the LD level. Gregori-Signes and Clavel-Arroitia (2015) achieved similar results across LD and LFP analyses they administered on students' essays and remark that LFP correlates well with another independent analysis of lexical richness.

Correlation analysis showed that all lexical richness measures are significantly correlated to each other albeit in varying degrees. The highest correlation values have been observed in LS meaning that any of them could be used interchangeably. Unsurprisingly, the highest correlation levels have been observed between the BNC/COCA and the GSL. This is an indication of the similarity of the two lists. Both lists are capable of distinguishing lyrics from different music genres and type or token level B2K analyses could be utilized to this end depending on the research setting. No large-scale difference has been observed between types and tokens; however, if an LD analysis is to be conducted alongside LS, a token-based approach might be suitable as LD is a measure of tokens. This was evident in the higher level of correlation between LD and tokens than types. P_Lex is also highly correlated with other LFP measures. This suggests that any of these measures could be applied to measure LS depending on the research setting.

As regards the LV, all methods are correlated as well. The lowest correlation was between TTR and Guiraud. Nevertheless, all measures could be used in the assessment of LV. HD-D and *Vocd* scores are not fully comparable to each other as they yield results on different scales although they are correlated. TTR, although posited positive correlations, should be treated with caution since it is sensitive to text length. As the song lyrics in MC are around similar length, it might have positively affected the measure in its performance.

The correlation among the LV values showed that Uber correlates the highest with other LV measures, a finding similar to that of Šišková (2012: 32). She found that Uber correlates well with HD-D (r=.8601, p<.05) and Guiraud (r=. 9235, p<.05) (p. 32). These results are similar to the present study in terms of the significantly high correlation levels, except for the fact that the results of the MC indicate an inverse order as Uber is correlated higher with HD-D (r=.926, p<0.01) than with Guiraud (r=.844, p<0.01). The correlation between Guiraud and HD-D was measured as r=.7511, p<.05 in Šišková's study and r=.828, p<0.01 in the present one which suggests that the findings are quite similar. She also found significant correlations between GSL B2K types and Guiraud (r=.71, p<.05) and GSL B2K types and Uber (r=.52, p<.05) (p. 34) which have been found as r=.558, p<0.01 and r=.576, p<0.01 in the present study. These values are arguably close to each other to some extent; however, the most significant finding between the two studies is that she found very low correlations between LD and Guiraud (r=.09, p<.05) and LD and Uber (r=0.26, p<.05). She also found no significant correlations between LD and LS results. The present study found relatively lower correlation degrees for the said pairs than the other measures; yet, they were all statistically significant. The reason for this difference could be stemming from the fact that Šišková's study contained fewer texts than the present one and done in an L2 setting. Additionally, the measurement tools were different. These factors might be the reason for the difference between the results.

Another study, which found a positive correlation between B2K and Uber, was Zheng (2016: 48). The correlation values in the said research varied between r=140 and r=906 (p<.017). Excluding these outliers, the lowest value observed was r=.573 which is similar to the present study. Finally, Daller and Xue (2007: 163) ran a similar correlation analysis between some of the values in the present study. Excluding the Guiraud Advanced values, which is not utilized in the present study, they found the highest correlation between D and Guiraud (r=.78, p<.01) followed by B2K and Guiraud (r=.73, p<.01). The finding for the first pair in the present study is (r=.828, p<0.01). It is not possible to compare the next pair in a definite way as the method of measurement in their study is not stated. Nevertheless, the correlation levels between R and LFP measures are between r=.528, p<0.01 and r=.564, p<0.01, which are lower than Daller and Xue's findings. They are still significantly correlated, though. Another interesting finding in their study is that TTR correlated negatively with B2K (r=-.37, p<.01). The possible reason for this is that the text lengths differed largely in their study whereas they were within specific boundaries in the present one.

Van Hout and Vermeer (2007: 112) found that Guiraud was positively correlated with *Vocd* (r=.61, = p<.01) but not correlated with TTR (r=.06). It must be noted that their sample size was quite low (n=32) featuring around 16,000 tokens (p. 111) in total compared to the present study featuring 600 songs and around 100,000 tokens. The reason for this difference might stem from the sampling size as well as the genre differences. They used spontaneous speech data acquired from children whereas the present study compared song lyrics which are written-to-be-sung. Daller and Phelan (2007: 242) also found a very low and insignificant correlation between TTR and Guiraud (r=.309) and a significant one between D and TTR (r=.483, p<.01). The latter pair has a positive correlation in the present paper (r=.832, p<.01). Their study was on 31 essays with an average token value of 280.77 (p. 238). The minimum (n=169) and maximum (n=563) token values are close to those in the present study. Thus, the reason for the difference might lie in the low number of samples and the difference between song lyrics and essays. They also conclude that TTR is not correlated with any measure. This is contrary to the present case. Applying TTR on long texts and using it in the comparison of texts of varying lengths is not likely to provide significant results. The positive correlation of TTR with other measures achieved with song lyrics could indicate that TTR is actually a solid measure which is capable of uncovering generic distinctions as long as repeats are removed and upper and lower token thresholds are set for lyrics. Hence, despite the common stance against the said measure, TTR could actually be used in music genre comparison.

As the last point of discussion, the results of the quadratic discriminant analysis were interesting to show that the class memberships obtained from lexical richness analysis overlap with the existing categorization only to a certain level (64.5%). The prediction rate was the highest for HMSC and lowest for TMSC. The clearest division was observed between HMSC and DMSC. Only 6% of the songs in HMSC were predicted as DMSC and 7% vice versa. This is parallel to the previous lexical richness analysis findings in that the highest difference has so far been observed between these two subcorpora. Perhaps the most interesting finding is that 29% of the songs in DMSC were predicted as TMSC. It indicates permeability between these two subgenres. This is also parallel to Weinstein's (2000) theory of Chaotic themes which are prominent in these two subgenres in particular. TMSC is again in the middle of the scale with 18% of its songs predicted as HMSC and 23% as DMSC. In this sense, the TMSC values are quite balanced between HMSC and DMSC.

The quadratic discriminant analysis was carried out with the purpose of assessing the match between the lexical values and musical generic classification. As mentioned earlier, it is not the aim of the research to devise an alternate way of music genre classification. Nonetheless, the 64.5% match is still interesting. It is advised that lexical analyses on music genres be carried out based on established music genres in an attempt to find out the lexical differences between them in an exploratory way. A reverse approach whereby musical genres are sought to be established through lexical (or any other lyrical or linguistic) analyses are not likely to provide reliable results given the outcome of the present corpus analysis.

5.5. Keyness

The results of the keyword analysis will be mostly compared to Taina's (2014) study as it is the closest paper to the present one. To start with the keyword numbers, the present study has identified 230 keywords for HMSC, 220 for TMSC and 288 for DMSC. According to Taina's (2014) finding, the more extreme subgenres contained fewer keywords. His study contained two more subgenres in addition to those investigated here; thus, only the relevant three subgenres have been taken into consideration. He found 23 keywords for heavy metal (the highest value), 18 for thrash metal and 9 for death metal (the lowest value) and these are inclusive of function words. The reason for this difference could be the fact that he had a rather smaller corpus (1/5 in size per subgenre) and he used a higher cut-off point although he does not mention which. The outcome in the present study was rather different in this sense as the most extreme subgenre, DMSC, yielded the most keywords both cumulatively and in content words. The order is reversed in the present study as HMSC contains fewer keywords than TMSC. However, from the viewpoint of key content words, the order is similar as it increases from HMSC to DMSC, thus resembling Taina's (2014). It is not suggested, however, that these findings indicate a pattern in the number of keywords in subgenres along the continuum of extremity. The distributions do display a similarity; yet, a quantitative deduction of a difference among the subgenres has to be treated with caution. Instead, qualitative investigation could be more useful to shed light on the generic differences.

The keywords have been found to come from different "domains" as suggested by Rayson (2008). The keywords are concordanced to look for differences in the contexts they occur in. It was stated earlier that there were two words in all corpus that are shared across all subcorpora: *soul* and *die*. *Soul* occurs 49 times in HMSC and 28 of these occurrences are in negative sense. For example, the occurrences of *soul* listed in Table 54 are in negative sense as they are parts of phrases such as *take one's soul*, *sell one's soul*, *steal one's soul* and others such as *soul is ill*, *soul is dying*, etc.

Table 54: Occurrences of *soul* in negative sense in HMSC.

heart will take your	**soul**	How could I have known?	heavy_us_03-01-01
forever? When your	**soul**	is dying alone And all the	heavy_us_08-01-10
never quite sure your	**soul**	is ill but you will not find	heavy_uk_02-02-03
salvation, selling your	**soul**	Lies and corruption,	heavy_us_05-01-07
and now I eat your	**soul**	Pray for blood to the	heavy_us_04-01-06
come to take your	**soul**	Satan has received	heavy_uk_01-01-08
Going to take your	**soul**	to Satan Suddenly	heavy_uk_01-01-08
She can steal your	**soul**	, with her mind control	heavy_us_01-01-05

On the other hand, there are neutral or positive occurrences of the same word some of which are listed in Table 55. Phrases such as *check one's soul* and *think about one's soul* seem neutral whereas *move one's soul* and *save one's soul* are of more positive nature.

Table 55: Occurrences of soul in neutral and positive senses in HMSC.

Hotel Check your	**soul**	at the door They have	heavy_us_06-02-02
music move your	**soul**	Burn the fire of rock	heavy_us_07-01-01
die Saving your	**soul**	by taking your	heavy_uk_06-01-02
thought about your	**soul**	- can it be saved?	heavy_uk_02-02-01

Soul has 57 hits in TMSC, 50 of which are used in negative sense. Some selected concordance lines are provided in Table 56. The word has been used in various phrases some of which are *give one's soul*, *metallization of one's soul*, *anger within one's soul*, etc.

Table 56: Occurrences of *soul* in negative senses in TMSC.

who gives his	**soul**	to Hell, must dare to	thrash_uk_07-01-01
goal fills his	**soul**	with a ruthless cry	thrash_us_06-02-02
The human	**soul**	Seen through the dark	thrash_uk_01-01-03
metallization of your inner	**soul**	, twisting and turning	thrash_us_06-01-08
the anger within my	**soul**	Bow to your master	thrash_uk_04-01-07

On very few occasions the word was used in neutral or positive sense as shown in Table 57.

Table 57: Occurrences of *soul* in neutral and positive senses in TMSC.

sweat and time A	**soul**	to save and I think it is mine	thrash_us_04-01-01
have mercy on my	**soul**	. Be gone foul beast that	thrash_uk_07-01-01

The distribution of the word *soul* in negative senses between HMSC and TMSC is noteworthy. The situation in DMSC is more or less a step forward from TMSC due to 65 hits of *soul,* 50 of which are used in negative sense. Although DMSC has more occurrences of *soul*, proportionally TMSC contains more instances in negative sense as seen in Table 58. It is found in phrases such as *rob someone of one's soul*, *curse one's soul*, *soul burning*, etc.

Table 58: Occurrences of *soul* in negative senses in DMSC.

of the Devil to rob it of its	**soul**	. Consume the lifeless,	death_us_02-01-10
Sorrow creeps throughout my	**soul**	All is lost, none have won	death_uk_02-01-02
Wrath of God - Satan Sin my	**soul**	, blesses with fire Throne	death_us_04-01-01
me insane Evil curse my	**soul**	Burning away Satan's hell	death_us_10-01-01
forged in the embers of my	**soul**	Burning with agony, the	death_uk_04-01-05

Die occurs 53 times in HMSC, 78 times in TMSC and 85 times in DMSC. Analyzed in the same method as *soul*, *die* is used 9 times in a positive or neutral context in HMSC. TMSC has 10 and DMSC has 6 positive/neutral instances. The use of *die* is obviously strongly associated with negative messages. DMSC expectedly has more occurrences of the word and the ratio of negative senses is the highest (79/85 [93%]). The ratio for HMSC is 44/53 (83%) and TMSC 68/78 (87%). The gradual increase in the negativity of *die* is reflected as in the other findings of the study. Some of the positive/neutral occurrences of *die* are given in Table 59 below.

Table 59: Occurrences of *die* in positive/neutral senses in MC

last As we fight not to	**die**	In this arena I who will	heavy_uk_03-01-03
Lords of time say never	**die**	And she said - do not	heavy_us_10-01-06
Who cares if we live or	**die**	Flies in a web we fall	thrash_uk_10-01-07
away I do not want to	**die**	Time moving slow, the	thrash_us_06-02-01
give Die, die, he cannot	**die**	Defying death with a	death_uk_10-01-04
a soul That can never	**die**	Following tales of	death_us_06-01-05

As seen in Table 59, *die* is used is phrases such as *never die*, *not to die*, *cannot die* and *not want to die*. These are rare examples positive/neutral uses of *die* in the whole MC. The remainder of all occurrences is in the negative sense. Some examples are provided in Table 60 below.

Table 60: Occurrences of *die* in negative senses in MC

will shiver - men will	**die**	A cast of millions - a	heavy_uk_06-01-07
- like warriors they	**die**	But gangland is alive	heavy_us_10-01-01
on just to see him	**die**	Death you want more	thrash_uk_08-01-02
eyes Now it is time to	**die**	Burning in my brain I	thrash_us_06-02-01
towards each other to	**die**	again too soon; They	death_uk_04-01-07
soon come with me to	**die**	And find that death is	death_us_09-01-05

The MC has an abundance of examples as given in xxx, where die is used in the negative sense. *Will die*, *time to die*, *you die* are common collocations. Death metal has the highest ratio of the word in general and in negative senses as mentioned above. The dominant ratio of negative sense words and the provided examples are hardly a surprise given the already acknowledged dark and offensive nature of death metal lyrics (Phillipov 2012, Purcell 2003).

Soul and *die* were the shared key content words across all subcorpora. The remaining ones will be dealt with separately subgenre by subgenre. To start with HMSC, it is interesting to note that the most key content word is *love*. It is interesting because of the massive negative perception regarding the genre.

Table 61: Occurrences of *love* in HMSC.

justice when you fall in	**love**	It gives you blindness	heavy_us_03-01-05
Echoing its refrain, my	**love**	still remains so Cry out in	heavy_us_09-01-08
feeling gone When our	**love**	was strong I have lost my	heavy_uk_10-02-02
realized before Your	**love**	is worth waiting for Do	heavy_us_03-01-07
again I let your lust for	**love**	excite me You took my	heavy_us_03-01-09
bitch, bitch, bitch, baby	**Love**	is a bitch, it is crazy	heavy_us_06-01-03
stays in the bed Making	**love**	the only thing we can do	heavy_us_01-01-08

It occurs in phrases such as *in love*, *my love*, *our love* and *your love*, which indicate a relationship with the opposite gender. It also bears sexual connotation as seen in phrases such as *lust for love*, *love is a bitch*, *making love*. These findings are supportive of Weinstein's (2000) claim that heavy metal bands mostly write lyrics on Dionysian themes, which involve sexual and debauchery related topics. The second key content word in HMSC is *never*. Although an adverb of frequency, which can also be considered a grammatical word, it is included in the study. *Never* is mostly used with *can* and *will* which indicate a stance against challenges and accentuate freedom. The third key content word is *let*, which is heavily used in structures such as *let me go*, *let us go*, *let you go*, *let it go,* etc. The use of *let* is definitive of heavy metal as it is used more frequently than the other subgenres in the study. *Night* is also worthy of note as it is the most key time expression in HMSC. It depicts a dark context which could also be interpreted in line with the Dionysian themes such as sex and drug use. HMSC contains four key content words which refer to time: *never, night, just, forever*. This is not the case with other subcorpora with the exception of *never* in DMSC. *Just* occurs fewer as an adverb of time than as an adverb meaning *exactly*. *Forever* could also be interpreted as a deterministic attitude against challenges. *Mazes*, *wheels* and *fly* are controversial keywords as *mazes* appears in only one song 20 times (heavy_us_02-01-07), and *wheels* occurs 36 times in total, 29 of which is in only one song (heavy_uk_08-01-04). Similarly, *fly* occurs 43 times in the HMSC, yet 27 times in one particular song (heavy_us_06-01-07). It is therefore not appropriate to treat these words as representative of the subgenre. Taina (2014: 71) has a similar observation where he discusses the keyword *you* in the following song:

Can you can you can you hear me
Can you can you can you see me
Can you can you can you hear me
Can you can you can you see me
<Thrash34>

Mazes, *wheels* and *fly* are similar to this example in that they are observed in only one each song, which is hardly representative of the subgenres. On the other hand, these words differ from *you* as given in the above example since *you* is already a very frequent word – a grammatical word – which would have probably been key again regardless of the effect of the song no <Thrash34>. In any case, *maze*, *wheels* and *fly* have to be treated with caution as they hardly look representative due to their poor distribution across songs. *Know* is key as it is used mostly as a token of curiosity (*Where you end up can you ever know* [heavy_us_08-01-01]) and certainty (*You know I like that sound* [heavy_uk_06-01-08]). The last key content word *cry* is used in both senses as shouting and weeping as in the examples: *I still hear the battle cry* (heavy_uk_07-01-06) and *Only so many tears you can cry* (heavy_us_03-01-06).

As most of the discussion on keyword revolve around Taina's (2014) findings as the closest study to the present one in terms of keyness, it would be appropriate to compare the key words of the HMSC to Taina's heavy metal corpus.

Table 62: Taina's (2014) keywords for three subgenres in the present study (adapted from Taina 2014: 49)

Traditional heavy metal	**Thrash metal**	**Death metal**
Re	Can	Thy
Gotta	Bring	Thee
Gonna	Death	Art
Got	Down	Cast
Yeah	Fuck	Peace
You	Nightmare	Of
T	Me	Beyond
Rock	Blood	Their
Ain	Wall	Soul
Shake	Shit	
It	Takes	
Baby	Hell	
Love	Cost	
If	You	
Girl	Show	
Roll	Murder	
Turning	Trapped	
King	Die	
Devil		
Same		
Angel		
Feeling		
Sweat		

As noted earlier, the fact that the number of key words is highest in heavy metal in Taina (2014) is a contradictory result to the present study as the highest number of keywords is observed in DMSC. In terms of the words themselves, only one key content word appears in both studies: *love*. Of the very key words in Taina's heavy metal corpus, in HMSC *rock* is the 175th keyword (+42.66), *baby* 178th (+42.49), *roll* 140th (+51.93), *turn* 206th (+35.79 – not *turning*), *devil* 150th (+49.13), *angel* 106th (+71.15), *feel* 47th (+136.04 – not *feeling*). *Shake*, *girl*, *king*, *same*, and *sweat* are not among the keywords of the HMSC. Taina's findings do not include keyness values of the words; therefore, it is not possible to fully compare the results.

The results obtained for heavy metal in both studies show that the keywords depend on the texts chosen from the genre, their number and the reference corpus used (the reference corpus is not stated in Taina [2014]). Although these findings may lead to the conclusion that there are not specific keywords that are typical to heavy metal, that is actually not the case. Despite these differences, both sets of keywords consist of relatively positive and Dionysian ones. Therefore, the keyness results of both studies confirm Weinstein's theory. This is important as her theory was based on an observation and judgement whereas the keyness results are of statistical nature.

The most key content word in TMSC is *blood* – which is not surprising owing to the aggressive nature of thrash metal. Blood is used mostly in negative sense except for very few occurrences such as *blood brother*. There are 97 instances of the word and of them are in *your blood* and 9 in *in blood*. Some examples are given in Table 63:

Table 63: Occurrences of *blood* in TMSC.

screen Covered in	**blood**	, does it make you feel	thrash_uk_08-01-02
throat, drink your	**blood**	, who cares when others	thrash_us_03-01-07

In blood is frequently used in phrases such as *covered in blood*, *signed in blood*, *bathe in blood*, etc. *Your blood* occurs in *through your blood*, *pulsing in your blood*, *drink your blood*, etc. These instances are supportive of Weinstein's (2000) theory that thrash metal uses Chaotic themes which center around violence, death, war, murder, and so on. Parallel to this finding is the fact that *death* shows up as the second most key content word. *Death* occurs 117 times in TMSC in total and 27 times in *of death*, which is the most frequent collocation. Some examples are shown in Table 64,

Table 64: Occurrences of *of death* in TMSC

Rotten egg air	**of death**	wrestles your	thrash_us_05-01-07
a wrecking ball	**of death**	and completely	thrash_us_04-01-04
your bringer	**of death**	Your child of chaos	thrash_uk_04-02-01
seen A choice	**of death**	made for man By	thrash_us_07-01-04
hide the face	**of death**	Oppression ruled by	thrash_us_09-01-02

The frequent use of this collocation is indicative of a more formal language as advocated by Taina (2014) as well who studied this very particular pattern (*of*) across five metal subgenres. In this regard, more formal language – through the use of more prepositions and nominalizations as also claimed by

Biber (1986: 395) – lead to a more lexically rich language which can be backed up using a keyness analysis as well. *Kill* occurs 72 times as the fourth most key content word after *die* which was discussed earlier as one of the two common key words. Following those, comes *life*, which seems contradictory to the main flow at first. It is used 169 times and mostly in *my life* (18) and *your life* (36), some of which are shown in Table 65.

Table 65: Occurrences of *my life* and *your life* in TMSC

an ordinary type of bloke,	**My life**	is mundane, I hate	thrash_uk_03-01-07
machines changed	**my life**	Forgotten how life	thrash_uk_06-01-04
feel great, painkiller in	**my life**	Cut bare flesh to the	thrash_us_03-01-02
, it is how I live	**my life**	I cannot take it any	thrash_us_06-01-02
Reject the pain I suffer	**My life**	was never yours to	thrash_uk_04-01-07
for the time, to take away	**your life**	. A fist full of fury,	thrash_uk_10-01-06
... as yesterday Fight for	**your life**	For blood and iron	thrash_us_08-01-07
ultimate defense Beg for	**your life**	, they bring your	thrash_uk_08-01-05
You shiver in fear for	**your life**	Uneasy with your	thrash_us_07-01-08
As terror strikes, give	**your life**	We drink your blood,	thrash_us_08-01-07
from you and me	**Your life**	is such a mess, you	thrash_uk_06-02-04
cash and bring to me	**your life**	Place your faith in	thrash_uk_01-01-01
in pieces Nothing of	**your life**	remains Rest in	thrash_uk_04-01-04
your eyes from the sight	**Your life**	thrown away years	thrash_uk_01-01-09
God to say: "I will take	**your life**	from you"? Flash	thrash_us_06-02-01
and we are going to take	**your life**	Kick in your face	thrash_us_03-01-02
They have come to take	**your life**	On through the dead	thrash_us_06-01-01
a mother's son to take	**your life**	They say he died ten	thrash_us_08-01-09
a mother's son to take	**your life**	They say he died ten	thrash_us_08-01-09
those who took	**your life**		thrash_uk_08-02-01
your crime They took	**your life**	but not your soul	thrash_uk_08-02-01
to forever cry Took	**your life**	, I do not feel bad,	thrash_us_03-01-07
can you die when	**your life**	is a lie?	thrash_uk_10-01-08
for all of you,	**your life**	is through There is...	thrash_us_08-01-03

My life has been used in both in positive (*it is how I leave my life, my life was never yours to steal*) and negative (*my life cut bare to, my life is mundane, machines changed my life*) contexts. In this regard, a stance on freedom and complaint can be observed in thrash songs. *Your life* is used more often than *my life* (exactly twice as many) in structures such as *fight for your life*, *beg for your life*, and most of all *take your life* (8 times). This clearly indicates a hostile attitude, which is in line with the general theme of thrash metal subgenre.

These examples portray a similar picture as Weinstein's (2000) theory that thrash metal draws heavily on Chaotic themes. *Hell* is an abstract word, which indicates torment and suffering even after death. The rest of the key words similarly depict a horrid and sinister tone (e.g., *burn*, *fear*, *dead*, *evil*, *pain*, *insane*, etc.). *Take* is different from others in that it is usually a neutral verb. 126 instances of the word are mostly realized in structures with negative connotations, such as *take away, take my soul, take your life,* etc. *Fight* is of importance since it highlights the reactionary and anti-social stance of thrash metal. Another significant finding is that TMSC is the only corpus to feature *fucking* as a very key word. It occurs in HMSC and DMSC as well, yet as the 211th and 74th key word respectively. The high rank it has achieved in TMSC is also indicative of the protest, uncompromising and relentless lyrical style of the genre. *Eyes* are used in structures as, *empty eyes*, *frozen eyes*, *dead eyes*, *psychotic eyes*, etc. *My eyes* and *your eyes* are very common structures occurring 14 and 18 times respectively. This can be interpreted as direct interaction with the listener and indirectly the system suggesting different views on life style and ideology.

Finally, *lies* is another key word indicating anger and confrontation to society's thoughts and beliefs. Unlike the Dionysian themes in heavy metal which makes use of personal relations more of then not involving sexual connotation, thrash metal takes on a social stance and defies the society on the basis of wars, corruption, inequality, etc., whereby *lies* is testament to this social attitude manifesting itself in lines such as,

Table 66: Occurrences of *lies* in TMSC

men preaching holy	**lies**	. Too blind to see the	thrash_uk_10-01-07
your never ending	**lies**	Your supremacy is fading,	thrash_uk_06-02-02
with their malicious	**lies**	, their unforgotten violence,	thrash_uk_07-01-06
Wicked smile, full of	**lies**	Head of snakes, approach	thrash_us_01-01-08
Opposition book of	**lies**	I am your darkness	thrash_uk_04-02-01
And now the world	**lies**	deathly still Ruled by	thrash_us_07-01-04
see through all the	**lies**	we do not walk away	thrash_us_08-02-01

The examples depict a disbelief and complaint of existing system, which is the common adversary in thrash metal. The key content words of thrash metal as found by Taina (2014) are *bring, **death**, **fuck**, nightmare, **blood**, wall, shit, **takes**,*

hell, *cost, show, murder, trapped* and ***die***. There are six matches with the top-30 key words in the present study, which are shown in, bold (see also Table 44). In fact, *fuck* in Taina's study appears as *fucking* and *takes* as *take* in the present study but they are the variations of the same lemma. *Murder* and *kill* can also be considered similar in this respect. Aside from the matching words, the other ones carry negative meanings as well in both studies. This, in turn, supports Weinstein's Chaotic theme theory.

Death is unsurprisingly the most key content word of DMSC occurring 150 times in the subcorpus. It is only natural for the word to be assigned as the name of the subgenre. The most frequent structure it occurs in is *of death* (33) as shown in Table 67,

Table 67: Occurrences of *death* in DMSC

in the fucking web of	**death**	(suffer) A purgatory of	death_us_06-01-09
Agonizing lobotomy of	**death**	Assumed consequence	death_uk_10-01-05
corpses, eternity of	**death**	confirmed Destined to	death_uk_10-02-01
darkest deep holes of	**death**	. Pain evolves slow,	death_us_09-02-04
soon fade Silence of	**death**	now heard By mutant	death_us_07-01-05
forth the throes of	**death**	Awaken in darkness to a	death_us_07-01-08
blood, concentrate of	**death**	Congregation is dead	death_us_04-01-05
the same! The sight of	**death**	drives me nuts, I must kill	death_us_02-01-08
to be free Shadow of	**death**	, to meet the bitter end	death_us_06-01-10
begun. The look of	**death**	in my eye, surely no one	death_us_02-01-01

This finding is parallel to that of Taina (2014) stating that noun phrase use through *of*+noun structures is more frequent in death metal than other genres. One way the present study differs from Taina (2014) is that *death* was not in his list of keywords for death metal. He argues that the reason for this could be the fact that the subject is dealt with using other words (p. 82). However, it is the second overall key word and the first key content word in the present study. This clearly indicates that the songs chosen affect the resulting keywords.

The next key content words are *flesh* and *blood*. In fact, a look at the first three key content words reveals the genre for what it is. The fascination for death, gore and violence, as proposed by Weinstein (2000), Purcell (2003), Phillipov (2012), etc., are overtly evident in the key words of DMSC. *Evil* and *Satan* are other keywords highly ranked at the top of the list. *Satan* is interesting here as it only occurs in DMSC at very high ranking – 8th. It is the 149th key word in HMSC and 32nd in TMSC. This finding is consistent with the evil and merciless attitude of the genre. The following keywords are comparable to those in TMSC to a great extent – *life*, *dead*, *soul*, *pain*, *kill*, *hell* and *fear*. The fact that these key words overlap across TMSC and DMSC could be considered a linguistic proof that a) these genres are under the same supra genre (extreme metal) and b) "death metal was created out of thrash metal" (Kahn-Harris 2000: 17). Both genres share similar aggressive traits; however, thrash metal uses these themes to oppose the system in an attempt to state is wrongness and express social matters explicitly with no regard of political correctness. Death metal, on the other hand, mostly uses aggression for the sake of aggression and relentlessly and uncompromisingly thrives on cruelty and evil. The following key content words could support this claim: *God*, *decay*, *mortal*, *grave*, *rotting*, *darkness*, *lust*. They have almost no social context and highly recall death. *God* is interesting as it ranks quite key albeit not as much as *Satan*. This could be viewed as mockery and metaphor, which indicates the fight between good and evil of which balance is tipped toward the dark side. Some sample lines featuring *God* are provided in Table 68.

Table 68: Occurrences of *God* in DMSC

righteous For your	**God**	is dead	death_us_04-01-02
inherit the meek Your	**god**	is dead Bound down,	death_uk_03-02-03
, body decays Your	**god**	is dead No god There	death_uk_05-01-06
evermore I am your	**god**	Bow down or die	death_uk_05-01-01
my slave I am your	**god**	now I am your saviour	death_uk_05-01-10
Killing to become your	**god**	So I have become your god	death_uk_08-01-02
As you wait for your	**god**	or the void Or the abyss	death_uk_03-01-05
You will meet your	**god**	Chained to torment	death_us_10-01-07

Your God is the most frequent phrase occurring 20 times and indicating that *God* is attributed to the adversary and not to the subgenre. Phases such as your *God is dead* and *I am your God* support the atheistic and anti-religious attitude of the genre. *Satan* is used 47 times in various structures. This frequent use is a part of the adversary view expressed in *God* as well which well corresponds to the existing Chaotic theory by Weinstein (2000) and the violent themes as stated by Arnett (1996), Buts and Buelens (2008), Philips and Cogan

(2009) and Frandsen (2011). *Now* and *forever* are two key adverbs in DMSC. It is interesting to note that *forever* is also key in HMSC. However, a concordance analysis shows that they occur in quite dissimilar contexts.

Table 69: Occurrences of *forever* and *now* in DMSC

malaise is alive in the air	**Forever**	by horror enslaved Gone	death_us_01-01-04
flesh and bone Trapped	**forever**	, caged within Mental	death_uk_01-01-05
A slave to the curse	**forever**	confined Shatter the	death_us_03-01-06
pain Fire and hell,	**forever**	Into the flames Sacrifice	death_us_10-02-01
where time stands still	**Forever**	life of pain Walking dead	death_us_10-01-02
finished nice and neat	**Forever**	lost they thought In the,	death_uk_06-02-03
I am the evil inside	**forever**	now I will reign Fallen,	death_uk_05-01-01
of pure repugnance.	**Forever**	scarring a once beautiful	death_uk_07-01-04
of madness Vision	**forever**	stained There is no use	death_us_06-01-09
, captured and	**now**	put to death	death_us_06-01-07
opened their eyes and	**now**	they will not be saved	death_us_06-01-06
fulfilled The flames are	**now**	burning hot Bodies are	death_us_10-01-09
shadows of the past are	**now**	forsaken From the dawn	death_uk_02-02-04
mind survive You are	**now**	ready for the kill Kill to	death_uk_05-01-08
death Man swept aside,	**now**	inevitable demise	death_uk_10-02-02
and balls Fall asleep,	**now**	for the attack In no time	death_uk_06-01-04
victory, into body-bags	**now**	scraped... Regnant	death_uk_03-02-01
continues - this battle	**now**	is won Mourn no more -	death_uk_02-01-01
Enjoy your begging	**now**	you... Die! Can never	death_us_05-01-09
, immortal, bestial,	**now**	he owns your soul.	death_us_02-01-03
old Stagnant blood	**now**	flowing free The mating	death_us_01-01-05
cut them all, your blood	**now**	flows Your kneecaps,	death_us_01-01-10
arteries gushing blood.	**Now**	it is time to feed on flesh,	death_us_02-01-01

As seen in the examples in Table 69, *forever* is frequently used in negative contexts referring to an unending state or continuation of unpleasant and horrific concepts. *Now* occurs 164 times in DMSC and emphasizes the instant realization of evil deeds on the adversary.

It is interesting that the keywords for death metal found by Taina (2014) and the present study are quite different and the only match is the word *soul*. As mentioned earlier, even *death* is not among Taina's death metal keyword list although it is the second key word in the present study. What is even more interesting is that the number of keywords is the lowest for death metal in Taina (2014) while it is the opposite in the present paper. These differences indicate that the corpus is very important in studying keywords. The controversies in Taina's attribution of bands to particular genres have been discussed earlier (Lyrics Corpora and Studies) and this may be a reason for different results as well as the fact that the present paper is based on subcorpora which are five times the size of those in Taina (2014). Obviously, this does not imply that the present study offers better or more reliable results. There may be flaws and limitations, too, which will be discussed in the limitations section. Besides, the differences with Taina (2014) mostly surface in death metal while similarities become dominant in heavy metal and thrash metal. Hence, the two studies may complement each other.

To sum up, the keyness analysis conducted in the present study is capable of distinguishing the generic differences across HMSC, TMSC and DMSC. The findings are highly compliant to Weinstein (2000) and Taina (2014). It must be noted that Weinstein does not reserve Dionysian themes exclusively to heavy metal and proposes some heavy metal albums to feature Chaotic themes as well. In fact, her argument of Chaotic themes revolves around examples from heavy metal bands such as *Black Sabbath*, *Iron Maiden*, *Judas Priest*, *Ozzy Osbourne*, etc. (pp. 40-3). Her judgement holds true for the songs she exemplifies; yet, the results of the keyness analysis in the present study suggest that thrash metal and death metal are more Chaotic on a broader scale.

While heavy metal displays a more Dionysian approach through the use of "lighter" words, thrash metal and death metal boast a dark, chaotic and evil selection. Based on these findings, it could be argued that the lyrical themes and domains of bands or genres could be identified with a keyness analysis. Although the key words might already be evident – as Weinstein found – keyness analysis offers a more scientific approach instead of human judgment. When used with concordancing, the structures give a better view of the words in context and enable the researcher to see the usage patterns more clearly.

6. CONCLUSION

This thesis study sought to compare three metal subgenres – heavy metal, thrash metal and death metal – with regard to their lexical richness levels through a sampling which included 200 songs from each subgenre. Three different methods of measurement for lexical richness have been applied at nine operational levels: lexical sophistication (GSL B2K tokens, GSL B2K types, BNC/COCA tokens and BNC/COCA types), lexical variation (TTR, Guiraud, Uber and HD-D) and lexical density. Additionally, a keyness analysis was conducted. It has been proposed that metal song lyrics – or lyrics from any other genre, for that matter – could be compared in terms of lexis using lexical richness and keyness analyses. All analyses yielded statistically significant results between heavy metal, thrash metal and death metal. The emerging pattern in all analyses was a gradual increase from heavy metal to death metal. Thrash metal was in between in all results. These results showed that a higher level of lexical richness is observable in heavier music lyrics.

It has been found that lexical richness and keyness analyses are powerful tools of music genre comparison. They can also be used in the comparison of other genres in combination with each other or in isolation. They can be preferred especially if/when a certain sentence structure is missing or punctuation is problematic in a way that renders a syntactic analysis difficult or impossible. In general, all lexical richness analyses are quantitative and data-driven where the input consists of tokens and types and their proportional distribution in texts. LFP and P_Lex were based on the GSL and the BNC/COCA, meaning the frequency analyses were carried out on established reference points rather than pure ratio calculation as is the case with other lexical richness analyses. Still, the reason for obtaining significant results might lie in the subgenres/bands/albums chosen despite the fact that the corpus is representative enough compared to similar studies.

In addition to the quantitative measures, keyness analysis provided qualitative insight to the study as an alternative to human judgments in genre comparison. The gradual increase of lexical richness is also evident in the keyness levels of the genres. However, heavy metal is separated from thrash metal and death metal more clearly than the lexical richness analyses. The themes which are prominent in heavy metal are different from those in thrash metal and death metal. The addition of this analysis was useful in this respect.

The comparison of results to other studies has been provided in the previous sections. However, the references for comparison consisted of various other texts including songs, which are not limited to metal. It must be noted that there is not a fully comparable study that is conducted in metal using the same methodology. The characteristics of metal lyrics have been laid out; nevertheless, for the most part, they are qualitative and meaning-driven remarks which do not involve qualitative measurement methods (see Walser 1993 and Weinstein 2000). The present study took on an exploratory approach in an attempt to distinguish metal subgenres thorough lexical analyses. Whether or not similar results will be obtained could be verified through replication of the study in the same manner or following more advanced ones as discussed in Further study.

The study showed that the existing genres could be analyzed using lexical methods and tangible and significant results could be obtained. It could be used as a reference as to which measures to use when comparing genres. It could be an alternate – or lighter – version of Biber's MDA to distinguish genres. Xiao and McEnery (2005) also attempted at finding a "low effort" alternative to MDA with keyness analysis and the results of the MDA and keyness analyses were similar. However, it is not one of the results of the present study that lexical analyses could be used to identify/assign genres in metal. Although this may be likely, it cannot be a proper way of genre association as other factors are eliminated. Instead, the present study has taken the genres for granted and sought to find if the existing generic borders existed in lexis or not and the results were supportive of a match between the metal subgenres.

6.1. Limitations

The present research inevitably has its limitations. As there was no fully comparable study in either the field of musicology or linguistics, the research setting in the study was compiled from various studies. Reflections on the study are provided under separate headings in the following sections.

6.1.1. Genre Definitions

This study analyzed three metal subgenres which were relatively easy to categorize. Not all subgenres, nevertheless, can be studied as easily in a scope similar to this. It would certainly cause problems if one were to compare symphonic black metal to avant-garde death metal, for instance. There is a need for clarity for the bands to be assigned to a specific subgenre through sufficient reference in the academic literature and media. A work-around for this issue could be opting for broader subgenres, e.g., taking extreme metal instead of thrash and death metal. However, Tsatsishvili (2011: 32) cautions against the umbrella genres by stating that "so called 'umbrella' genres such as avant-garde, industrial, metalcore and NU-metal […] feature very diverse musical characteristics".

Although all the analyses showed significant differences between the subgenres, it must be noted that comparing close subgenres might need caution as they will most probably yield similar results. In other words, it could be argued, as a foresight, that the more branched a subgenre is (e.g., symphonic melodic black metal), the less likely it is to find significant differences with its neighboring or parent subgenre (e.g., melodic black metal).

As a final remark, the justification of a band being assigned to a specific genre cannot always be achieved through the academic literature. Thus, looking for references in the literature, as carried out in the present study, could be limiting in establishing a comprehensive corpus. Not all bands could be mentioned in academic texts and this fact should not eliminate them from academic studies. Instead, media and reliable music sources could be used as reference points in genre specification.

6.1.2. British/American Bands

This study included NS bands on purpose in order to eliminate the L2 factor. It could be argued that all bands could have been chosen based on their country of origin, such as choosing only American bands. This might be plausible; yet, it could be contrary to the evolution of metal. Heavy metal being originated from the UK makes it almost mandatory to choose British bands to be the representative of the subgenre. Alternatively, the number of bands to be included in the corpus might be increased to feature equal number of bands from both USA and UK (or even from Canada, Australia, New Zealand, etc.) for each subgenre. Given these facts, it seems that setting up a corpus stands on an edge, which could easily be tipped off unless a balance is achieved in all aspects.

6.1.3. Representation of the subgenres

The study was designed to include the bands, which have at least three full-length albums. This was useful in attaining representation as well as having back-up albums when one or two albums of a band do not meet the criteria of having sufficient number of songs that contain no fewer than 200 words. It might seem a strict cut-off point, which leaves out crucial bands, yet, it was necessary to have an equal balance and representation.

Another point is that, the number of releases of the bands does not always offer parallelism since the productivity of the bands varies. There was no limit on the release years of the albums in the corpus. The effects of the productivity and the maturity of the bands are open to discussion, though. For example, in their comparative analysis on the poems written by suicidal and non-suicidal poets, Stirman and Pennebaker (2001) address the issue of different vocabulary profiles depending on the periods of the poets' careers. The present study did not take into consideration the linear progression of the lyrical maturity of specific bands or subgenres. Instead, it took on a cumulative approach and included albums spanning across various years. Besides, release years of particular songs might not be a sound indicator of their specific lexical traits due to the fact that those songs could have been written earlier than their release years. Album-wise, not all songs in an album could necessarily be written in the same year. It could be an accumulation of various levels of 'maturity' spanning across many years.

More comprehensive results could be attained by incorporating all the albums by each band; yet, then again, this would have its own limitations. If the bands in the corpus have changed their styles in their careers even in one album, that may affect the results. Additionally, not every band has the same or similar number of albums, so including the full discography of a band with twenty full-length releases together with one with only three is likely to disrupt the balance of the corpus.

The results obtained in the study might be restricted to the bands and albums chosen and therefore might not fully represent the subgenres despite the utmost care taken in the selection of the most representative bands. Thus, it may not be appropriate to suggest a general feature for a specific subgenre. Instead, the study rather hints at a 'tendency' that could (or could not) reflect a common trait for the subgenres. Verifications or refutations to the findings of the study could be made through the inclusion of more bands from each subgenre to achieve more generalizable results.

6.1.4. Lyrics

The lyrics of the albums in the corpus were retrieved from *Encyclopaedia Metallum* and they were scanned for typos. One method could have been exactly copying the text off the website, keeping the typos, and proceeding accordingly. Yet, this approach might have led to different results, ones which the artists have no responsibility for. Therefore, the lyrics were corrected through spellchecking features of the word processor and sometimes cross-checking with other resources with utmost care to find out the exact words. These corrections were made to the best of the author's capability; however, there might have been errors. It must be noted that a major mistake leading to a whole different result is not quite likely; yet, the human factor should always be taken into consideration.

All analyses are well-established; yet, applying them on song lyrics was rather more challenging than plain texts. The removal of repeats and chorus parts was an issue which was dealt with a custom method proposing two categories: exact repeats and partial repeats. It could be argued that the repeats be kept to maintain fidelity to the original artwork but the final decision was to remove them as there are different amounts of repeats in each song which might have eventually affected the results. The method devised in the present study is adopted from similar studies mentioned in the respective sections. Same study or similar ones could be conducted with an alternate setting in which the original lyrics are maintained in their entirety.

6.1.5. Analyses

Given the time constraints and the availability of measurement methods, not all lexical richness measures could be included. However, all the lexical richness methods were chosen to be representative of each respective methodology. Although it is not very likely to achieve massively different results due to the facts obtained from the correlation analysis, other methods could be applied to find out whether different frequency lists or measurement methods/tools will yield different results.

Although a good way of uncovering otherwise hidden features of texts, keyword analysis is not without its limitations. A major point here is the semantic aspect and synonymy. The analysis is carried out on individual words irrespective of their meanings and each word is considered a unique entity. However, some words are indeed semantically related and could be used instead of one another. Baker (2006: 143) voices this issue as follows:

> [T]ext producers may sometimes try to avoid repetition by using alternatives to a word, so it could be the case that it is not a word itself which is particularly important, but the general meaning or sense that it refers to. For example, it could be the case that the notion of 'largeness' is key in one text when compared to another, and this is demonstrated by the writer using a range of words such as big, huge, large, great, giant, massive, etc. - none of which occur in great numbers, but taken as a cumulative whole, would actually appear as key.

This study was carried out in a lexis-oriented methodology and grammatical and semantic aspects were disregarded. In fact, this was the point in the study. The aim was to test if it is possible to differentiate between the genres only through a lexical approach. In this sense, the grammatical and semantic aspects could be thought of as additional factors – as in the MDA – and could be used in a complementary way as further study.

Another point is that, repeats are not equally distributed across songs, so the ones with more repeats (partial ones) may cause a particular word to be key although it might occur in only one song, such as in *tailgunner* in HMSC. A cut-off limit was established in the keyness analysis and only top-30 words were taken into consideration. There are almost 400 keywords in each subgenre and it is virtually impossible to concordance and analyze each and every one of them given the constraints on time and scope.

Secondly, as indicated earlier, the choice of the reference corpus definitely has an effect on the results. However, it would have been the case no matter which other reference corpus had been used. As Scott (2009) has shown, there is, in fact, no bad reference corpus. One would get similar results from any reference corpus even when a very "absurd" one is used. This argument could also be used in defense of the choice of the BNC as the reference corpus. Despite the fact that it reflects British English and is rather dated, it is still a widely used corpus. The software used does not feature the BNC/COCA list; therefore, the analysis was conducted with the BNC. In this respect, US bands were benchmarked against a British English reference corpus. However, difference at a large scale is not expected as the results of the keyness analysis do not vary much as pointed out by Scott (2009). Even though a few words might be different across different reference corpora, the general outcome would still be able to point out to differences across the subgenres.

6.2. Further study

The present study included only NS bands to maintain equivalency in linguistic competence. This was necessary in order to be able to eliminate yet another factor at play. However, it is important to reflect that many bands representing major subgenres have to be excluded as a result of this selection criterion. Morris and Cobb (2004) examined the differences between NS and NNS students' essays using *Vocabprofile* and Kormos (2011) underlines that L1 writers outperform L2 writers with regard to syntactic complexity. Such studies in the literature could be replicated to achieve similar indicators for metal music. Further studies could, and should, include NNS bands as well, so that a more comprehensive corpus with a stronger capacity to represent the genres could be established.

Alternatively, further analyses could be carried out with bands of the same L2 origin – e.g., an all-German corpus. This may allow the L1 factor to be standardized for each band and achieve more meaningful results. On the other hand, the educational levels of the bands (song writers) might affect the outcome, but this notion is a universal one. Establishing criteria which involve educational background of the artists sounds quite far-fetched.

The analyses in the present study were limited to 200 songs from each subgenre. The reason for not being able to add more bands was the concerns raised in the Representation of the subgenres. It may not be plausible to include a large selection of albums using the setting applied in the present research because of the time constraints. Therefore, an alteration might be necessary in the corpus building criteria.

Finally, the results obtained in the present study might be dependent on the chosen measurement methods. The reasons for the selection of these methods have already been discussed in METHODOLOGICAL APPROACH along with the justifications of why some other alternatives have not been used. However, it cannot be overlooked that various other methods of measuring lexical richness could be applied depending on the research design in order to support/refute the results or propose an alternative viewpoint. Moreover, the lexical richness results provided by the present study could be tested on different levels of complexity apart from lexis, such as syntax, semantics, etc.

7. REFERENCES

Aarts, Flor and Jan MG Aarts. *English Syntactic Structures: Functions and categories in Sentence Analysis Vol. 1*. Pergamon, 1982.

About the Awards - RIAA. n.d. 10 December 2016. <https://www.riaa.com/gold-platinum/about-awards/>.

Abrams, Zsuzsanna I. and David R. Byrd. "The effects of pre-task planning on L2 writing: Mind-mapping and chronological sequencing in a 1st-year German class." *System* 63 (2016): 1-12.

Advanced Search - Encyclopaedia Metallum: The Metal Archives. n.d. 28 March 2017. <http://www.metal-archives.com/search/advanced/searching/bands?bandName=&genre=death&country=&yearCreationFrom=&yearCreationTo=&status=1&themes=&location=&bandLabelName=#bands>.

Ågren, Malin, Jonas Granfeldt and Suzanne Schlyter. "The growth of complexity and accuracy in L2 French: Past observations and recent applications of developmental stages." *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Ed. Alex Housen, Folkert Kuiken and Ineke Vedder. John Benjamins Publishing Company, 2012.

Ai, Haiyang and Xiaofei Lu. "A web-based system for automatic measurement of lexical complexity." *Paper presented at the 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10)*. Amherst, MA, 2010.

Aijmer, Karin and Diana Lewis. "Introduction." Aijmer, Karin and Diana Lewis. *Contrastive Analysis of Discourse-pragmatic Aspects of Linguistic Genres*. Springer International Publishing AG, 2017.

Ailes, Drew. *Six Nu-Metal Bands You Shouldn't Be Ashamed of Liking*. 05 September 2016. 13 October 2017.

<http://www.phoenixnewtimes.com/music/six-nu-metal-bands-you-shouldnt-be-ashamed-of-liking-6625165>.

Allen, Edward D., et al. "Comprehension and text genre: An analysis of secondary school foreign language readers." *The Modern Language Journal* 72.2 (1988): 163-172.

Anderson, J. Charles. "Judging the frequency of English words." *Applied Linguistics* 28.3 (2007): 383-409.

Anthony, Laurence. "Antconc Version 3.5.7." Tokyo: Waseda University, 2018.

—. "AntWordProfiler 1.4.0w." Tokyo: Waseda University, 2014.

Archer, Dawn, Jonathan Culpeper and Rayson Rayson. "Love – 'a familiar or a devil'? An Exploration of Key domains in Shakespeare's Comedies and tragedies." *What's in a Word-list? Investigating Word Frequency and Keyword Extraction*. Ed. Dawn Archer. Ashgate, 2009. 137-157.

Arnaud, Pierre J. L. "The Lexical Richness of L2 Written Productions and the Validity of Vocabulary Tests." Essex Univ., Colchester (England). Dept. of Language and Linguistics. *Practice and Problems in Language Testing. Papers from the International Symposium on Language Testing (7th, Colchester, England, 1984). Occasional Papers, No. 29*. Ed. T. Culhane, C. Klein-Braley and D. K. Stevenson. 1984. 14-28.

Arnett, Jeffrey Jensen. *Metalheads: heavy metal music and adolescent alienation*. Denver, CO: Westview Press, Inc., 1996.

Asch, Andrew. *THE HEAVY WAIT IS O V E R; A new generation of metal music bringing in listeners; Four decades: Poster boys of metal; TEN BANDS THE EXTREME-MUSIC ENTHUSIAST SHOULD KNOW*. 17 January 1999. 20 May 2016. <http://search.proquest.com/docview/252594342?accountid=17219>.

Assunção, Carlos and Carla Araújo. "A Lexicometric Analysis of the Poems from O Guardador de Rebanhos." *Acta Scientiarum. Language and Culture* 41 (2019).

Aucouturier , Jean-Julien and François Pachet. "Representing musical genre: A state of the art." *Journal of New Music Research* 32.1 (2003): 83-93.

Baayen, R. Harald and Rochelle Lieber. "Word frequency distributions and lexical semantics." *Computers and the Humanities* 30 (1997): 281-291.

Baayen, R. Harald, Petar Milin and Michael Ramscar. "Frequency in lexical processing." *Aphasiology* 30.11 (2016): 1174-1220.

Baker, Paul. "Corpora and Gender Studies." *Corpus Applications in Applied Linguistics*. Ed. Ken Hyland, Chau Meng Huat and Michael Handford. Continuum, 2012.

—. "Querying Keywords: Questions of Difference, Frequency, and Sense in Keywords Analysis." *Journal of English Linguistics* 32.4 (2004): 346-359.

—. "The BE06 Corpus of British English and recent language change." *International Journal of Corpus Linguistics* 14.3 (2009): 312-337.

—. *Using corpora in discourse analysis*. Continuum, 2006.

"Banger - Facebook.com." 26 November 2015. *Let's say someone was going to make a Mathcore branch of the Metal Tree. And it looked like this. Agree? #askingforafriend.* 16 December 2015. <https://www.facebook.com/banger/photos/a.10152961867329286.1073741826.42842679285/10153692755049286/?type=3&theater>.

"Banger - Facebook.com." 12 November 2015. 28 March 2017. <https://www.facebook.com/banger/photos/a.10152961867329286.1073741826.42842679285/10153667327259286/?type=3>.

"Banger - Facebook.com." 19 November 2015. *New and improved Industrial Metal Chart! Thanks all for joining the Lock Horns live debate. Watch the archive anytime at: https://www.youtube.com/watch?v=kPiwPSh0ofU.* 16 December 2016. <https://www.facebook.com/banger/photos/a.10152961867329286.1073741826.42842679285/10153679506499286/?type=3&theater>.

"Banger - Facebook.com." 4 December 2015. *Presenting....Banger's first Folk Metal chart! Thanks to Natalie Zed and everyone who made their picks known. Great debate today! Watch it again here: http://bit.ly/BangerTV-FolkMetal.* 16 December 2015. <https://www.facebook.com/banger/photos/a.10152961867329286.1073741826.42842679285/10153704990069286/?type=3&theater>.

"Banger - Facebook.com." 10 December 2015. *The new improved Prog Metal chart! Thanks to Dan from Between The Buried And Me and everyone who joined us live. Archive is coming soon. We're just fixing some of the buffering buggery and will have a clean version up shortly.* 16 December 2015. <https://www.facebook.com/banger/photos/a.10152961867329286.1073741826.42842679285/10153717854639286/?type=3&theater>.

"Banger - Facebook.com." 12 November 2015. *Well that was fun. Thanks for participating in the Doom debate. But now we're even more confused! Gonna have to keep talking about this one before we start carving it in stone!* 16 December 2015. <https://www.facebook.com/banger/photos/a.10152961867329286.1073741826.42842679285/10153667327259286/?type=3&theater>.

"Banger - Facebook.com." 05 November 2015. *You spoke. We listened. Here's your new Metalcore branch of Banger's Heavy Metal Family Tree. Also introducing...Crossover! Thanks everyone who contributed on FB, YT, Twitter, etc. And yes, even Reddit. (OK, some of Reddit.).* 16 December 2015. <https://www.facebook.com/banger/photos/a.10152961867329286.1073741826.42842679285/10153655261599286/?type=3&theater>.

Bardel, Camilla, Anna Gudmundson and Christina Lindqvist. "Aspects of lexical sophistication in advanced learners' oral production: Vocabulary Acquisition and Use." *Studies in Second Language Acquisition* 34 (2012): 269-290.

Bauer, Laurie and Paul Nation. "Word Families." *International Journal of Lexicography* 6.4 (1993): 253-279.

Bauman, J. and B. Culligan. *About the General Service List*. 1995. 13 October 2016. <http://jbauman.com/aboutgsl.html>.

Beers, Scott F. and William E. Nagy. "Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre?" *Reading & Writing* 22.2 (2009): 185-200.

Berger, Harris M. *Metal, Rock, and Jazz: Perception and the Phenomenology of Musical Experience*. Hanover, NH: Wesleyan University Press, 1999.

Bértoli-Dutra, Patrícia. "Multi-Dimensional analysis of pop songs." *Multi-Dimensional Analysis, 25 Years On: A tribute to Douglas Biber*. Ed. Tony Berber Sardinha and Marcia Veirano Pinto. John Benjamins Publishing Company, 2014.

Biber, Douglas. *Dimensions of register variation: A cross-linguistic comparison*. New York: Cambridge University Press, 1995.

—. "Methodological Issues Regarding Corpus-based Analyses of Linguistic Variation." *Literary and Linguistic Computing* 5.4 (1990): 257–269.

—. "On the complexity of discourse complexity: A multidimensional analysis." *Discourse Processes* 15.2 (1992): 133-163.

—. "Representativeness in Corpus Design." *Literary and Linguistic Computing* 8.4 (1993): 243–257,.

—. "Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings." *Language* 62.2 (1986): 384-414.

—. *University language: A corpus-based study of spoken and written registers*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 2006.

—. *Variation across speech and writing*. New York: Cambridge University Press, 1988.

Biber, Douglas, Susan Conrad and Randi Reppen. *Corpus linguistics: Investigating language structure and use*. Cambridge University Press, 1998.

Breyer, Lisa and Michael Schemmann. "Comparing national adult education policies on a global level. An exploration of patterns of difference and similarity in international policy contexts by method of lexicometric analysis." *International Journal of Lifelong Education* 37.6 (2018): 749-762.

Brezina, Vaclav and Dana Gablasova. "Is There a Core General Vocabulary? Introducing the New General Service List." *Applied Linguistics* 36.1 (2013): 1–22.

Brown, Andy R. "Heavy Genealogy: Mapping the Currents, Contraflows and Conflicts of the Emergent Field of Metal Studies, 1978-2010." *Journal for Cultural Research* 15.3 (2011): 213-242.

Brown, Dale. "The Frequency Model of Vocabulary Learning and Japanese Learners." *Vocabulary Learning and Instruction* 1.1 (2012): 20-28.

Browne, Charles. "A new general service list: The better mousetrap we've been looking for." *Vocabulary Learning and Instruction* 3.2 (2014): 1-10.

Browne, Charlie, Brent Culligan and Joseph Phillips. *The New General Service List*. 2013. 08 July 2015. <http://www.newgeneralservicelist.org/>.

Brysbaert, Marc, Boris New and Emmanuel Keuleers . "Adding part-of-speech information to the SUBTLEX-US word frequencies." *Behavior Research Methods* 44 (2012): 991-997.

Buckethead - Encyclopaedia Metallum: The Metal Archives. n.d. 12 December 2017. <https://www.metal-archives.com/bands/Buckethead/3540384034>.

Bulté, Bram and Alex Housen. "Defining and operationalising L2 complexity." *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA*. Ed. Alex Housen, Folkert Kuiken and Ineke Vedder.

Amsterdam / Philadelphia: John Benjamins Publishing Company, 2012. 21-46.

Burns, Amelia Regina, Olimpia Matarazzo and Lucia Abbamonte. "Corpus Linguistics and the Appraisal Framework for Retrieving Emotion and Stance – The Case of Samsung's and Apple's Facebook Pages." *Recent Advances of Neural Network Models and Applications*. Ed. Simone Bassis, Anna Esposito and Francesco Carlo Morabito. Vols. Smart Innovation, Systems and Technologies. Springer, 2014. 26 vols. 283-293.

Buts, Jeroen and Gert Buelens. "The Thematical and Stylistic Evolution of Heavy Metal Lyrics and Imagery From the 70s to Present Day." Master's Thesis. UGent. Faculteit Letteren en Wijsbegeerte, 2008.

Caldwell-Harris, Catherine, Jonathan Berant and Shimon Edelman. "Measuring Mental Entrenchment of Phrases with Perceptual Identification, Familiarity Ratings, and Corpus Frequency Statistics." *Frequency Effects in Language Representation*. Ed. Dagmar Divjak and Stefan Th. Gries. De Gruyter Mouton, 2012.

Camiciottoli, Belinda Crawford. *The language of business studies lectures: a corpus-assisted analysis*. John Benjamins Publishing Company, 2007.

Campbell, Karlyn Kohrs and Kathleen Hall Jamieson. "Form and genre in rhetorical criticism: An introduction." *Form and genre: Shaping rhetorical action*. Ed. Karlyn Kohrs Campbell and Kathleen Hall Jamieson. Virginia: The Speech Communication Association, 1978. 9-32.

Carroll, John B. "Measurement properties of subjective magnitude estimates of word frequency." *Journal of Verbal Learning and Verbal Behavior* 10.6 (1971): 722-729.

Chandler, Daniel. "An Introduction to Genre Theory." 1997. 11 July 2017. <http://faculty.washington.edu/farkas/HCDE510-Fall2012/Chandler_genre_theoryDFAnn.pdf>.

Christe, Ian. *Sound of the beast: The complete headbanging history of heavy metal*. Epub edition. Harper Collins, 2010.

Chung, Teresa Mihwa and Paul Nation. "Technical vocabulary in specialised texts." *Reading in a Foreign Language* 15.2 (2003): 103-116.

Cobb, T. and Marlise Horst. "Is there room for an academic word list in French?" *Vocabulary in a Second Language: Selection, acquisition, and testing*. Ed. Paul Bogaards and Batia Laufer. Amsterdam: John Benjamins Publishing Company, 2004. 15-38.

Cobb, T. *Wb VP Classic [computer program]*. n.d. 04 November 2014. <http://www.lextutor.ca/vp/eng/>.

Cobb, Thomas. "FREQUENCY 2.0: Incorporating homoforms and multiword units in pedagogical frequency lists." *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis*. Ed. Camilla Bardel, Christina Lindqvist and Batia Laufer. European Second Language Association, 2013.

Corpus of American Soap Operas. n.d. 13 04 2019. <https://www.english-corpora.org/soap/>.

Corpus of Global Web-Based English (GloWbE). n.d. 13 04 2019. <https://www.english-corpora.org/glowbe/>.

Covington, Michael A. and Joe D. McFall. "Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR)." *Journal of Quantitative Linguistics* 17.2 (2010): 94-100.

Coxhead, Averil. "A New Academic Word List." *TESOL Quarterly* 34.2 (2000): 213-238.

—. "An Academic Word List." *ELI Occasional Publications* No. 18 (1998).

Coxhead, Averil and Murielle Demecheleer. "Investigating the technical vocabulary of Plumbing." *English for Specific Purposes* 51 (2018): 84-97.

Coxhead, Averil. "The Academic Word List 10 Years On: Research and Teaching Implications." *TESOL Quarterly* 45.2 (2011): 355-362.

Criado, Raquel and Aquilino Sánchez. "Lexical Frequency, Textbooks and Learning from a Cognitive Perspective. A Corpus-Based Sample Analysis of ELT Materials." *Revista Española de Lingüística Aplicada* 1 (2012): 77-94.

Crossley, Scott A. and Max Louwerse. "Multi-dimensional register classification." *International Journal of Corpus Linguistics* 12.4 (2007): 453–478.

Crossley, Scott A., Tom Cobb and Danielle S. McNamara. "Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications." *System* 41 (2013): 965-981.

Culpeper, Jonathan. "Keywords and Characterization: An Analysis of Six Characters in Romeo and Juliet." Hoover, David L., Jonathan Culpeper and Kieran O'Halloran. *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*. Routledge, 2014.

Daller, Helmut and David Phelan. "What is in a teacher's mind? Teacher ratings of EFL essays and different aspects of lexical richness." *Modelling and Assessing Vocabulary Knowledge*. Ed. Helmut Daller, James Milton and Jeanine Treffers-Daller. Cambridge University Press, 2007.

Daller, Helmut and Huijuan Xue. "Lexical richness and the oral proficiency of Chinese EFL students: A comparison of different measures." *Modelling and Assessing Vocabulary Knowledge*. Ed. Helmut Daller, James Milton and Jeanine Treffers-Daller. Cambridge: Cambridge University Press, 2007.

Daller, Helmut, Roeland van Hout and Jeanine Treffers-Daller. "Lexical richness in the spontaneous speech of bilinguals." *Applied Linguistics* 24.2 (2003): 197-222.

Dang, Thi Ngoc Yen and Stuart Webb. "Making an essential word list for beginners." *Making and Using Word Lists for Language Learning and Testing*. Ed. I. S. P. Nation. John Benjamins Publishing Company, 2016.

Davies, Mark and Dee Gardner. *A Frequency Dictionary of Contemporary American English: Word sketches, collocates, and thematic lists*. Routledge, 2010.

Davies, Mark. "The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights." *International Journal of Corpus Linguistics* 14.2 (2009): 159-190.

—. *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. 2008-. 09 October 2017. <https://corpus.byu.edu/coca/>.

Davisson, Justin. "Extreme Politics and Extreme Metal: Strange Bedfellows or Fellow Travellers." *The Metal Void: First Gatherings*. Ed. Niall W. R. Scott. Oxford: Inter-Disciplinary Press, 2010. 175-210.

De Jong, Nivja H., et al. "The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers." *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Ed. Alex Housen, Folkert Kuiken and Ineke Vedder. John Benjamins Publishing Company, 2012.

Devitt, Amy J. "Generalizing about Genre: New Conceptions of an Old Concept." *College Composition and Communication* 44.4 (1993): 573-586.

DiMaggio, Paul. "Classification in Art." *American Sociological Review* 52.4 (1987): 440-455.

Dukes, Richard L., et al. "Expressions of love, sex, and hurt in popular songs: a content analysis of all-time greatest hits." *The Social Science Journal* 40.4 (2003): 643-650.

Eggins, Suzanne. *Introduction to systemic functional linguistics*. London: Continuum International Publishing Group, 2004.

Ehara, Yo, et al. "Formalizing Word Sampling for Vocabulary Prediction as Graph-based Active Learning." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2014. 1374-1384.

Eiter, Alexander. "'Haters gonna Hate': A Corpus Linguistic Analysis of the Use of Non-Standard English in Pop Songs." University of Innsbruck, 2017. Bachelor Thesis.

Ellis, Rod and Fangyuan Yuan. "The effects of careful within-task planning on oral and written task performance." *Planning and Task Performance in a Second Language*. Ed. Rod Ellis. John Benjamins Publishing Company, 2005.

—. "The effects of planning on fluency, complexity, and accuracy in second language narrative writing." *Studies in Second Language Acquisition* 26.1 (2004): 59-84.

Engber, Cheryl A. "The relationship of lexical proficiency to the quality of ESL compositions." *Journal of Second Language Writing* 4.2 (1995): 139-155.

English Web Corpora (Ententen Corpus). n.d. 12 July 2017. <https://www.sketchengine.co.uk/ententen-corpus/>.

Falk, Johanna. "We will rock you: A diachronic corpus-based analysis of linguistic features in rock lyrics." Linnaeus University, 2012.

Fast, Susan. *In the Houses of the Holy: Led Zeppelin and the Power of Rock Music*. Oxford: Oxford University Press, 2001.

Feinberg, Jonathan. *Wordle*. 2008. 08 July 2017. <http://www.wordle.net/>.

Fell, Michael and Caroline Sporleder. "Lyrics-based Analysis and Classification of Music." *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics, August 23-29, 2014*. Dublin, 2014. 620-631.

Fergadiotis, Gerasimos. "Modeling lexical diversity across language sampling and estimation techniques." Arizona State University, 2011. PhD Dissertation.

Fergadiotis, Gerasimos, Heather H. Wright and Thomas M. West. "Measuring Lexical Diversity in Narrative Discourse of People With Aphasia." *American Journal of Speech-Language Pathology* 22.2 (2013): 397-408.

Fergadiotis, Gerasimos, Heather Harris Wright and Samuel B. Green. "Psychometric Evaluation of Lexical Diversity Indices: Assessing Length Effects." *Journal of Speech, Language, and Hearing Research* 58.3 (2015): 840-852.

"File:Metal Genealogy.jpg." n.d. *Wikimedia Commons, the free media repository.* 05 August 2016. <https://commons.wikimedia.org/w/index.php?title=File:Metal_Genealogy.jpg&oldid=226286866>.

Findler, Nicholas V. and Heino Viil. "A Few Steps Towards Computer Lexicometry." *American Journal of Computational Linguistics* (1974).

Fish, Stanley Eugene. *Is There a Text in This Class?: The Authority of Interpretive Communities*. Cambridge, MA: Harvard University Press, 1980.

Fitzpatrick, Tess. "Productive vocabulary tests and the search for concurrent validity." *Modelling and Assessing Vocabulary Knowledge*. Ed. Helmut Daller, James Milton and Jeanine Treffers-Daller. Cambridge: Cambridge University Press, 2007. 116-133.

Forsyth, Richard S. "Pops and Flops: Some Properties of Famous English Poems." *Empirical Studies of the Arts* 18.1 (2000): 49-67.

Frandsen, Daniel. "Two Steps Past Insanity: The Expression of Aggression in Death Metal Music." *Can I Play with Madness? Metal, Dissonance, Madness and Alienation*. Ed. Colin A. McKinnon, Niall Scott and Kristen Sollee. Oxford: Inter-Disciplinary Press, 2011. 35-40.

Fraysse-Kim , Soon Hee. "Keywords in Korean national consciousness: A corpus-based analysis of school textbooks." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. 2010.

Frear, Mark Wain and John Bitchener. "The effects of cognitive task complexity on writing complexity." *Journal of Second Language Writing* 30 (2015): 45-57.

Gabrielatos , Costas. "Keyness Analysis: nature, metrics and techniques." (2018).

Gardner, Dee and Mark Davies. "A new academic vocabulary list." *Applied Linguistics* 35.1 (2013): 305-327.

Garnier, Mélodie and Norbert Schmitt. "The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses." *Language Teaching Research* 6 (2015): 645-666.

Gellerstam, Martin. "Modern Swedish Text Corpora." *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Ed. Jan Svartvik. Berlin: De Gruyter, 1992.

Genre. n.d. 02 May 2014. <http://www.merriam-webster.com/dictionary/genre>.

Gerbig, Andrea. "Key words and key phrases in a corpus of travel writing: From Early modern English literature to contemporary “blooks”." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. John Benjamins Publishing Company, 2010.

Gilabert, Roger. "The Simultaneous Manipulation of Task Complexity Along Planning Time and [+/-Here-and-Now]: Effects on L2 Oral Production." *Investigating Tasks in Formal Language Learning*. Ed. María del Pilar García Mayo. Multilingual Matters Ltd, 2007.

Gilner, Leah. "A primer on the General Service List." *Reading in a Foreign Language* 23.1 (2011): 65-83.

Gilner, Leah and Franc Morales. "Corpus-Based Frequency Profiling: Migration to a Word List Based on the British National Corpus." *The Buckingham Journal of Language and Linguistics* 1 (2008a): 41-57.

—. "Elicitation and application of a phonetic description of the General Service List." *System* 36.4 (2008b): 517-533.

Glukhova, L. "Semantic Structure of the Lexical Unit Educational Management in Modern English (on the basis of corpora BNC and COCA)." *The Advanced Science Journal* (2014): 245-248.

Gracyk, Theodore. "Heavy metal: Genre? Style? Subculture?" *Philosophy Compass* 11.12 (2016): 775–785.

Graesser, Arthur C., et al. "Advanced Outlines, Familiarity, and Text Genre on Retention of Prose." *The Journal of Experimental Education* 48.4 (1980): 281-290.

—. "Coh-Metrix: Analysis of text on cohesion and language." *Behavior Research Methods, Instruments, & Computers* 36 (2004): 193-202.

Grant, Lynn E. "Frequency of 'core idioms' in the British National Corpus (BNC)." *International Journal of Corpus Linguistics* 10.4 (2005): 429-451.

Gregori-Signes, Carmen and Begoña Clavel-Arroitia. "Analysing lexical density and lexical diversity in university students' written discourse." *Procedia - Social and Behavioral Sciences* 198 (2015): 546-556.

Groom, Nicholas. "Closed-class keywords and corpus-driven discourse analysis." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. John Benjamins Publishing Company, 2010.

Guo, Liang, Scott A. Crossley and Danielle S. McNamara. "Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study." *Assessing Writing* 18.3 (2013): 218-238.

Halliday, M.A.K. and Christian M.I.M. Matthiessen. *An introduction to functional grammar*. 3rd Edition. Hodder Arnold, 2004.

Halliday, M.A.K. and J.R. Martin. *Writing science: Literacy and discursive power*. The Falmer Press, 1993.

Halliday, M.A.K. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold Publishers, 1978.

Halliday, M.A.K. "Lexicology." *Lexicology and Corpus Linguistics: An Introduction*. Ed. M.A.K. Halliday, et al. London: Continuum, 2004. 1-22.

—. *Spoken and written language*. Ed. Frances Christie. Oxford: Oxford University Press, 1989.

Hanafiah, Ridwan and Muhammad Yusuf. "Lexical density and grammatical intricacy in linguistic thesis abstract: A qualitative content analysis." *Proceedings of English Education International Conference* 1.2 (2016): 43-46.

Hancıoğlu, Nilgün, Steven Neufeld and John Eldridge. "Through the looking glass and into the land of lexico-grammar." *English for Specific Purposes* 27.4 (2008): 459-479.

Handford, Michael. *Professional Communication and Corpus Linguistics*. Ed. Ken Hyland, Chau Meng Huat and Michael Handford. Continuum, 2012.

Handford, Michael. "What corpora have to tell us about specialised genres." *The Routledge Handbook of Corpus Linguistics*. Ed. Anne O'Keeffe and Michael McCarthy. 2010.

Hardy, Jack A. "Multi-Dimensional Analysis of Academic Discourse." *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Ed. Paul Baker and Tony McEnery. Palgrave Macmillan, 2015.

Hasan, Ali Sa'ud. "Variation in spoken discourse in and beyond the english foreign language classroom : a comparative study." Aston University, 1988. PhD thesis.

Hayes, Donald P., Loreen T. Wolfer and Michael F. Wolfe. "Schoolbook Simplification and Its Relation to the Decline in SAT-Verbal Scores." *American Educational Research Journal* 33.2 (1996): 489-508.

Heatley, Alex, I.S.P. Nation and Averil Coxhead. *RANGE and FREQUENCY programs*. 2002. 10 October 2017. <https://www.victoria.ac.nz/lals/about/staff/paul-nation>.

Herdan, Gustav. *The advanced theory of language as choice and chance*. Berlin: Springer, 1966.

Hickam, Brian and Jeremy Wallach. "Female Authority and Dominion: Discourse and Distinctions of Heavy Metal Scholarship." *Journal for Cultural Research* 15.3 (2011): 255-277.

Hickam, Brian and Thomas Atwood. "Filling the Void: The Heavy Metal and Hardcore Punk Archiving Project." *Heavy Fundametalisms: Music, Metal and Politics Conference, 04 November 2008*. Salzburg, Austria, 2008.

Hinds, Elizabeth Jane Wall. "The devil sings the blues: Heavy metal, Gothic fiction and "postmodern" discourse." *Journal of Popular Culture* 26.3 (1992): 151-164.

Hlaváčová, Jaroslava. "New Approach to Frequency Dictionaries - Czech Example." *The 5th international conference on Language Resources and Evaluation, LREC*. Genoa, 22-28 May 2006, 2006. 373-378.

Hoffmann, Sebastian and Hans-Martin Lehman. "Collocational evidence from the British National Corpus." *Corpora galore: analyses and techniques in describing English*. Ed. John M. Kirk. Amsterdam: Rodopi, 2000.

Holmes, D. I. "A Stylometric Analysis of Mormon Scripture and Related Texts." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 155.1 (1992): 91-120.

Holmes, David I. "Authorship Attribution." *Computers and the Humanities* 28 (1994): 87-106.

Holmes, Richard. "Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines." *English for Specific Purposes* 16.4 (1997): 321-337.

Hoover, David L. "Another Perspective on Vocabulary Richness." *Computers and the Humanities* 37.2 (2003): 151-178.

Hsu, Wenhua. "The most frequent BNC/COCA mid- and low-frequency word families in English-medium traditional Chinese medicine (TCM) textbooks." *English for Specific Purposes* 51 (2018): 98-110.

—. "The most frequent opaque formulaic sequences in English-medium college textbooks." *System* 47 (2014): 146-161.

Hsueh-Chao, Marcella Hu and Paul Nation. "Unknown Vocabulary Density and Reading Comprehension." *Reading in a Foreign Language* 13.1 (2000): 403-430.

Hunston, Susan. "Afterword: The Problems of Applied Linguistics." *Corpus Applications in Applied Linguistics*. Ed. Ken Hyland, Chau Meng Huat and Michael Handford. Continuum, 2012.

Hutcherson, Ben and Ross Haenfler. "Musical genre as a gendered process: Authenticity in extreme metal." *Studies in Symbolic Interaction*. Ed. Norman K. Denzin. Vol. 35. Emerald Group Publishing Limited, 2010. 101-121.

Hyland, Ken and Polly Tse. "Is There an "Academic Vocabulary"?" *TESOL Quarterly* 41.2 (2007): 235-253.

Hyland, Ken. "Corpora and Academic Discourse." *Corpus Applications in Applied Linguistics*. Ed. Ken Hyland, Chau Meng Huat and Michael Handford. Continuum, 2012.

Hyland, Ken. "Genre and discourse analysis in language for specific purposes." *The encyclopaedia of applied linguistics*. Ed. Carol Chapelle. Oxford: Wiley, 2012.

Hyltenstam , Kenneth. "Lexical characteristics of near-native second-language learners of Swedish." *Journal of Multilingual and Multicultural Development* 9.1 (1988): 67-84.

Ignat, Camelia, et al. "A tool set for the quick and efficient exploration of large document collections." *Proceedings of the Symposium on Safeguards and Nuclear Material Management* (2006).

Ishikawa, Tomohito. "The Effect of Manipulating Task Complexity Along the [±Here-and-Now] Dimension on L2 Written Narrative Discourse."

Investigating Tasks in Formal Language Learning*. Ed. María del Pilar García Mayo. Multilingual Maters Ltd, 2007.

iWeb: The 14 Billion Word Web Corpus. 05 2018. 13 04 2019. <https://www.english-corpora.org/iweb/>.

Jarvis, Scott. "Short texts, best-fitting curves and new measures of lexical diversity." *Language Testing* 19.1 (2002): 57-84.

Jarvis, Scott, et al. "Exploring multiple profiles of highly rated learner compositions." *Journal of Second Language Writing* 12.4 (2003): 377-403.

Johansson, Stig. "Some Aspects of the Vocabulary of Learned and Scientific English." *Gothenburg Studies in English 42* (1978).

Johansson, Victoria. "Lexical diversity and lexical density in speech and writing: a developmental perspective." *Working Papers, Department of Linguistics and Phonetics, Lund University* 53 (2008): 61-79.

Johnson, Mark D. "Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis." *Journal of Second Language Writing* 37 (2017): 13-38.

—. "Planning in second language writing." Northern Arizona University, 2011. PhD Dissertation.

Kahn-Harris, Keith. "Roots'?: The Relationship between the Global and the Local within the Extreme Metal Scene." *Popular Music* 19.1 (2000): 13-30.

—. *The Metal Future*. 20 December 2013. 13 October 2016. <http://souciant.com/2013/12/the-metal-future/>.

—. *Too Much Metal*. 29 November 2013. 17 October 2016. <http://souciant.com/2013/11/too-much-metal/>.

Kahraman, Volkan. *Analysis of Songs and Implications for English Language Teaching*. Master's Thesis. Uludağ Üniversitesi. Bursa, 2009.

Katagiri, Noriaki and Goh Kawai. "Lexical Analyses of Native and Non-Native English Language Instructor Speech Based on a Six-Month Co-Taught Classroom Video Corpus." *INTERSPEECH 2008 9th Annual Conference of the International Speech Communication Association*. Brisbane, Australia, 2008. 1984-1987.

Kennedy, Graeme. *An introduction to corpus linguistics*. Longman, 1998.

Kennedy, Graeme. "Preferred ways of putting things with implications for language teaching." *Directions in corpus linguistics: proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Ed. Jan Svartvik. Berlin, 1992.

Kilgarriff, Adam. "Using Word Frequency Lists to Measure Corpus Homogeneity and Similarity between Corpora." *Language Engineering for Document Analysis and Recognition. Proceedings, AISB Workshop*. Falmer, 1997.

Kirkness, Alison and Lindsay Neill. "Choosing texts for today's students: Do they understand language?" *Journal of Hospitality, Leisure, Sports and Tourism Education (Pre-2012)* 8.2 (2009): 4-16.

Kitteringham, Sarah. "Extreme Conditions Demand Extreme Responses: The Treatment of Women in Black Metal, Death Metal, Doom Metal, and Grindcore." Master’s Thesis. University of Calgary, 2014.

Klee, Thomas, et al. "Utterance Length and Lexical Diversity in Cantonese-Speaking Children With and Without Specific Language Impairment." *Journal of Speech, Language, and Hearing Research* 47.6 (2004): 1396–1410.

Klepper, Stefan de, et al. *Cultural heritage and history in the European metal scene*. Utrecht: Belvedere Educational Network [etc.], 2007.

Kneer, Julia, et al. "Goethe vs. Rammstein: Who is Allowed to Play with Madness? The Influence of Musical Taste on Prejudice against Heavy Metal Lyrics." *Can I Play with Madness? Metal, Dissonance, Madness and Alienation* (2011): 3-14.

Knees, Peter, Markus Schedl and Gerhard Widmer. "Multiple Lyrics Alignment: Automatic Retrieval of Song Lyrics." *6th International Conference on Music Information Retrieval (ISMIR 2005), 11 - 15 September 2005*. London, UK, 2005. 564-569.

Koester, Almut. "Corpora and Workplace Discourse." *Corpus Applications in Applied Linguistics*. Ed. Ken Hyland, Chau Meng Huat and Michael Handford. 2012.

Koizumi, Rie and Yo In'nami. "Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens." *System* 40.4 (2012): 554-564.

Kojima, Masumi and Junko Yamashita. "Reliability of lexical richness measures based on word lists in short second language productions." *System* 42 (2014): 23-33.

Konow, David. *Bang Your Head: The Rise and Fall of Heavy Metal*. NY: Three Rivers Press, 2002.

Kormos, Judit and Zoltán Dörnyei. "The interaction of linguistic and motivational variables in second language task performance." *Zeitschrift für interkulturellen Fremdsprachenunterricht* 9.2 (2004).

Kormos, Judit. "Task complexity and linguistic and discourse features of narrative writing performance." *Journal of Second Language Writing* 20.2 (2011): 148-161.

Kreyer, Rolf and Joybrato Mukherjee. "The Style of Pop Song Lyrics: A Corpus-linguistic Pilot Study." *Anglia - Journal of English Philology* 125.1 (2007): 31-58.

Kuiken, Folkert and Ineke Vedder. "Cognitive task complexity and written output in Italian and French as a foreign language." *Journal of Second Language Writing* 17.1 (2008): 48-60.

—. "Task complexity and measures of linguistic performance in L2 writing." *International Review of Applied Linguistics in Language Teaching* 45.3 (2007): 261-284.

Kuiken, Folkert, Maria Mos and Ineke Vedder. "Cognitive task complexity and second language writing performance." *EUROSLA Yearbook* 5.1 (2005): 195-222.

Kübler, Sandra and Heike Zinsmeister. *Corpus Linguistics and Linguistically Annotated Corpora*. Bloomsbury, 2015.

Kwary, Deny A. and Almira F. Artha. "The Academic Article Word List for Social Sciences." *MEXTESOL Journal* 41.4 (2017).

Kwary, Deny A. and Jurianto. "Selecting and creating a word list for english language teaching." *Teaching English with Technology* 17.1 (2017): 60-72.

Laufer, Batia and Paul Nation. "Vocabulary Size and Use: Lexical Richness in L2 Written Production." *Applied Linguistics* 16.3 (1995): 307-322.

Laufer, Batia. "Lexical Frequency Profiles." *The Encyclopedia of Applied Linguistics*. Ed. Carol A. Chapelle. Wiley, 2012.

—. "Lexical Frequency Profiles: From Monte Carlo to the Real World : A Response to Meara (2005)." *Applied Linguistics* 4.1 (2005): 582-588.

Laufer, Batia. "What percentage of text-lexis is essential for comprehension?" *Special Language: From Humans Thinking to Thinking Machines Multilingual matters*. Ed. Christer Laurén and Marianne Nordman. Multilingual Matters, 1989. 316-323.

Lee, David Y. W. "Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC." *Language Learning and Technology* 5.3 (2001): 37-72.

Leech, Geoffrey and Roger Fallon. "Computer corpora – what do they tell us about culture?" *ICAME Journal* 16 (1992): 29-50.

Leech, Geoffrey. "Corpora and theories of linguistic performance." *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4 - 8 August 1991*. Ed. Jan Svartvik. 1992.

Leech, Geoffrey, Paul Rayson and Andrew Wilson. *Word Frequencies in Written and Spoken English based on the British National Corpus*. Routledge, 2001.

Leech, Geoffrey, Stig Johansson and Knut Hofland. *The Lancaster-Oslo/Bergen Corpus (LOB Corpus)*. 1970-1978. University of Oslo Department of English. 12 October 2017. <http://www.helsinki.fi/varieng/CoRD/corpora/LOB/index.html>.

Lei, Lei and Dilin Liu. "A new medical academic word list: A corpus-based study with enhanced methodology." *Journal of English for Academic Purposes* 22 (2016): 42-53.

León, María Araceli Losey. "Corpus-based Contrastive Analysis of Keywords and Collocations across Sister Specialized Subcorpora in the Maritime Transport Field." *Procedia - Social and Behavioral Sciences* 198 (2015): 526-534.

Leone, Paola. "General spoken language and school language: Key words and discourse patterns in history textbooks." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. John Benjamins Publishing Company, 2010.

Levkina, Mayya and Roger Gilabert. "The effects of cognitive task complexity on L2 oral production." *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Ed. Alex Housen, Folkert Kuiken and Ineke Vedder. John Benjamins Publishing Company, 2012.

Li, Yili. "Linguistic characteristics of ESL writing in task-based e-mail activities." *System* 28.2 (2000): 229-245.

Li, Yongyan and David D. Qian. "Profiling the Academic Word List (AWL) in a financial corpus." *System* 38.3 (2010): 402-411.

Lieb, Maggie. "English Through Popular Songs: Reinforcing High-frequency Vocabulary and Enhancing All Four Skills." *Teachers Helping Teachers, The Proceedings of 2011 Conferences, Seminars and Workshops*. Ed. Brent Jones and Richard Silver. Japan Association for Language Teaching (JALT), 2011. 19-32.

Lightman, Erin J., et al. "Using Computational Text Analysis Tools to Compare the Lyrics of Suicidal and Non-Suicidal Songwriters." *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2007. 1217-1222.

Linnarud, Moira. "Lexis in composition: A performance analysis of Swedish learners' written English." *CWK Gleerup* 74 (1986).

Lissón, Paula and Nicolas Ballier. "Investigating Lexical Progression through Lexical Diversity Metrics in a Corpus of French L3." *Discours* 23 (2018).

Liu, Dilin. "The Most Frequently Used English Phrasal Verbs in American and British English: A Multicorpus Examination." *TESOL Quarterly* 45.4 (2011): 661-688.

Logan, Beth, Andrew Kositsky and Pedro Moreno. "Semantic analysis of song lyrics." *2004 IEEE International Conference on Multimedia and Expo (ICME) 27-30 June 2004*. Taipei, Taiwan, 2004. 827-830.

Lu, Xiaofei. "A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development." *TESOL Quarterly* 45.1 (2011): 36-62.

—. "The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives." *The Modern Language Journal* 96.2 (2012): 190-208.

Lu, Xiaofei, David Alexander Gamson and Sarah Anne Eckert. "Lexical difficulty and diversity of American elementary school reading textbooks: Changes over the past century." *International Journal of Corpus Linguistics* 19.1 (2014): 94-117.

Maas, Heinz-Dieter. "Zusammenhang zwischen Wortschatzumfang und Länge eines Textes." *Zeitschrift für Literaturwissenschaft und Linguistik* 2.8 (1972): 73-79.

Malvern, David and Brian Richards. "Measures of Lexical Richness." *The Encyclopedia of Applied Linguistics*. Ed. Carol A. Chapelle. Blackwell Publishing Ltd., 2012.

Malvern, David, et al. *Lexical Diversity and Language Development: Quantification and Assessment*. Palgrave Macmillan, 2004.

Martinez, Ron. "A framework for the inclusion of multi-word expressions in ELT." *ELT Journal* 67.2 (2013): 184-198.

Martinez, Ron and Norbert Schmitt. "A Phrasal Expressions List." *Applied Linguistics* 33.3 (2012): 299-320.

Masrai, Ahmed and James Milton. "Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement." *Journal of English for Academic Purposes* 31 (2018): 44-57.

Matthews, Joshua. "Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge." *System* 72 (2017): 23-36.

Mayer, Rudolf, Robert Neumayer and Andreas Rauber. "Rhyme and Style Features for Musical Genre Classification by Song Lyrics." *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR), 14-18 September 2008*. Philadelphia, PA, USA, 2008. 337-342.

McCarthy, Michael and Jeanne McCarten. "Corpora and Materials Design ." *C orpus A pplications in Applied Linguistics*. Ed. Ken Hyland, Chau Meng Huat and Michael Handford. Continuum, 2012.

McCarthy, Philip M. and Scott Jarvis. "MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment." *Behavior Research Methods* 42 (2010): 381-392.

—. "vocd: A theoretical and empirical evaluation." *Language Testing* 24.4 (2007): 459-488.

McCauley, Stewart M. and Morten H. Christiansen. "Individual Differences in Chunking Ability Predict On-line Sentence Processing." *CogSci* (2015): 1553-1558.

McCrostie, James. "Investigating the Accuracy of Teachers' Word Frequency Intuitions." *RELC Journal* 38.1 (2007): 53-66.

McEnery, Tony and Andrew Hardie. *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press, 2012.

McEnery, Tony, Richard Xiao and Yukio Tono. *Corpus-based Language Studies: An Advanced Resource Book*. Routledge, 2006.

McLean, Stuart and Brandon Kramer. "The Creation of a New Vocabulary Levels Test." *Shiken* 19.2 (2015): 1-11.

McNamara, Danielle S. and Arthur C. Graesser. "Coh-Metrix: An Automated Tool for Theoretical and Applied Natural Language Processing." *Applied Natural Language Processing: Identification, Investigation and Resolution*. Ed. Philip M. McCarthy and Chutima Boonthum-Denecke. IGI Global, 2012.

McNamara, Danielle S., Yasuhiro Ozuru and Randy G. Floyd. "Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge." *International Electronic Journal of Elementary Education* 4.1 (2011): 229-257.

Meara, Paul and Huw Bell. "P-Lex: A Simple and Effective Way of Describing the lexical Characteristics of Short L2 Tests." *Prospect* 16.3 (2001): 5-19.

Meara, Paul. "Lexical Frequency Profiles: A Monte Carlo Analysis." *Applied Linguistics* 26.1 (2005): 32-47.

—. "P_Lex v.3.0." 2018.

—. "Tintin and the World Service: A look at lexical environments." *IATEFL Annual Conference Report*. 1993. 32-37.

Medero, Julie and Mari Ostendorf. "Identifying Targets for Syntactic Simplification." *ISCA International Workshop on Speech and Language Technology in Education (SLaTE 2011)*. 2011. 69-72.

Meyer, Charles F. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press, 2002.

Meylan, Stephan and Susanne Gahl. "The Divergent Lexicon: Lexical Overlap Decreases With Age in a Large Corpus of Conversational Speech." *Proceedings of the Annual Meeting of the Cognitive Science Society* 36.36 (2014): 1006-1011.

Michalke, M. "koRpus." *Entschuldigen Sie, dass ich Ihnen einen komplizierten Artikel schreibe, für einen lesbaren habe ich keine Zeit" -- Textanalyse mit den R-Paketen koRpus & tm.plugin.koRpus*. Marburg: Paper presented at the Tagung experimentell arbeitender Psychologen (TeaP), 2018.

Miethaner, Ulrich. *I Can Look Through Muddy Water: Analyzing Earlier African American English in Blues Lyics (BLUR)*. Frankfurt a.M.: Peter Lang, 2005.

Milizia , Denise. "Keywords and phrases in political speeches." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. John Benjamins Publishing Company, 2010.

Miller, Carolyn R. "Genre as social action." *Quarterly Journal of Speech* 70.2 (1984): 151-167.

Miller, Don and Douglas Biber. "Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition." *International Journal of Corpus Linguistics* 20.1 (2015): 30–53.

Miller, G. A., E. B. Newman and E. A. Friedman. "Length-frequency statistics for written English." *Information and Control* 1.4 (1958): 370-389.

Miller, Jim. *An introduction to English syntax*. Edinburgh: Edinburgh University Press, 2002.

Mishrell, Kirk W. "Rockin' The Tritone: Gender, Race & The Aesthetics of Aggressive Heavy Metal Subcultures." Master's Thesis. Georgia State University, 2012.

Mittell, Jason. "A Cultural Approach to Television Genre Theory." *Cinema Journal* 40.3 (2001): 3-24.

Moberg, Marcus. "Faster for the master!: exploring issues of religious expression and alternative Christian identity within the Finnish Christian metal music scene." Doctoral Dissertation. 2009.

Moessner, Lilo. "Genre, Text Type, Style, Register: A Terminological Maze?" *European Journal of English Studies* 5.2 (2001): 131-138.

Moore, Doug. *Death Metal English*. 08 November 2013. 28 October 2014. <http://www.invisibleoranges.com/death-metal-english>.

Morford, Jill Patterson and James MacFarlane. "Frequency Characteristics of American Sign Language." *Sign Language Studies* 3.2 (2003): 213-225.

Morris, Lori and Tom Cobb. "Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees." *System* 32.1 (2004): 75-87.

Morrison, Craig. *American Popular Music: Rock and Roll*. NY: Facts On File, Inc., 2006.

Mudrian, Albert and John Peel. *Choosing Death: The Improbable History of Death Metal and Grindcore*. Los Angeles, CA: Feral House, 2004.

Muncie, James. "Process writing and vocabulary development: comparing Lexical Frequency Profiles across drafts." *System* 30.2 (2002): 225-235.

Murphey, Tim. "The Discourse of Pop Songs." *TESOL Quarterly* 26.4 (1992): 770-774.

Nation, Paul. "A study of the most frequent word families in the British National Corpus." *Vocabulary in a Second Language: Selection, acquisition, and* *testing*. Ed. Paul Bogaards and Batia Laufer. Amsterdam: John Benjamins Publishing Company, 2004. 3-13.

Nation, Paul and Hwang Kyongho. "Where would general service vocabulary stop and special purposes vocabulary begin?" *System* 23.1 (1995): 35-41.

Nation, Paul and Laurence Anthony. "Mid-frequency readers." *Journal of Extensive Reading* 1 (2013): 5-16.

Nation, Paul and Robert Waring. "Vocabulary size, text coverage and word lists." *Vocabulary: Description, Acquisition and Pedagogy*. Ed. Norbert Schmitt and Michael McCarthy. Cambridge: Cambridge University Press, 1997.

Nation, Paul and Stuart Webb. *Researching and Analyzing Vocabulary*. Boston: Heinle, Cencage Learning, 2011.

Nation, Paul. "Fundamental issues in modelling and assessing vocabulary knowledge." *Modelling and Assessing Vocabulary Knowledge*. Ed. Helmut Daller, James Milton and Jeanine Treffers-Daller. Cambridge: Cambridge University Press, 2007. 35-43.

—. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press, 2001.

Nesi, Hilary. "An English Spoken Academic Wordlist." *Proceedings of the 10th EURALEX International Congress*. Ed. Anna Braasch and Claus Povlsen. Copenhagen, Denmark: Center for Sprogteknologi, 2002. 351-357.

Neufeld, Steve and Ali Billuroğlu. "In search of the critical lexical mass: How 'general' is the GSL? How 'academic' is the AWL?" 2005. *Lextutor.ca.* 11 November 2014. <http://www.lextutor.ca/vp/tr/BNL_Rationale.doc>.

Newman, John and Sally Rice. "Transitivity schemas of English EAT and DRINK in the BNC." *Corpora in Cognitive Linguistics*. Ed. Stefan Th. Gries and Anatol Stefanowitsch. Berlin: Mouton de Gruyter, 2006. 225-260.

Nguyen, Ngoc To Ngan and Thi Thu Thao Nguyen. "Lexical Density and Readability of Non-English Majored Freshmen's Writing in Vietnamese Context." (2016).

Nippold, Marilyn A., et al. "Conversational Versus Expository Discourse: A Study of Syntactic Development in Children, Adolescents, and Adults." *Journal of Speech, Language, and Hearing Research* 48.5 (2005): 1048-1064.

NOW Corpus (News on the Web). n.d. 13 04 2019. <https://www.english-corpora.org/now/>.

O'Donnell, Matthew Brook. "The adjusted frequency list: A method to produce clustersensitive frequency lists." *ICAME Journal* 35 (2011): 135-169.

O'Keeffe, Anne. "Corpora and Media Studies." *Corpus Applications in Applied Linguistics*. Ed. Ken Hyland, Chau Meng Huat and Michael Handford. Continuum, 2012.

O'Donnell, Mick. "UAM Corpus Tool v. 3.3." Madrid: Universidad Autónoma de Madrid, 2016. Computer Application.

Okamoto, Mayumi. "Is corpus word frequency a good yardstick for selecting words to teach? Threshold levels for vocabulary selection." *System* 51 (2015): 1-10.

Olinghouse, Natalie G. and Joshua Wilson. "The relationship between vocabulary and writing quality in three genres." *Reading and Writing* 26 (2013): 45-65.

Olivo, Warren. "Phat Lines: Spelling Conventions in Rap Music." *Written Language & Literacy* 4.1 (2001): 67-85.

O'Loughlin, Kieran. "Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test." *Language Testing* 12.2 (1995): 217-237.

Ong, Justina and Lawrence Jun Zhang. "Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing." *Journal of Second Language Writing* 19.4 (2010): 218-233.

Öztürk, Meral. "Vocabulary Load of English Song Lyrics for EFL Learners." *Journal of Foreign Language Education and Technology* 2.2 (2017): 54-73.

Pachet, François and Daniel Cazaly. "A taxonomy of musical genres." *Proceedings of Content-Based Multimedia Information Access (RIAO '00)*. Paris: Le Centre de Hautes Etudes Internationales d'Informatique Documentaire (C.I.D.), 2000. 1238-1245.

Pallotti, Gabriele. "CAF: Defining, Refining and Differentiating Constructs." *Applied Linguistics* 30.4 (2009): 590-601.

Park, Soo-Kyung. "Lexical Analysis of Korean University Students' Narrative and Argumentative Essays." *English Teaching* 68.3 (2013): 131-157.

Pennebaker, J. W., et al. "Linguistic Inquiry and Word Count: LIWC2015." Austin, TX: Pennebaker Conglomerates (www.LIWC.net), 2015.

Perrone, Pierre. *Paul Gray: Bassist and songwriter with controversial nu-metal band Slipknot*. 30 May 2010. 13 October 2017. <http://www.independent.co.uk/news/obituaries/paul-gray-bassist-and-songwriter-with-controversial-nu-metal-band-slipknot-1987378.html>.

Petrie, Keith J., James W. Pennebaker and Borge Sivertsen. "Things we said today: A linguistic analysis of the Beatles." *Psychology of Aesthetics, Creativity, and the Arts* 2.4 (2008): 197-202.

Pettijohn, Terry F. and Donald F. Sacco. "The Language of Lyrics: An Analysis of Popular Billboard Songs Across Conditions of Social and Economic Threat." *Journal of Language and Social Psychology* 28.3 (2009): 297-311.

Philip, Gill. "Metaphorical keyness in specialised corpora." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. John Benjamins Publishing Company, 2010. 185-203.

Phillipov, Michelle. *Death Metal and Music Criticism: Analysis at the Limits*. Plymouth: Lexington Books, 2012.

Phillips, William and Brian Cogan. *Encyclopedia of heavy metal music*. Westport, CT: Greenwood Press, 2009.

Piccoli, Sean. "Extreme-metal bands come in a variety of alloys." 23 November 2003. *Sun-sentinel.com.* 23 April 2010. <http://proquest.umi.com/pqdlink?did=460436781&Fmt=7&clientId=42772&RQT=309&VName=PQD>.

Pieslak, Jonathan. "Re-casting Metal: Rhythm and Meter in the Music of Meshuggah." *Music Theory Spectrum* 29.2 (2007): 219-245.

Pillsbury, Glenn T. *Damage Incorporated: Metallica and the Production of Musical Identity*. New York: Routledge, 2006.

Pojanapunya, Punjaporn and Richard Watson Todd. "Lexical Differences and Similarities in the Humanities and the Sciences." *Proceedings from The 7th International Conference on Language and Communication ICLC2015 "Advancing Research in Language, Communication and Culture: Towards a Sustainable Asia"*. Osaka, 2015. 156-171.

Polio, Charlene G. "Measures of Linguistic Accuracy in Second Language Writing Research." *Language Learning* 47.1 (1997): 101-143.

Purcell, Natalie J. *Death Metal: The Passion and Politics of a Subculture*. Jefferson, NC: McFarland & Company, Inc., Publishers, 2003.

Quero, Betsy and Averil Coxhead. "Using a Corpus-Based Approach to Select Medical Vocabulary for an ESP Course: The Case for High-Frequency Vocabulary." *Key Issues in English for Specific Purposes in Higher Education*. Ed. Yasemin Kırkgöz and Kenan Dikilitaş. Springer, 2018. 51-75.

Quirk, Randolph, et al. *A Comprehensive Grammar of the English Language*. NY: Longman, 1985.

Rayson, Paul and Roger Garside. "Comparing corpora using frequency profiling." *Proceedings of the Workshop on Comparing Corpora*. Association for Computational Linguistics, 2000. 1-6.

Rayson, Paul. "Corpus Analysis of Key Words." *The Encyclopedia of Applied Linguistics*. Ed. Carol A. Chapelle. Wiley, 2012.

—. "From key words to key semantic domains." *International Journal of Corpus Linguistics* 13.4 (2008): 519 - 549.

—. "Matrix: A statistical method and software tool for linguistic analysis through corpus comparison." Lancaster University, 2003. PhD Dissertation.

Read, John and Paul Nation. "An investigation of the lexical dimension of the IELTS Speaking Test." *International English Language Testing System (IELTS) Research Reports* 6 (2006): 1-25.

Read, John. *Assessing Vocabulary*. Cambridge University Press, 2000.

Reda, Ghsoon. "English coursebooks: prototype texts and basic vocabulary norms." *ELT Journal* 57.3 (2003): 260-268.

Reid, Joy. "Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective." *Second language writing: Research insights for the classroom*. Ed. Barbara Kroll. NY: Cambridge University Press, 1990. 191-210.

Research Uses of Vocabprofile. n.d. 03 November 2014. <http://www.lextutor.ca/vp/research.html>.

Révész, Andrea, Nektaria-Efstathia Kourtali and Diana Mazgutova. "Effects of Task Complexity on L2 Writing Behaviors and Linguistic Complexity." *Language Learning* 67.1 (2019): 208-241.

Richards, Brian and David Malvern. "Validity and threats to the validity of vocabulary measurement." *Modelling and Assessing Vocabulary Knowledge*. Ed. Helmut Daller, James Milton and Jeanine Treffers-Daller. Cambridge University Press, 2007.

Riches, Gabrielle. "Embracing the Chaos: Mosh Pits, Extreme Metal Music and Liminality." *Journal for Cultural Research* 15.3 (2011): 315-332.

Ringbom, Håkan. "High-frequency verbs in the ICLE corpus." *Explorations in Corpus Linguistics*. Ed. Antoinette Renouf. Amsterdam: Rodopi, 1998.

Ringeling, Tobi. "Subjective Estimations as a Useful Alternative to Word Frequency Counts." *Interlanguage Studies Bulletin* 8.1 (1984): 59-69.

Rissanen, Matti. "The diachronic corpus as a window to the history of English." *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Ed. Jan Svartvik. De Gruyter, 1992.

Robinson, Peter. "Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty." *International Review of Applied Linguistics in Language Teaching* 45.3 (2007): 193-213.

Roman, Luke and Monica Roman. *Encyclopedia of Greek and Roman Mythology*. New York: Facts On File, Inc., 2010.

Rules & Guidelines - Encyclopaedia Metallum: The Metal Archives. n.d. 28 March 2017. <https://www.metal-archives.com/content/rules>.

Rutherford, Brian A. "Genre Analysis of Corporate Annual Report Narratives: A Corpus Linguistics–Based Approach." *Journal of Business Communication* 42.4 (2005): 349-378.

Sano, Motoki and Takehiko Maruyama. *Proceedings of ISFC 35: Voices Around the World*. Ed. Canzhong Wu, Christian M.I.M. Matthiessen and Maria Herke. Sydney, 2008. 359-364.

Schmitt, Norbert and Bruce Dunham. "Exploring native and non-native intuitions of word frequency." *Second Language Research* 15.4 (1999): 389-411.

Schmitt, Norbert and Diane Schmitt. "A reassessment of frequency and vocabulary size in L2 vocabulary teaching." *Language Teaching* 47.4 (2014): 484-503.

Scholz, Ronny. "Lexicometry: A Quantifying Heuristic for Social Scientists in Discourse Studies." *Quantifying Approaches to Discourse for Social*

Scientists. Postdisciplinary Studies in Discourse. Ed. Ronny Scholz. Palgrave Macmillan, 2019. 123-153.

Scott, M. "WordSmith Tools version 7." Stroud, 2017. Lexical Analysis Software.

Scott, Mike and Christopher Tribble. *Textual patterns: Key words and corpus analysis in language education*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 2006.

Scott, Mike. "In Search of a Bad Reference Corpus." *What's in a Word-list?: Investigating Word Frequency and Keyword Extraction*. Ed. Dawn Archer. Ashgate Publishing, Ltd, 2009.

Scott, Mike. "Problems in investigating keyness, or clearing the undergrowth and marking out trails…." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. John Benjamins Publishing Company, 2010.

Seidlhofer, Barbara. "Corpora and English as a Lingua Franca." *Corpus Applications in Applied Linguistics*. Ed. Ken Hyland, Chau Meng Huat and Michael Handford. Continuum, 2012.

Shuker, Roy. *Understanding Popular Music*. 2nd Edition. London: Routledge, 2001.

Simonton, Dean Keith. "Lexical choices and aesthetic success: A computer content analysis of 154 Shakespeare sonnets." *Computers and the Humanities* 24.4 (1990): 251-264.

Sinclair, John. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

Sinclair, John. "The Automatic Analysis of Corpora." *Directions in corpus linguistics: proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Ed. Jan Svartvik. De Gruyter, 1992.

Singh, Sameer. "A Pilot Study on Gender Differences in Conversational Speech on Lexical Richness Measures." *Literary and Linguistic Computing* 16.3 (2001): 251-264.

—. "Linguistic computing in speech and language disorders." *Proceedings of the 5th International Conference on Speech Science and Technology*. 1994.

Šišková, Zdislava. "Lexical Richness in EFL Students' Narratives." *Language Studies Working Papers* 4 (2012): 26-36.

Skehan, Peter and Pauline Foster. "Complexity, accuracy, fluency and lexis in task-based performance: A synthesis of the Ealing research." *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Ed. Alex Housen, Folkert Kuiken and Ineke Vedder. John Benjamins Publishing Company, 2012.

Skehan, Peter. "Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis." *Applied Linguistics* 30.4 (2009): 510–532.

Sonbul, Suhad. "Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing." *Bilingualism: Language and Cognition* 18.3 (2014): 419-437.

Sophiadi, Angela. "The Song Remains the Same…or Not? A Pragmatic Approach to the Lyrics of Rock Music." *Major Trends in Theoretical and Applied Linguistics 2: Selected Papers from the 20th ISTAL*. Ed. Nikolaos Lavidas, Thomaï Alexiou and Areti Maria Sougari. London: Versita, 2014.

Spracklen, Karl, Andy R. Brown and Keith Kahn-Harris. "Metal Studies? Cultural Research in the Heavy Metal Scene." *Journal for Cultural Research* 15.3 (2011): 209-2012.

Stamatatos, E., Nikos D. Fakotakis and George K. Kokkinakis. "Text genre detection using common word frequencies." *Proceedings of the 18th Conference on Computational Linguistics*. Association for Computational Linguistics, 2000. 808-814.

Stats - Encyclopaedia Metallum: The Metal Archives. n.d. 28 March 2017. <https://www.metal-archives.com/stats>.

Stelzner, Chuck. "Death Metal/Throat Vocal Analysis." Unpublished paper. 06 October 2016. <https://courses.physics.illinois.edu/phys406/sp2017/Student_Projects/Spring05/Chuck_Stelzner/Chuck_Stelzner_P498POM_Final_Report.pdf>.

Stirman, Shannon Wiltsey and James W. Pennebaker. "Word Use in the Poetry of Suicidal and Nonsuicidal Poets." *Psychosomatic Medicine* 63.4 (2001): 517-522.

Strother, Eric Scott. "Unlocking the Paradox of Christian Metal Music." Doctoral Dissertation. University of Kentucky, 2013.

Stubbs, Michael. "Three concepts of keywords." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. John Benjamins Publishing Company, 2010.

Subgenre. 2011. Houghton Mifflin Harcourt Publishing Company. 17 October 2016. <https://www.thefreedictionary.com/subgenre>.

Sundberg, Ross. "Bande à Part: The Development of a Music Application for L2 French Learners." Master's Thesis. Concordia University, 2015.

Sutarsyah, Cucu, Paul Nation and Graeme Kennedy. "How Useful Is Eap Vocabulary for Esp? A Corpus Based Case Study." *RELC Journal* 25.2 (1994): 34-50.

Swales, John M. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press, 1990.

Tagg, Caroline. "Texting and Corpora." *Corpus Applications in Applied Linguistics*. Ed. Ken Hyland, Chau Meng Huat and Michael Handford. Continuum, 2012.

Taina, Jesse. "Keywords in heavy metal lyrics: A data-driven corpus study into the lyrics of five heavy metal subgenres." *Master's Thesis* (2014).

Tegge, Friederike. "The lexical coverage of popular songs in English language teaching." *System* 67 (2017): 87-98.

Teubert, Wofgang. "Language and Corpus Linguistics." *Lexicology and Corpus Linguistics: An Introduction*. Ed. M.A.K. Halliday, et al. London: Continuum, 2004. 73-112.

Teubert, Wolfgang and Anna Čermáková. "Directions in Corpus Linguistics." *Lexicology and Corpus Linguistics: An Introduction*. Ed. M.A.K. Halliday, et al. Continuum, 2004.

The British National Corpus. Vers. 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. 12 July 2017. <http://www.natcorp.ox.ac.uk/>.

The Wikipedia Corpus. n.d. 13 04 2019. <https://www.english-corpora.org/wiki/>.

Tidball, Françoise and Jeanine Treffers-Daller. "Exploring measures of vocabulary richness in semi-spontaneous French speech." *Modelling and Assessing Vocabulary Knowledge*. Ed. Helmut Daller, James Milton and Jeanine Treffers-Daller. Cambridge University Press, 2007.

To, Vinh, Si Fan and Damon Thomas. "Lexical density and readability: A case study of English textbooks." *Internet Journal of Language, Culture and Society* 37 (2013): 61-71.

Tognini-Bonelli, Elena. *Corpus Linguistics at Work*. John Benjamins Publishing Company, 2001.

Tonkyn, Alan. "Measuring and perceiving changes in oral complexity, accuracy and fluency: Examining instructed learners' short-term gains." *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Ed. Alex Housen, Folkert Kuiken and Ineke Vedder. John Benjamins Publishing Company, 2012.

Toolan, Michael. *Narrative Progression in the Short Story: A corpus stylistic approach*. John Benjamins Publishing Company, 2009.

Torruella, Joan and Ramon Capsada. "Lexical Statistics and Tipological Structures: A Measure of Lexical Richness." *Procedia - Social and Behavioral Sciences* 95 (2013): 447-454.

Tristl, Christiane, Martin Müller and Veit Bachmann. "Lexicometric Analysis: A Methodological Prelude." *Perceptions of the EU in Eastern Europe and Sub-Saharan Africa*. Ed. Veit Bachmann and Martin Müller. Palgrave Macmillan, 2015.

Tsatsishvili, Valeri. "Automatic subgenre classification of heavy metal music." Master's Thesis. University of Jyväskyla, Finland, 2011.

Tweedie, Fiona J. and R. Harald Baayen. "How Variable May a Constant be? Measures of Lexical Richness in Perspective." *Computers and the Humanities* 32 (1998): 323-352.

Ure, Jean. "Lexical density and register differentiation." *Applications of Linguistics*. Ed. G. E. Perren and J. L. M. Trim. Cambridge University Press, 1971. 443-452.

Van Elferen, Isabella. "'And machine created music': Cybergothic music and the phantom voices of the technological uncanny." *Digital Material: Tracing New Media in Everyday Life and Technology*. Ed. Marianne Van den Boomen, et al. Amsterdam: Amsterdam University Press, 2009. 121-132.

van Hout, Roeland and Anne Vermeer. "Comparing measures of lexical richness." *Modelling and Assessing Vocabulary Knowledge*. Ed. Helmut Daller, James Milton and Jeanine Treffers-Daller. Cambridge: Cambridge University Press, 2007.

van Zeeland, Hilde and Norbert Schmitt. "Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension?" *Applied Linguistics* 34.4 (2013): 457-479.

Vermeer, Anne. "Coming to grips with lexical richness in spontaneous speech data." *Language Testing* 17.1 (2000): 65-83.

Verspoor, Marjolijn, Monika S. Schmid and Xiaoyan Xu. "A dynamic usage based perspective on L2 writing." *Journal of Second Language Writing* 21.3 (2012): 239-263.

Vidaković, Ivana and Fiona Barker. "Lexical development across second language proficiency levels: a corpus-informed study." *Language, Learning, and Context: Proceedings of the 42nd Annual Meeting of the British Association for Applied Linguistics*. Ed. Andrew Harris and Adam Brandt. London: Scitsiugnil Press, 2009. 143-146.

Vongpumivitch, Viphavee, Ju-yu Huang and Yu-Chia Chang. "Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers." *English for Specific Purposes* 28.1 (2009): 33-41.

Vögelin, Cristina, et al. "The influence of lexical features on teacher judgements of ESL argumentative essays." *Assessing Writing* 39 (2019): 50-63.

Walser, Robert. *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*. Hanover, NH: Wesleyan University Press, 1993.

Ward, Steven. *Geoff Barton, behind the wheel: Former Kerrang! editor discovers a different kind of speed*. n.d. 08 October 2014. <http://rockcriticsarchives.com/interviews/geoffbarton/geoffbarton.html>.

Warren, Martin. "Corpora: Specialized." Ed. Carol A. Chapelle. Wiley, 2012.

Warren, Martin. "Identifying aboutgrams in engineering texts." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. John Benjamins Publishing Company, 2010.

Webb, Stuart and Michael P. H. Rodgers. "The Lexical Coverage of Movies." *Applied Linguistics* 30.3 (2009a): 407-427.

—. "Vocabulary Demands of Television Programs." *Language Learning* 59.2 (2009b): 335-366.

Weinstein, Deena. *Heavy Metal: A Cultural Sociology*. New York: Lexington, 1991.

—. *Heavy Metal: The Music and Its Culture*. Revised Edition. London: Da Capo Press, 2000.

Weinstein, Deena. "The Empowering Masculinity of British Heavy Metal." *Heavy Metal Music in Britain*. Ed. Gerd Bayer. Surrey, UK: Ashgate, 2009. 17-31.

Werner, Valentin. "Love is all around: a corpus-based study of pop lyrics." *Corpora* 7.1 (2012): 19-50.

West, Michael. *A General Service List of English Words with Semantic Frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology*. London: Longman, 1953.

Wimmer, Gejza and Gabriel Altmann. "Review Article: On Vocabulary Richness." *Journal of Quantitative Linguistics* 6.1 (1999): 1-9.

Xiao, Zhonghua and Anthony McEnery. "Two Approaches to Genre Analysis: Three Genres in Modern American English." *Journal of English Linguistics* 33.1 (2005): 62-82.

Yang , Wenhsien. "Keyness in academic textbook blurbs: Lexical variations across disciplines." *2012 Language and Language Teaching Conference Proceedings*. 2012. 63-73.

Yang, Ming-Nuan. "A nursing academic word list." *English for Specific Purposes* 37 (2015): 27-38.

Yang, Weiwei. "Mapping the Relationships among the Cognitive Complexity of Independent Writing Tasks, L2 Writing Quality, and Complexity, Accuracy and Fluency of L2 Writing." Georgia State University, 2014. Dissertation.

Yates, Joanne and Wanda J. Orlikowski. "Genres of Organizational Communication: A Structurational Approach to Studying Communication and Media." *The Academy of Management Review* 17.2 (1992): 299-326.

Yoon, Hyung-Jo and Charlene Polio. "The Linguistic Development of Students of English as a Second Language in Two Written Genres." *TESOL Quarterly* 51.2 (2017): 275-301.

Yoon, Hyung-Jo. "Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality." *System* 66 (2017): 130-141.

Yu, Guoxing. "Lexical Diversity in Writing and Speaking Task Performances." *Applied Linguistics* 31.2 (2009): 236-259.

Zheng, Yongyan. "The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English." *System* 56 (2016): 40-53.

Zora, Subhi and Catherine Johns-Lewis. "Lexical Density in Interview and Conversation." *York Papers in Linguistics* 14 (1989): 89-100.

ALBUMS

Angel Witch. *Angel Witch*, Bronze Records, 1980.

Angel Witch. *Screamin' n' Bleedin'*, Killerwatt Records, 1985.

Annihilated. *Scorched Earth Policy*, Killer Metal Records, 2010.

Anthrax. *State of Euphoria*, Island Records, 1989.

Anthrax. *Spreading the Disease*, Island Records, 1985.

Armored Saint. *March of the Saint*, Chrysalis, 1984.

Autopsy. *Macabre Eternal*, Peaceville Records, 2011.

Benediction. *Grind Bastard*, Nuclear Blast, 1998.

Black Sabbath. *Paranoid*, Vertigo, 1970.

Black Sabbath. *Master of Reality*, Vertigo, 1971.

Blaze Bayley. *The Man Who Would Not Die*, Blaze Bayley Recordings, 2008.

Bolt Thrower. *...For Victory*, Earache Records, 1994.

Bolt Thrower. *The IVth Crusade*, Earache Records, 1992.

Cannibal Corpse. *Eaten Back to Life*, Metal Blade Records, 1990.

Carcass. *Heartwork*, Earache Records, 1993.

Carcass. *Necroticism - Descanting the Insalubrious*, Earache Records, 1991.

Dāmim. *The Difference Engine*, Candlelight Records, 2007.

Dāmim. *Purity: The Darwinian Paradox*, Candlelight Records, 2005.

Dead Beyond Buried. *Condemned to Misery*, Siege of Amida Records, 2007.

Death. *The Sound of Perseverance*, Nuclear Blast, 1998.

Death. *Symbolic*, Roadrunner Records, 1995.

Death Angel. *The Dream Calls for Blood*, Nuclear Blast, 2013.

Death Angel. *Frolic Through the Park*, Enigma, 1988.

Dee Snider. *For the Love of Metal*, Napalm Records, 2018.

Deicide. *Scars of the Crucifix*, Earache Records, 2004.

Deicide. *Deicide*, R/C Records, 1990.

Desecration. *Cemetery Sickness*, Metal Age Productions, 2014.

Desecration. *Pathway to Deviance*, Copro Records, 2002.

Dokken. *Back for the Attack*, Elektra Records, 1987.

Dyscarnate. *And So It Came to Pass*, Siege of Amida Records, 2012.

Dyscarnate. *Enduring the Massacre*, Siege of Amida Records, 2010.

Evile. *Infected Nations*, Earache Records, 2009.

Evile. *Enter the Grave*, Earache Records, 2007.

Exodus. *Blood In Blood Out*, Nuclear Blast, 2014.

Exodus. *Bonded by Blood*, Music for Nations, 1985.

Fist. *Storm*, Demolition Records, 2005.

Fist. *Turn the Hell On*, MCA Records, 1980.

Flotsam and Jetsam. *The End of Chaos*, AFM Records, 2019.

Fozzy. *Chasing the Grail*, Riot! Entertainment, 2010.

Gama Bomb. *Tales from the Grave in Space*, Earache Records, 2009.

Girlschool. *Girlschool*, Progressive International, 1992.

Girlschool. *Take a Bite*, GWR Records, 1988.

Girlschool. *Play Dirty*, Bronze Records, 1983.

Girlschool. *Demolition*, Bronze Records, 1980.

Gorerotted. *A New Dawn for the Dead*, Metal Blade Records, 2005.

Gorerotted. *Only Tools and Corpses*, Metal Blade Records, 2003.

Impellitteri. *Venom*, Frontiers Records, 2015.

Iron Maiden. *No Prayer for the Dying*, EMI, 1990.

Judas Priest. *Painkiller*, Columbia Records, 1990.

Judas Priest. *British Steel*, CBS, 1980.

Lawnmower Deth. *Return of the Fabulous Metal Bozo Clowns*, Earache Records, 1992.

Lawnmower Deth. *Ooh Crikey It's... Lawnmower Deth*, Earache Records, 1990.

Malevolent Creation. *The Will to Kill*, Nuclear Blast, 2002.

Malevolent Creation. *Retribution*, R/C Records, 1992.

Massacre. *Back from Beyond*, Century Media Records, 2014.

Megadeth. *Rust in Peace*, Capitol Records, 1990.

Megadeth. *Peace Sells... But Who's Buying?*, Capitol Records, 1986.

Metallica. *Ride the Lightning,* Megaforce Records, 1984.

Metallica. *Kill 'Em All,* Megaforce Records, 1983.

Monstrosity. *Rise to Power*, Conquest Music, Inc., 2003.

Monstrosity. *Millennium*, Conquest Music, Inc., 1996.

Morbid Angel. *Domination*, Earache Records, 1995.

Morbid Angel. *Covenant*, Earache Records, 1993.

Napalm Death. *Words from the Exit Wound*, Earache Records, 1998.

Napalm Death. *Fear, Emptiness, Despair*, Earache Records, 1994.

Necrosanct. *Incarnate*, Black Mark Production, 1992.

Necrosanct. *Equal in Death*, Metalworks, 1990.

Nuclear Assault. *Survive,* I.R.S. Records, 1988.

Nuclear Assault. *Game Over,* Combat Records, 1986.

Obituary. *World Demise*, Roadrunner Records, 1994.

Obituary. *The End Complete*, R/C Records, 1992.

Onslaught. *Sounds of Violence*, AFM Records, 2011.

Onslaught. *Killing Peace*, Candlelight Records, 2007.

Overkill. *The Years of Decay,* Megaforce Records, 1989.

Overkill. *Feel the Fire,* Megaforce Records, 1985.

Possessed. *Revelations of Oblivion*, Nuclear Blast, 2019.

Possessed. *Seven Churches*, Combat Records, 1985.

Psython. *Hatred*, Independent, 2017.

Psython. *...Outputs*, Independent, 2016.

Quiet Riot. *QR III*, Pasha, 1986.

Quiet Riot. *Metal Heath*, Pasha, 1983.

Racer X. *Technical Difficulties*, Mercury Records, 1999.

Re-Animator. *That Was Then... This Is Now*, Under One Flag, 1992.

Re-Animator. *Condemned to Eternity*, Under One Flag, 1990.

Sabbat. *Mourning Has Broken*, Noise Records, 1991.

Sabbat. *Dreamweaver*, Noise Records, 1989.

Sabbat. *History of a Time to Come*, Noise Records, 1988.

Saxon. *Denim and Leather*, Carrere, 1981.

Saxon. *Wheels of Steel*, Carrere, 1980.

Sebastian Bach. *Kicking & Screaming*, Frontiers Records, 2011.

Slayer. *Seasons in the Abyss*, Def American Recordings, 1990.

Slayer. *South of Heaven*, Def Jam Recordings, 1988.

Tank. *Filth Hounds of Hades*, Kamaflage Records, 1982.

Tank. *Power of the Hunter*, Kamaflage Records, 1982.

Testament. *Low*, Atlantic Records, 1994.

Toranaga. *Righteous Retribution*, Chuffing Hell Records, 2013.

Toranaga. *God's Gift*, Chrysalis, 1990.

Tygers of Pan Tang. *Spellbound*, MCA Records, 1981.

Tygers of Pan Tang. *Wild Cat*, MCA Records, 1980.

Virgin Steele. *Guardians of the Flame*, Mongol Horde, 1983.

Virgin Steele. *Virgin Steele*, Maze Music, 1982.

W.A.S.P.. *Babylon*, Demolition Records, 2009.

W.A.S.P.. *The Headless Children*, Capitol Records, 1989.

Xentrix. *For Whose Advantage*, Roadracer Records, 1990.

Xentrix. *Shattered Existence*, Roadracer Records, 1989.

DOCUMENTARIES

Get Thrashed: The Story of Thrash Metal. 2006. Dir. Rick Ernst. Saigon1515 Productions, LLC.

Heavy: The Story of Metal, Episode 2: "British Steel". 2006. Dir. Michael John Warren. VH1, 441 Productions.

Heavy: The Story of Metal, Episode 4: "Seek and Destroy". 2006. Dir. Michael John Warren. VH1, 441 Productions.

Heavy Metal: Louder Than Life. 2006. Dir. Dick Carruthers. Metropolis.

Heavy Metal Britannia. 2010. Dir. Chris Rodley. British Broadcasting Corporation (BBC).

Metal: A Headbanger's Journey. 2005. Dir. Sam Dunn, Scot McFayden and Jessica Joy Wise. Banger Films.

Metal Evolution. 2011. Dir Sam Dunn and Scot McFayden. Banger Films.

Metal Evolution, Episode 3: "Early Metal Part 2: UK Division". 2011. Dir Sam Dunn and Scot McFayden. Banger Films.

Metal Evolution, Episode 6: "Thrash". 2011. Dir Sam Dunn and Scot McFayden. Banger Films.

Metal Evolution, Episode 8: Nu Metal. 2012. Dir Sam Dunn and Scot McFayden. Banger Films.

Metal Evolution: "Extreme Metal: The Lost Episode". 2014. Dir Sam Dunn and Scot McFayden. Banger Films.

Seven Ages of Rock, Episode 4: "Never Say Die". 2007. Dir. Anna Gravelle. BBC Worldwide and VH1 Classic.

Seven Ages of Rock, Episode 5: "We Are the Champions". 2007. Dir. Anna Gravelle. BBC Worldwide and VH1 Classic.

Seven Ages of Rock: Heavy Metal 1970-1991, DVD, directed by Anna Gravelle (BBC 2/Vh1 Classic co-production, 2007) – EPISODE 5 (We Are the Champions)

8. APPENDICES

A. Metal: A Headbanger's Journey

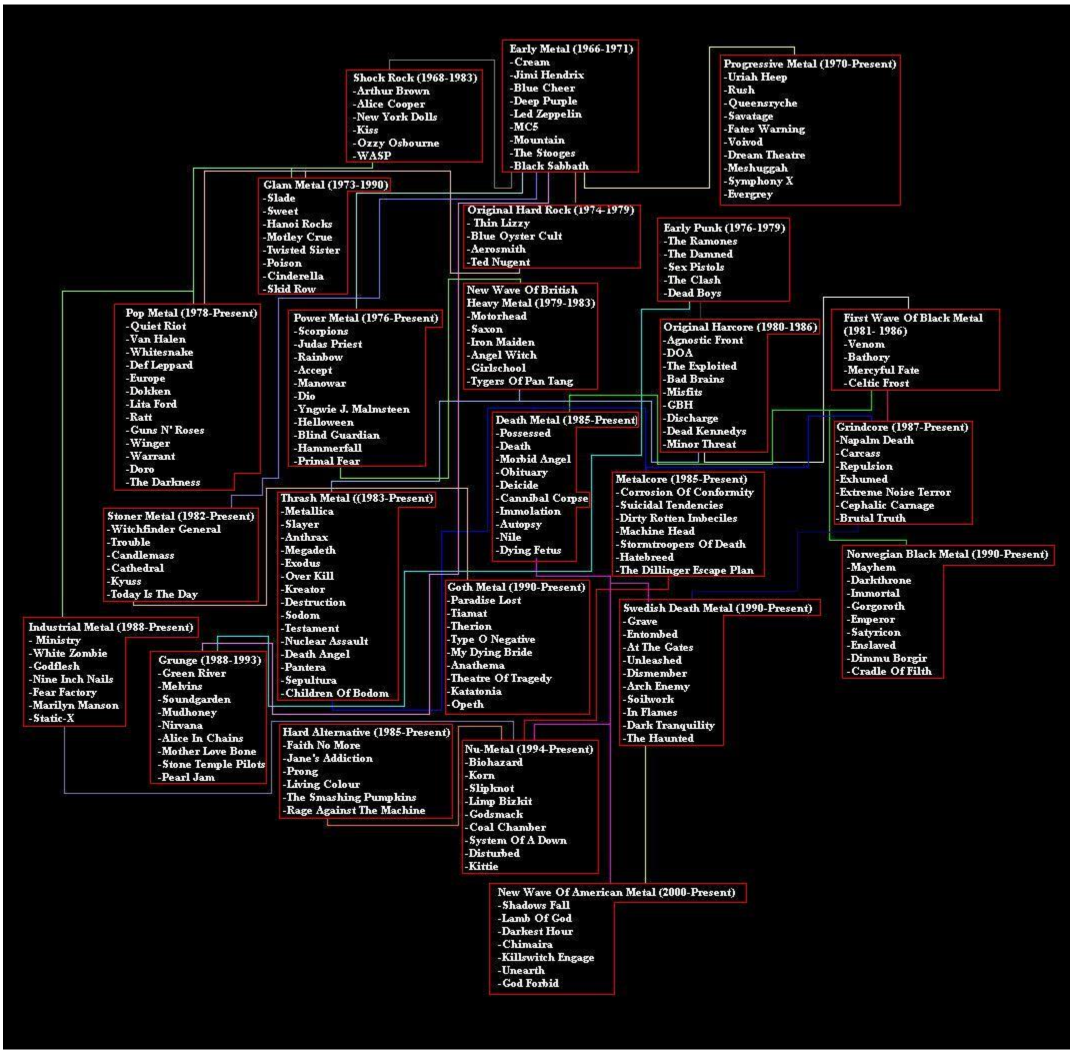

Figure 4: Metal Genealogy created by Metalprince based on *Metal: A Headbanger's Journey* published on the Wikipedia Page of the movie.

B. Metal Evolution

Figure 5: Metal Genealogy Screenshot from Metal Evolution (1/3).

Figure 6: Metal Genealogy Screenshot from Metal Evolution (2/3).

Figure 7: Metal Genealogy Screenshot from Metal Evolution (3/3).

C. Metal Evolution Charts from Banger Films Facebook Page

Figure 8: Updated Metalcore and Crossover chart published on Banger Facebook Page.

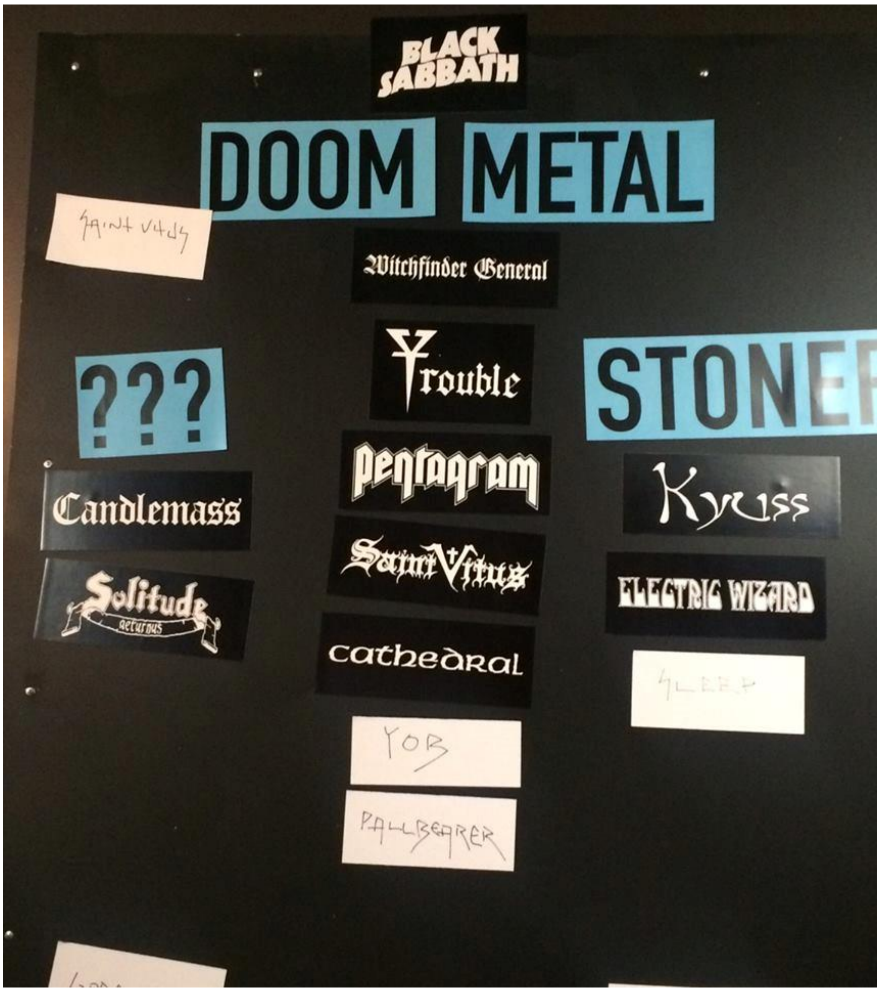

Figure 9: Updated doom metal and stoner chart published on Banger Facebook Page.

Figure 10: Updated industrial metal chart published on Banger Facebook Page.

Figure 11: Updated Mathcore chart published on Banger Facebook Page.

Figure 12: Updated folk metal chart published on Banger Facebook Page.

Figure 13: Updated prog metal chart published on Banger Facebook Page.

D. HEAVY METAL SUBCORPUS (HMSC)

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
1	Angel Witch	heavy_uk_01	UK	4	Angel Witch (1980)	heavy_uk_01-01	heavy_uk_01-01-01	1. Angel Witch	141	1,561
							heavy_uk_01-01-02	2. Atlantis	156	
							heavy_uk_01-01-03	3. White Witch	147	
							heavy_uk_01-01-04	4. Confused	104	
							heavy_uk_01-01-05	6. Gorgon	199	
							heavy_uk_01-01-06	7. Sweet Danger	121	
							heavy_uk_01-01-07	8. Free Man	145	
							heavy_uk_01-01-08	9. Angel Of Death	167	
					Screamin' n' Bleedin' (1985)	heavy_uk_01-02	heavy_uk_01-02-01	1. Who's to Blame	219	
							heavy_uk_01-02-02	2. Child of the Night	162	
2	Black Sabbath	heavy_uk_02	UK	19	Paranoid (1970)	heavy_uk_02-01	heavy_uk_02-01-01	1. War Pigs	144	1,620
							heavy_uk_02-01-02	2. Paranoid	141	
							heavy_uk_02-01-03	4. Iron Man	139	
							heavy_uk_02-01-04	5. Electric Funeral	140	
							heavy_uk_02-01-05	6. Hand of Doom	249	
							heavy_uk_02-01-06	8. Fairies Wear Boots	110	
					Master of Reality (1971)	heavy_uk_02-02	heavy_uk_02-02-01	2. After Forever	306	
							heavy_uk_02-02-02	4. Children of the Grave	144	
							heavy_uk_02-02-03	6. Lord of This World	142	
							heavy_uk_02-02-04	7. Solitude	105	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
3	Blaze Bayley	heavy_uk_03	UK	6	The Man Who Would Not Die (2008)	heavy_uk_03-01	heavy_uk_03-01-01	1. The Man Who Would Not Die	164	1,482
							heavy_uk_03-01-02	2. Blackmailer	118	
							heavy_uk_03-01-03	3. Smile Back at Death	262	
							heavy_uk_03-01-04	5. Samurai	119	
							heavy_uk_03-01-05	6. Crack in the System	113	
							heavy_uk_03-01-06	8. At the End of the Day	144	
							heavy_uk_03-01-07	9. Waiting for My Life to Begin	152	
							heavy_uk_03-01-08	10. Voices from the Past	127	
							heavy_uk_03-01-09	11. The Truth Is One	152	
							heavy_uk_03-01-10	12. Serpent Hearted Man	131	
4	Fist	heavy_uk_04	UK	3	Storm (2005)	heavy_uk_04-01	heavy_uk_04-01-01	1. Fe Fi Fo Fum	124	1,628
							heavy_uk_04-01-02	2. Guardian Angel	121	
							heavy_uk_04-01-03	3. Acid Rock	101	
							heavy_uk_04-01-04	4. Storm	185	
							heavy_uk_04-01-05	5. Name Rank & Serial Number	163	
							heavy_uk_04-01-06	6. Try a Little Love on Me	165	
							heavy_uk_04-01-07	7. Brain Damage	134	
							heavy_uk_04-01-08	8. Supercallousflagellistic-expertcunnilingus	290	
							heavy_uk_04-01-09	10. Never Get Me Up (In One of Those)	130	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
					Turn the Hell On (1980)	heavy_uk_04-02	heavy_uk_04-02-01	1. Hole in the Wall Gang	215	
5	Girlschool	heavy_uk_05	UK	13	Girlschool (1992)	heavy_uk_05-01	heavy_uk_05-01-01	1. My Ambition	112	1,257
							heavy_uk_05-01-02	2. One More	110	
							heavy_uk_05-01-03	4. Wild at Heart	163	
							heavy_uk_05-01-04	6. We Came	119	
							heavy_uk_05-01-05	8. Sitting Pretty	111	
							heavy_uk_05-01-06	9. On My Way	139	
							heavy_uk_05-01-07	10. Take Me I'm Yours	106	
					Demolition (1980)	heavy_uk_05-02	heavy_uk_05-02-01	1. Demolition Boys	113	
					Play Dirty (1983)	heavy_uk_05-03	heavy_uk_05-03-01	2. High and Dry	154	
					Take a Bite (1988)	heavy_uk_05-04	heavy_uk_05-04-01	1. Action	130	
6	Iron Maiden	heavy_uk_06	UK	16	No Prayer for the Dying (1990)	heavy_uk_06-01	heavy_uk_06-01-01	1. Tailgunner	126	1,596
							heavy_uk_06-01-02	2. Holy Smoke	232	
							heavy_uk_06-01-03	3. No Prayer for the Dying	125	
							heavy_uk_06-01-04	4. Public Enema Number One	148	
							heavy_uk_06-01-05	5. Fates Warning	146	
							heavy_uk_06-01-06	6. The Assassin	159	
							heavy_uk_06-01-07	7. Run Silent Run Deep	209	
							heavy_uk_06-01-08	8. Hooks in You	172	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							heavy_uk_06-01-09	9. Bring Your Daughter... to the Slaughter	172	
							heavy_uk_06-01-10	10. Mother Russia	107	
7	Judas Priest	heavy_uk_07	UK	19	Painkiller (1990)	heavy_uk_07-01	heavy_uk_07-01-01	1. Painkiller	137	1,507
							heavy_uk_07-01-02	4. Leather Rebel	123	
							heavy_uk_07-01-03	6. Night Crawler	185	
							heavy_uk_07-01-04	7. Between the Hammer & the Anvil	119	
							heavy_uk_07-01-05	8. A Touch of Evil	122	
							heavy_uk_07-01-06	10. One Shot at Glory	136	
					British Steel (1980)	heavy_uk_07-02	heavy_uk_07-02-01	1. Rapid Fire	149	
							heavy_uk_07-02-02	4. Grinder	153	
							heavy_uk_07-02-03	7. Living After Midnight	134	
							heavy_uk_07-02-04	6. You Don't Have to Be Old to Be Wise	249	
8	Saxon	heavy_uk_08	UK	23	Wheels of Steel (1980)	heavy_uk_08-01	heavy_uk_08-01-01	1. Motorcycle Man	132	1,654
							heavy_uk_08-01-02	2. Stand Up and Be Counted	140	
							heavy_uk_08-01-03	3. 747 (Strangers in the Night)	187	
							heavy_uk_08-01-04	4. Wheels of Steel	306	
							heavy_uk_08-01-05	5. Freeway Mad	112	
							heavy_uk_08-01-06	6. See the Light Shining	163	
							heavy_uk_08-01-07	7. Street Fighting Gang	203	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							heavy_uk_08-01-08	8. Suzie Hold On	147	
							heavy_uk_08-01-09	9. Machine Gun	115	
					Denim and Leather (1981)	heavy_uk_08-02	heavy_uk_08-02-01	1. Princess of the Night	149	
9	Tank	heavy_uk_09	UK	6	Filth Hounds of Hades (1982)	heavy_uk_09-01	heavy_uk_09-01-01	1. Shellshock	106	1,402
							heavy_uk_09-01-02	2. Struck by Lightning	133	
							heavy_uk_09-01-03	3. Run Like Hell	198	
							heavy_uk_09-01-04	4. Blood, Guts and Beer	162	
							heavy_uk_09-01-05	5. That's What Dreams are Made Of	122	
							heavy_uk_09-01-06	6. Turn Your Head Around	116	
							heavy_uk_09-01-07	7. Heavy Artillery	125	
							heavy_uk_09-01-08	9. Filth Hounds of Hades	133	
							heavy_uk_09-01-09	10. (He Fell in Love with a) Stormtrooper	147	
					Power of the Hunter (1982)	heavy_uk_09-02	heavy_uk_09-02-01	1. Walking Barefoot over Glass	160	
10	Tygers of Pan Tang	heavy_uk_10	UK	11	Wild Cat (1980)	heavy_uk_10-01	heavy_uk_10-01-01	1. Euthanasia	114	1,267
							heavy_uk_10-01-02	2. Slave to Freedom	130	
							heavy_uk_10-01-03	3. Don't Touch Me There	104	
							heavy_uk_10-01-04	4. Money	142	
							heavy_uk_10-01-05	5. Killers	139	
							heavy_uk_10-01-06	6. Fireclown	105	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							heavy_uk_10-01-07	8. Suzie Smiled	130	
							heavy_uk_10-01-08	9. Badger Badger	107	
					Spellbound (1981)	heavy_uk_10-02	heavy_uk_10-02-01	2. Take It	167	
							heavy_uk_10-02-02	10. Don't Stop By	129	
11	Armored Saint	heavy_us_01	USA	7	March of the Saint (1984)	heavy_us_01-01	heavy_us_01-01-01	1. March of the Saint	116	1,487
							heavy_us_01-01-02	2. Can U Deliver	154	
							heavy_us_01-01-03	3. Mad House	168	
							heavy_us_01-01-04	4. Take a Turn	148	
							heavy_us_01-01-05	5. Seducer	148	
							heavy_us_01-01-06	6. Mutiny on the World	117	
							heavy_us_01-01-07	7. Glory Hunter	141	
							heavy_us_01-01-08	8. Stricken by Fate	203	
							heavy_us_01-01-09	9. Envy	140	
							heavy_us_01-01-10	10. False Alarm	152	
12	Dee Snider	heavy_us_02	USA	4	For the Love of Metal (2018)	heavy_us_02-01	heavy_us_02-01-01	1. Lies Are a Business	141	1,739
							heavy_us_02-01-02	2. Tomorrow's No Concern	107	
							heavy_us_02-01-03	3. I Am the Hurricane	173	
							heavy_us_02-01-04	4. American Made	110	
							heavy_us_02-01-05	5. Roll over You	246	
							heavy_us_02-01-06	6. I'm Ready	162	
							heavy_us_02-01-07	7. Running Mazes	185	
							heavy_us_02-01-08	8. Mask	195	
							heavy_us_02-01-09	9. Become the Storm	206	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							heavy_us_02-01-10	10. The Hardest Way	214	
13	Dokken	heavy_us_03	USA	12	Back for the Attack (1987)	heavy_us_03-01	heavy_us_03-01-01	1. Kiss of Death	173	2,100
							heavy_us_03-01-02	2. Prisoner	172	
							heavy_us_03-01-03	3. Night by Night	197	
							heavy_us_03-01-04	4. Standing in the Shadows	215	
							heavy_us_03-01-05	5. Heaven Sent	260	
							heavy_us_03-01-06	7. So Many Tears	238	
							heavy_us_03-01-07	8. Burning like a Flame	232	
							heavy_us_03-01-08	9. Lost Behind the Wall	129	
							heavy_us_03-01-09	10. Stop Fighting Love	224	
							heavy_us_03-01-10	11. Cry of the Gypsy	260	
14	Fozzy	heavy_us_04	USA	7	Chasing the Grail (2010)	heavy_us_04-01	heavy_us_04-01-01	1. Under Blackened Skies	238	1,757
							heavy_us_04-01-02	2. Martyr No More	187	
							heavy_us_04-01-03	3. Grail	148	
							heavy_us_04-01-04	4. Broken Soul	160	
							heavy_us_04-01-05	5. Let the Madness Begin	169	
							heavy_us_04-01-06	6. Pray for Blood	162	
							heavy_us_04-01-07	7. New Day's Dawn	224	
							heavy_us_04-01-08	8. God Pounds His Nails	173	
							heavy_us_04-01-09	9. Watch Me Shine	172	
							heavy_us_04-01-10	10. Paraskavedekatriaphobia (Friday the 13th)	124	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
15	Impellitteri	heavy_us_05	USA	11	Venom (2015)	heavy_us_05-01	heavy_us_05-01-01	1. Venom	112	1,525
							heavy_us_05-01-02	2. Empire of Lies	127	
							heavy_us_05-01-03	3. We Own the Night	179	
							heavy_us_05-01-04	4. Nightmare	163	
							heavy_us_05-01-05	5. Face the Enemy	208	
							heavy_us_05-01-06	6. Domino Theory	120	
							heavy_us_05-01-07	7. Jehovah	144	
							heavy_us_05-01-08	8. Rise	136	
							heavy_us_05-01-09	9. Time Machine	131	
							heavy_us_05-01-10	10. Holding On	205	
16	Quiet Riot	heavy_us_06	USA	14	Metal Heath (1983)	heavy_us_06-01	heavy_us_06-01-01	1. Metal Health	235	1,936
							heavy_us_06-01-02	3. Don't Wanna Let You Go	156	
							heavy_us_06-01-03	5. Love's a Bitch	247	
							heavy_us_06-01-04	6. Breathless	169	
							heavy_us_06-01-05	7. Run for Cover	201	
							heavy_us_06-01-06	9. Let's Get Crazy	211	
							heavy_us_06-01-07	10. Thunderbird	220	
					QR III (1986)	heavy_us_06-02	heavy_us_06-02-01	1. Main Attraction	186	
							heavy_us_06-02-02	3. Twilight Hotel	128	
							heavy_us_06-02-03	10. Slave to Love	183	
17	Racer X	heavy_us_07	USA	5	Technical Difficulties (1999)	heavy_us_07-01	heavy_us_07-01-01	2. Fire of Rock	207	2,094
							heavy_us_07-01-02	3. Snakebite	156	
							heavy_us_07-01-03	5. Miss Mistreater	191	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							heavy_us_07-01-04	6. Bolt in My Heart	301	
							heavy_us_07-01-05	7. 17th Moon	249	
							heavy_us_07-01-06	8. Waiting	128	
							heavy_us_07-01-07	9. Poison Eyes	201	
							heavy_us_07-01-08	11. God of the Sun	205	
							heavy_us_07-01-09	12. Give It to Me	203	
							heavy_us_07-01-10	13. The Executioner's Song	253	
18	Sebastian Bach	heavy_us_08	USA	3	Kicking & Screaming (2011)	heavy_us_08-01	heavy_us_08-01-01	1. Kicking & Screaming	156	1,663
							heavy_us_08-01-02	2. My Own Worst Enemy	136	
							heavy_us_08-01-03	3. TunnelVision	190	
							heavy_us_08-01-04	4. Dance on Your Grave	132	
							heavy_us_08-01-05	5. Caught in a Dream	164	
							heavy_us_08-01-06	6. As Long as I Got the Music	233	
							heavy_us_08-01-07	7. I'm Alive	187	
							heavy_us_08-01-08	8. Dirty Power	163	
							heavy_us_08-01-09	9. Live the Life	100	
							heavy_us_08-01-10	10. Dream Forever	202	
19	Virgin Steele	heavy_us_09	USA	17	Guardians of the Flame (1983)	heavy_us_09-01	heavy_us_09-01-01	1. Don't Say Goodbye (Tonight)	136	1,692
							heavy_us_09-01-02	3. Life of Crime	129	
							heavy_us_09-01-03	4. The Redeemer	229	
							heavy_us_09-01-04	6. Guardians of the Flame	191	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							heavy_us_09-01-05	7. Metal City	115	
							heavy_us_09-01-06	8. Hell or High Water	148	
							heavy_us_09-01-07	9. Go All the Way	124	
							heavy_us_09-01-08	10. A Cry in the Night	197	
					Virgin Steele (1982)	heavy_us_09-02	heavy_us_09-02-01	3. Dead End Kids	151	
							heavy_us_09-02-02	9. Living in Sin	272	
20	W.A.S.P.	heavy_us_10	USA	16	The Headless Children (1989)	heavy_us_10-01	heavy_us_10-01-01	1. The Heretic (The Lost Child)	309	2,191
							heavy_us_10-01-02	3. The Headless Children	185	
							heavy_us_10-01-03	4. Thunderhead	289	
							heavy_us_10-01-04	5. Mean Man	239	
							heavy_us_10-01-05	6. The Neutron Bomber	128	
							heavy_us_10-01-06	8. Forever Free	157	
							heavy_us_10-01-07	9. Maneater	182	
							heavy_us_10-01-08	10. Rebel in the F.D.G.	289	
					Babylon (2009)	heavy_us_10-02	heavy_us_10-02-01	2. Live to Die Another Day	221	
							heavy_us_10-02-02	8. Godless Run	192	
	TOTAL				**33 albums**				**33,158**	**33,158**

E. THRASH METAL SUBCORPUS (TMSC)

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
1	Annihilated	thrash_uk_01	UK	5	Scorched Earth Policy (2010)	thrash_uk_01-01	thrash_uk_01-01-01	1. Blood of the Martyr	238	1,387
							thrash_uk_01-01-02	2. Scorched Earth Policy	124	
							thrash_uk_01-01-03	3. Dark Eyes of the Mind	119	
							thrash_uk_01-01-04	4. Predator	219	
							thrash_uk_01-01-05	5. The Burning of the Southern Cross	100	
							thrash_uk_01-01-06	6. A Cruel Twist of Fate	103	
							thrash_uk_01-01-07	7. Full Circle	132	
							thrash_uk_01-01-08	8. Despair and Retribution	105	
							thrash_uk_01-01-09	9. Death and Decay	113	
							thrash_uk_01-01-10	10. Until the Bitter End	134	
2	Evile	thrash_uk_02	UK	4	Enter the Grave (2007)	thrash_uk_02-01	thrash_uk_02-01-01	1. Enter the Grave	139	1,420
							thrash_uk_02-01-02	2. Thrasher	112	
							thrash_uk_02-01-03	4. Man Against Machine	151	
							thrash_uk_02-01-04	5. Burned Alive	152	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							thrash_uk_02-01-05	6. Killer from the Deep	146	
							thrash_uk_02-01-06	8. Schizophrenia	135	
							thrash_uk_02-01-07	9. Bathe in Blood	175	
							thrash_uk_02-01-08	10. Armoured Assault	125	
					Infected Nations (2009)	thrash_uk_02-02	thrash_uk_02-02-01	3. Nosophoros	130	
							thrash_uk_02-02-02	4. Genocide	155	
3	Lawnmower Deth	thrash_uk_03	UK	3	Ooh Crikey It's... Lawnmower Deth (1990)	thrash_uk_03-01	thrash_uk_03-01-01	2. Betty Ford's Clinic	105	1,553
							thrash_uk_03-01-02	4. Sheep Dip	126	
							thrash_uk_03-01-03	5. Lancer with Your Zancer	257	
							thrash_uk_03-01-04	7. Flying Killer Cobs from the Planet Bob	127	
							thrash_uk_03-01-05	10. Rad Dude	145	
							thrash_uk_03-01-06	11. Sumo Rabbit and His Inescapable Trap of Doom	155	
							thrash_uk_03-01-07	15. Icky Ficky	105	
							thrash_uk_03-01-08	18. Satan's Trampoline	101	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
					Return of the Fabulous Metal Bozo Clowns (1992)	thrash_uk_03-02	thrash_uk_03-02-01	1. The Return of the Fabulous Metal Bozo Clown	281	
							thrash_uk_03-02-02	4. Feetcleaner	151	
4	Onslaught	thrash_uk_04	UK	6	Sounds of Violence (2011)	thrash_uk_04-01	thrash_uk_04-01-01	2. Born for War	231	1,904
							thrash_uk_04-01-02	3. The Sound of Violence	211	
							thrash_uk_04-01-03	4. Code Black	159	
							thrash_uk_04-01-04	5. Rest in Pieces	218	
							thrash_uk_04-01-05	6. Godhead	132	
							thrash_uk_04-01-06	7. Hatebox	191	
							thrash_uk_04-01-07	8. Antitheist	228	
							thrash_uk_04-01-08	9. Suicideology	202	
					Killing Peace (2007)	thrash_uk_04-02	thrash_uk_04-02-01	1. Burn	193	
							thrash_uk_04-02-01	2. Killing Peace	139	
5	Psython	thrash_uk_05	UK	3	Hatred (2017)	thrash_uk_05-01	thrash_uk_05-01-01	1. Jörmungandr	181	1,852
							thrash_uk_05-01-02	2. Battery Life	175	
							thrash_uk_05-01-03	3. Teeth	154	
							thrash_uk_05-01-04	4. H.A.T.E.	208	
							thrash_uk_05-01-05	5. Chai Latte	147	
							thrash_uk_05-01-06	6. Ten Pounds	197	
							thrash_uk_05-01-07	7. One in Deed	119	
							thrash_uk_05-01-08	8. Hashtrap	252	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							thrash_uk_05-01-09	9. Old Man	252	
					…Outputs (2016)	thrash_uk_05-02	thrash_uk_05-02-01	2. Careless Whispers	167	
6	Re-Animator	thrash_uk_06	UK	3	Condemned to Eternity (1990)	thrash_uk_06-01	thrash_uk_06-01-01	1. Low Life	184	1,971
							thrash_uk_06-01-02	2. Chain of Command	117	
							thrash_uk_06-01-03	6. Buried Alive	126	
							thrash_uk_06-01-04	7. Techno Fear	104	
							thrash_uk_06-01-05	9. Say Your Prayers	153	
					That Was Then... This Is Now (1992)	thrash_uk_06-02	thrash_uk_06-02-01	1. Take Me Away	241	
							thrash_uk_06-02-02	4. Hope	297	
							thrash_uk_06-02-03	5. Last Laugh	333	
							thrash_uk_06-02-04	6. Kick Back	178	
							thrash_uk_06-02-05	7. Listen Up	238	
7	Sabbat	thrash_uk_07	UK	3	History of a Time to Come (1988)	thrash_uk_07-01	thrash_uk_07-01-01	2. A Cautionary Tale	350	2,326
							thrash_uk_07-01-02	3. Hosanna in Excelsis	174	
							thrash_uk_07-01-03	4. Behind the Crooked Cross	261	
							thrash_uk_07-01-04	5. Horned Is the Hunter	377	
							thrash_uk_07-01-05	6. I for an Eye	272	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							thrash_uk_07-01-06	7. For Those Who Died	166	
							thrash_uk_07-01-07	9. The Church Bizarre	290	
					Dreamweaver (1989)	thrash_uk_07-02	thrash_uk_07-02-01	3. Advent of Insanity	174	
					Mourning Has Broken (1991)	thrash_uk_07-03	thrash_uk_07-03-01	3. Paint the World Black	128	
							thrash_uk_07-03-02	5. The Voice of Time	134	
8	Toranaga	thrash_uk_08	UK	3	God's Gift (1990)	thrash_uk_08-01	thrash_uk_08-01-01	1. The Shrine	322	2,022
							thrash_uk_08-01-02	2. Psychotic	247	
							thrash_uk_08-01-03	3. Sword of Damocles	144	
							thrash_uk_08-01-04	4. Hammer to the Skull	152	
							thrash_uk_08-01-05	5. Food of the Gods	214	
							thrash_uk_08-01-06	6. Disciples	136	
							thrash_uk_08-01-07	8. Black Is the Mask	174	
					Righteous Retribution (2013)	thrash_uk_08-02	thrash_uk_08-02-01	2. Traitors Gate	235	
							thrash_uk_08-02-02	3. Cynical Eyes	167	
							thrash_uk_08-02-03	4. The Ultimate Act of Betrayal	231	
9	Gama Bomb	thrash_uk_09	UK	6	Tales from the	thrash_uk_09-01	thrash_uk_09-01-01	1. Slam Anthem	258	2,026

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
					Grave in Space (2009)		thrash_uk_09-01-02	2. New Eliminators of Atlantis B.C.	186	
							thrash_uk_09-01-03	3. Three Witches	206	
							thrash_uk_09-01-04	4. Last Ninjas Unite	216	
							thrash_uk_09-01-05	5. Escape from Scarecrow Mountain	238	
							thrash_uk_09-01-06	6. Mussolini Mosh	103	
							thrash_uk_09-01-07	7. We Respect You	202	
							thrash_uk_09-01-08	8. Apocalypse 1997	212	
							thrash_uk_09-01-09	9. Return to Blood Castle	206	
							thrash_uk_09-01-10	10. Polterghost	199	
10	Xentrix	thrash_uk_10	UK	4	Shattered Existence (1989)	thrash_uk_10-01	thrash_uk_10-01-01	1. No Compromise	187	1,447
							thrash_uk_10-01-02	2. Balance of Power	151	
							thrash_uk_10-01-03	3. Crimes	152	
							thrash_uk_10-01-04	4. Back in the Real World	111	
							thrash_uk_10-01-05	5. Dark Enemy	144	
							thrash_uk_10-01-06	6. Bad Blood	109	
							thrash_uk_10-01-07	7. Reasons for Destruction	133	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							thrash_uk_10-01-08	8. Position of Security	161	
							thrash_uk_10-01-09	9. Heaven Cent	123	
					For Whose Advantage (1990)	thrash_uk_10-02	thrash_uk_10-02-01	1. Questions	176	
11	Anthrax	thrash_us_01	USA	12	Spreading the Disease (1985)	thrash_us_01-01	thrash_us_01-01-01	1. A.I.R.	209	1,710
							thrash_us_01-01-02	2. Lone Justice	197	
							thrash_us_01-01-03	3. Madhouse	137	
							thrash_us_01-01-04	4. S.S.C. / Stand or Fall	162	
							thrash_us_01-01-05	5. The Enemy	188	
							thrash_us_01-01-06	6. Aftershock	125	
							thrash_us_01-01-07	7. Armed and Dangerous	152	
							thrash_us_01-01-08	8. Medusa	158	
							thrash_us_01-01-09	9. Gung-Ho	164	
					State of Euphoria (1988)	thrash_us_01-02	thrash_us_01-02-01	1. Be All, End All	218	
12	Death Angel	thrash_us_02	USA	8	Frolic Through the Park (1988)	thrash_us_02-01	thrash_us_02-01-01	1. 3rd Floor	102	1,667
							thrash_us_02-01-02	2. Road Mutants	194	
							thrash_us_02-01-03	3. Why You Do This	192	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							thrash_us_02-01-04	6. Confused	189	
							thrash_us_02-01-05	7. Guilty of Innocence	147	
							thrash_us_02-01-06	8. Open Up	173	
							thrash_us_02-01-07	9. Shores of Sin	140	
							thrash_us_02-01-08	11. Mind Rape	114	
					The Dream Calls for Blood (2013)	thrash_us_02-02	thrash_us_02-02-01	1. Left for Dead	279	
							thrash_us_02-02-02	2. Son of the Morning	137	
13	Exodus	thrash_us_03	USA	11	Bonded by Blood (1985)	thrash_us_03-01	thrash_us_03-01-01	1. Bonded by Blood	146	1,622
							thrash_us_03-01-02	2. Exodus	190	
							thrash_us_03-01-03	3. And Then There Were None	133	
							thrash_us_03-01-04	4. A Lesson in Violence	140	
							thrash_us_03-01-05	5. Metal Command	181	
							thrash_us_03-01-06	6. Piranha	124	
							thrash_us_03-01-07	7. No Love	154	
							thrash_us_03-01-08	8. Deliver Us to Evil	203	
							thrash_us_03-01-09	9. Strike of the Beast	202	
					Blood In	thrash_us_03-02	thrash_us_03-02-01	1. Black 13	149	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
					Blood Out (2014)					
14	Flotsam and Jetsam	thrash_us_04	USA	14	The End of Chaos (2019)	thrash_us_04-01	thrash_us_04-01-01	1. Prisoner of Time	153	1,707
							thrash_us_04-01-02	2. Control	170	
							thrash_us_04-01-03	3. Recover	113	
							thrash_us_04-01-04	4. Prepare for Chaos	219	
							thrash_us_04-01-05	5. Slowly Insane	117	
							thrash_us_04-01-06	6. Architects of Hate	104	
							thrash_us_04-01-07	7. Demolition Man	194	
							thrash_us_04-01-08	8. Unwelcome Surprise	224	
							thrash_us_04-01-09	9. Snake Eye	236	
							thrash_us_04-01-10	10. Survive	177	
15	Megadeth	thrash_us_05	USA	15	Rust in Peace (1990)	thrash_us_05-01	thrash_us_05-01-01	1. Holy Wars... The Punishment Due	260	2,196
							thrash_us_05-01-02	3. Take No Prisoners	220	
							thrash_us_05-01-03	4. Five Magics	155	
							thrash_us_05-01-04	5. Poison Was the Cure	142	
							thrash_us_05-01-05	6. Lucretia	124	
							thrash_us_05-01-06	7. Tornado of Souls	321	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							thrash_us_05-01-07	9. Rust in Peace... Polaris	233	
					Peace Sells... But Who's Buying? (1986)	thrash_us_05-02	thrash_us_05-02-01	2.The Conjuring	198	
							thrash_us_05-02-02	4. Devils Island	264	
							thrash_us_05-02-03	8. My Last Words	279	
16	Metallica	thrash_us_06	USA	11	Kill 'Em All (1983)	thrash_us_06-01	thrash_us_06-01-01	2. The Four Horsemen	205	1,649
							thrash_us_06-01-02	3. Motorbreath	143	
							thrash_us_06-01-03	4. Jump in the Fire	191	
							thrash_us_06-01-04	6. Whiplash	203	
							thrash_us_06-01-05	7. Phantom Lord	112	
							thrash_us_06-01-06	8. No Remorse	183	
							thrash_us_06-01-07	9. Seek & Destroy	149	
							thrash_us_06-01-08	10. Metal Militia	147	
					Ride the Lightning (1984)	thrash_us_06-02	thrash_us_06-02-01	2. Ride The Lightning	164	
							thrash_us_06-02-02	3. For Whom the Bell Tolls	152	
17	Nuclear Assault	thrash_us_07	USA	6	Game Over (1986)	thrash_us_07-01	thrash_us_07-01-01	2. Sin	204	1,600
							thrash_us_07-01-02	3. Cold Steel	118	
							thrash_us_07-01-03	5. Radiation Sickness	106	
							thrash_us_07-01-04	7. After the	136	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
								Holocaust		
							thrash_us_07-01-05	9. Stranded in Hell	208	
							thrash_us_07-01-06	10. Nuclear War	101	
							thrash_us_07-01-07	12. Vengeance	160	
							thrash_us_07-01-08	13. Brain Death	205	
					Survive (1988)	thrash_us_07-02	thrash_us_07-02-01	2. Brainwashed	186	
							thrash_us_07-02-02	5. Fight to Be Free	176	
18	Overkill	thrash_us_08	USA	19	Feel the Fire (1985)	thrash_us_08-01	thrash_us_08-01-01	1. Raise the Dead	175	1,770
							thrash_us_08-01-02	2. Rotten to the Core	153	
							thrash_us_08-01-03	3. There's No Tomorrow	173	
							thrash_us_08-01-04	4. Second Son	164	
							thrash_us_08-01-05	5. Hammerhead	156	
							thrash_us_08-01-06	6. Feel the Fire	278	
							thrash_us_08-01-07	7. Blood and Iron	114	
							thrash_us_08-01-08	8. Kill at Command	182	
							thrash_us_08-01-09	9. Overkill	236	
					The Years of Decay (1989)	thrash_us_08-02	thrash_us_08-02-01	1. Time to Kill	139	
19	Slayer	thrash_us_09	USA	12	Seasons in the Abyss (1990)	thrash_us_09-01	thrash_us_09-01-01	1. War Ensemble	226	1,603
							thrash_us_09-01-02	2. Blood Red	104	
							thrash_us_09-01-03	3. Spirit in Black	179	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							thrash_us_09-01-04	5. Dead Skin Mask	138	
							thrash_us_09-01-05	6. Hallowed Point	139	
							thrash_us_09-01-06	7. Skeletons of Society	192	
							thrash_us_09-01-07	8. Temptation	161	
							thrash_us_09-01-08	9. Born of Fire	211	
							thrash_us_09-01-09	10. Seasons in the Abyss	115	
					South of Heaven (1988)	thrash_us_09-02	thrash_us_09-02-01	1. South of Heaven	138	
20	Testament	thrash_us_10	USA	12	Low (1994)	thrash_us_10-01	thrash_us_10-01-01	1. Low	138	1,367
							thrash_us_10-01-02	2. Legions (In Hiding)	137	
							thrash_us_10-01-03	3. Hail Mary	149	
							thrash_us_10-01-04	4. Trail of Tears	135	
							thrash_us_10-01-05	5. Shades of War	122	
							thrash_us_10-01-06	6. P.C.	153	
							thrash_us_10-01-07	7. Dog Faced Gods	115	
							thrash_us_10-01-08	8. All I Could Bleed	159	
							thrash_us_10-01-09	10. Chasing Fear	156	
							thrash_us_10-01-10	11. Ride	103	
	TOTAL				**37 albums**				**34,799**	**34,799**

F. DEATH METAL SUBCORPUS (DMSC)

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
1	Benediction	death_uk_01	UK	7	Grind Bastard (1998)	death_uk_01-01	death_uk_01-01-01	1. Deadfall	254	1,880
							death_uk_01-01-02	2. Agonised	159	
							death_uk_01-01-03	3. West of Hell	160	
							death_uk_01-01-04	4. Magnificat	173	
							death_uk_01-01-05	5. Nervebomb	170	
							death_uk_01-01-06	7. Grind Bastard	362	
							death_uk_01-01-07	8. Shadow World	155	
							death_uk_01-01-08	9. The Bodiless	121	
							death_uk_01-01-09	10. Carcinoma Angel	132	
							death_uk_01-01-10	11. We the Freed	194	
2	Bolt Thrower	death_uk_02	UK	8	…For Victory (1994)	death_uk_02-01	death_uk_02-01-01	3. When Glory Beckons	101	1,312
							death_uk_02-01-02	4. …for Victory	144	
							death_uk_02-01-03	7. Silent Demise	105	
							death_uk_02-01-04	10. Armageddon Bound	120	
					The IVth Crusade (1992)	death_uk_02-02	death_uk_02-02-01	3. Embers	177	
							death_uk_02-02-02	5. As the World Burns	122	
							death_uk_02-02-03	6. This Time It's	100	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
								War		
							death_uk_02-02-04	7. Ritual	154	
							death_uk_02-02-05	8. Spearhead	163	
							death_uk_02-02-06	9. Celestial Sanctuary	126	
3	Carcass	death_uk_03	UK	6	Necroticism - Descanting the Insalubrious (1991)	death_uk_03-01	death_uk_03-01-02	2. Corporal Jigsore Quandary	257	2,065
							death_uk_03-01-03	3. Symposium of Sickness	397	
							death_uk_03-01-04	4. Pedigree Butchery	194	
							death_uk_03-01-05	5. Incarnated Solvent Abuse	128	
							death_uk_03-01-06	6. Carneous Cacoffiny	246	
							death_uk_03-01-07	8. Forensic Clinicism / The Sanguine Article	324	
					Heartwork (1993)	death_uk_03-02	death_uk_03-02-01	2. Carnal Forge	110	
							death_uk_03-02-02	4. Heartwork	123	
							death_uk_03-02-03	5. Embodiment	163	
							death_uk_03-02-04	7. Arbeit macht Fleisch	123	
4	Dāmim	death_uk_04	UK	3	Purity: The Darwinian	death_uk_04-01	death_uk_04-01-01	1. Spiritual Void	139	1,573
							death_uk_04-01-02	2. City of Envy	205	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
					Paradox (2005)		death_uk_04-01-03	3. No God with Me	171	
							death_uk_04-01-04	4. Come to Dust	137	
							death_uk_04-01-05	7. Body Temples of Sorrow	126	
							death_uk_04-01-06	10. Fortunes of Need	103	
							death_uk_04-01-07	11. Frightening and Obscene	175	
					The Difference Engine (2007)	death_uk_04-02	death_uk_04-02-01	1. The Difference Engine	196	
							death_uk_04-02-02	2. Eyeballing	184	
							death_uk_04-02-03	3. Outside	137	
5	Dead Beyond Buried	death_uk_05	UK	3	Condemned to Misery (2007)	death_uk_05-01	death_uk_05-01-01	2. Bow Down or Die	167	1,611
							death_uk_05-01-02	3. Condemned to Misery	157	
							death_uk_05-01-03	4. Aura of I	169	
							death_uk_05-01-04	6. Dissalution	141	
							death_uk_05-01-05	7. Organic Dementia	147	
							death_uk_05-01-06	8. God Is Dead	104	
							death_uk_05-01-07	9. Spear of Longinus	119	
							death_uk_05-01-08	10. Innocence	204	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
								Erased		
							death_uk_05-01-09	11. Life Slowly Tortured	182	
							death_uk_05-01-10	12. Rape Your Kingdom	221	
6	Desecration	death_uk_06	UK	8	Cemetery Sickness (2014)	death_uk_06-01	death_uk_06-01-01	1. Cemetery Sickness	132	1,288
							death_uk_06-01-02	4. Recipes of Horror	148	
							death_uk_06-01-03	6. Cunt Full of Maggots	130	
							death_uk_06-01-04	7. Cabletie Castrator	130	
							death_uk_06-01-05	8. Mortuary Debauchery	110	
							death_uk_06-01-06	10. Cut Up & Fed to the Dog	124	
					Pathway to Deviance (2002)	death_uk_06-02	death_uk_06-02-01	1. Cleaver, Saw and Butcher's Knife	126	
							death_uk_06-02-02	2. Offer the Flesh	122	
							death_uk_06-02-03	3. King of the Missing	141	
							death_uk_06-02-04	4. Bloody Human Carvery	125	
7	Dyscarnate	death_uk_07	UK	3	Enduring the	death_uk_07-01	death_uk_07-01-01	1. An Axe to	213	1,970

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
					Massacre (2010)			Grind		
							death_uk_07-01-02	2. Despised and Disgraced	185	
							death_uk_07-01-03	3. Extinguishing the Face of Heaven	167	
							death_uk_07-01-04	4. Yielding the Iron Fist	174	
							death_uk_07-01-05	5. Judecca	192	
							death_uk_07-01-06	6. The Vitruvian Plan	159	
							death_uk_07-01-07	7. Those Who Trespass Against Us	267	
							death_uk_07-01-08	8. Enduring the Massacre	222	
					And So It Came to Pass (2012)	death_uk_07-02	death_uk_07-02-01	2. In the Face of Armageddon	174	
							death_uk_07-02-02	3. Cain Enable	217	
8	Gorerotted	death_uk_08	UK	3	A New Dawn for the Dead (2005)	death_uk_08-01	death_uk_08-01-01	1. ...and Everything Went Black	287	2,124
							death_uk_08-01-02	2. Pain as a Prelude to Death	126	
							death_uk_08-01-03	3. Nervous Gibbering Wreck	215	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							death_uk_08-01-04	4. Adding Insult to Injury	209	
							death_uk_08-01-05	5. Fable of Filth	119	
							death_uk_08-01-06	6. Dead Drunk	257	
							death_uk_08-01-07	7. A Very Grave Business	275	
							death_uk_08-01-08	8. Horrorday in Haiti	265	
							death_uk_08-01-09	9. Selection and Dissection of Parts for Resurrection	206	
					Only Tools and Corpses (2003)	death_uk_08-02	death_uk_08-02-01	2. Hacked in the Back Dumped in a Sack	165	
9	Napalm Death	death_uk_09	UK	15	Fear, Emptiness, Despair (1994)	death_uk_09-01	death_uk_09-01-01	5. More than Meets the Eye	107	1,162
							death_uk_09-01-02	6. Primed Time	124	
							death_uk_09-01-03	7. State of Mind	102	
							death_uk_09-01-04	8. Armageddon X 7	111	
							death_uk_09-01-05	10. Fasting on Deception	123	
							death_uk_09-01-06	11. Throwaway	127	
					Words from	death_uk_09-02	death_uk_09-02-01	1. The Infiltraitor	112	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
					the Exit Wound (1998)		death_uk_09-02-02	2. Repression Out of Uniform	126	
							death_uk_09-02-03	4. Trio-Degradable / Affixed by Disconcern	121	
							death_uk_09-02-04	6. Devouring Depraved	109	
10	Necrosanct	death_uk_10	UK	3	Equal in Death (1990)	death_uk_10-01	death_uk_10-01-01	1. Arachneurosis	215	2,224
							death_uk_10-01-02	2. Pretentious Priests	169	
							death_uk_10-01-03	3. Trial by Fire	159	
							death_uk_10-01-04	4. Besieged Citadel	296	
							death_uk_10-01-05	6. In Death	173	
							death_uk_10-01-06	7. Vigilante	276	
							death_uk_10-01-07	8. Equananimous Deterioration	381	
							death_uk_10-01-08	9. Necronomicon	313	
					Incarnate (1992)	death_uk_10-02	death_uk_10-02-01	1. Ritual Acts	118	
							death_uk_10-02-02	2. Inevitable Demise	124	
11	Autopsy	death_us_01	USA	7	Macabre Eternal	death_us_01-01	death_us_01-01-01	1. Hand of Darkness	144	1,725

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
					(2011)		death_us_01-01-02	2. Dirty Gore Whore	178	
							death_us_01-01-03	3. Always About to Die	148	
							death_us_01-01-04	4. Macabre Eternal	172	
							death_us_01-01-05	5. Deliver Me from Sanity	114	
							death_us_01-01-06	6. Seeds of the Doomed	125	
							death_us_01-01-07	7. Bridge of Bones	235	
							death_us_01-01-08	8. Born Undead	114	
							death_us_01-01-09	9. Sewn into One	177	
							death_us_01-01-10	11. Sadistic Gratification	318	
12	Cannibal Corpse	death_us_02	USA	14	Eaten back to Life (1990)	death_us_02-01	death_us_02-01-01	1. Shredded Humans	317	2,252
							death_us_02-01-02	2. Edible Autopsy	193	
							death_us_02-01-03	4. Mangled	381	
							death_us_02-01-04	5. Scattered Remains, Splattered Brains	146	
							death_us_02-01-05	6. Born in a Casket	145	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							death_us_02-01-06	7. Rotting Head	183	
							death_us_02-01-07	8. The Undead Will Feast	320	
							death_us_02-01-08	9. Bloody Chunks	227	
							death_us_02-01-09	10. A Skull Full of Maggots	101	
							death_us_02-01-10	11. Buried in the Backyard	239	
13	Death	death_us_03	USA	7	Symbolic (1995)	death_us_03-01	death_us_03-01-01	1. Symbolic	139	1,289
							death_us_03-01-02	2. Zero Tolerance	124	
							death_us_03-01-03	3. Empty Words	157	
							death_us_03-01-04	4. Sacred Serenity	103	
							death_us_03-01-05	5. 1,000 Eyes	106	
							death_us_03-01-06	7. Crystal Mountain	123	
							death_us_03-01-07	8. Misanthrope	127	
							death_us_03-01-08	9. Perennial Quest	171	
					The Sound of Perseverance (1998)	death_us_03-02	death_us_03-02-01	1. Scavenger of Human Sorrow	120	
							death_us_03-02-02	2. Bite the Pain	119	
14	Deicide	death_us_04	USA	12	Deicide (1990)	death_us_04-01	death_us_04-01-01	2. Sacrificial Suicide	165	1,719
							death_us_04-01-02	3. Oblivious to Evil	143	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							death_us_04-01-03	4. Dead by Dawn	160	
							death_us_04-01-04	6. Deicide	161	
							death_us_04-01-05	7. Carnage in the Temple of the Damned	110	
							death_us_04-01-06	8. Mephistopheles	107	
							death_us_04-01-07	10. Crucifixation	136	
					Scars of the Crucifix (2004)	death_us_04-02	death_us_04-02-01	1. Scars of the Crucifix	213	
							death_us_04-02-02	2. Mad at God	368	
							death_us_04-02-03	3. Conquered by Sodom	156	
15	Malevolent Creation	death_us_05	USA	13	Retribution (1992)	death_us_05-01	death_us_05-01-01	1. Eve of the Apocalypse	187	1,900
							death_us_05-01-02	2. Systematic Execution	151	
							death_us_05-01-03	3. Slaughter of Innocence	241	
							death_us_05-01-04	4. Coronation of Our Domain	238	
							death_us_05-01-05	5. No Flesh Shall Be Spared	249	
							death_us_05-01-06	6. The Coldest Survive	183	
							death_us_05-01-07	7. Monster	142	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							death_us_05-01-08	8. Mindlock	156	
							death_us_05-01-09	9. Iced	248	
					The Will to Kill (2002)	death_us_05-02	death_us_05-02-01	1. The Will to Kill	105	
16	Massacre	death_us_06	USA	3	Back from Beyond (2014)	death_us_06	death_us_06-01-01	2. As We Wait to Die	184	1,543
							death_us_06-01-02	3. Ascension of the Deceased	104	
							death_us_06-01-03	4. Hunter's Blood	144	
							death_us_06-01-04	5. Darkness Fell	154	
							death_us_06-01-05	6. False Revelation	185	
							death_us_06-01-06	7. Succumb to Rapture	169	
							death_us_06-01-07	8. Remnants of Hatred	136	
							death_us_06-01-08	9. Shield of the Son	176	
							death_us_06-01-09	10. The Evil Within	153	
							death_us_06-01-10	11. Sands of Time	138	
17	Monstrosity	death_us_07	USA	6	Millennium (1996)	death_us_07-01	death_us_07-01-01	1. Fatal Millennium	220	1,570
							death_us_07-01-02	2. Devious Instinct	146	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							death_us_07-01-03	3. Manic	202	
							death_us_07-01-04	4. Dream Messiah	148	
							death_us_07-01-05	6. Manipulation Strain	164	
							death_us_07-01-06	7. Slaves and Masters	106	
							death_us_07-01-07	9. Stormwinds	171	
							death_us_07-01-08	10. Seize of Change	157	
					Rise to Power (2003)	death_us_07-02	death_us_07-02-01	1.The Exordium	120	
							death_us_07-02-02	2. Awaiting Armageddon	136	
18	Morbid Angel	death_us_08	USA	10	Domination (1995)	death_us_08-01	death_us_08-01-01	1. Dominate	135	1,499
							death_us_08-01-02	2. Where the Slime Live	166	
							death_us_08-01-03	3. Eyes to See, Ears to Hear	129	
							death_us_08-01-04	5. Nothing but Fear	187	
							death_us_08-01-05	6. Dawn of the Angry	135	
							death_us_08-01-06	7. This Means War	184	
							death_us_08-01-07	8. Caesar's Palace	123	
							death_us_08-01-08	10. Inquisition	193	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
								(Burn with Me)		
					Covenant (1993)	death_us_08-02	death_us_08-02-01	1. Rapture	115	
							death_us_08-02-02	3. World of Shit (The Promised Land)	132	
19	Obituary	death_us_09	USA	10	The End Complete (1992)	death_us_09-01	death_us_09-01-01	1. I'm in Pain	116	1,282
							death_us_09-01-02	2. Back to One	104	
							death_us_09-01-03	5. Sickness	186	
							death_us_09-01-04	7. Killing Time	104	
							death_us_09-01-05	8. The End Complete	110	
					World Demise (1994)	death_us_09-02	death_us_09-02-01	1. Don't Care	136	
							death_us_09-02-02	3. Burned In	131	
							death_us_09-02-03	4. Redefine	140	
							death_us_09-02-04	5. Paralyzing	112	
							death_us_09-02-05	6. Lost	143	
20	Possessed	death_us_10	USA	3	Seven Churches (1985)	death_us_10-01	death_us_10-01-01	1. The Exorcist	189	1,871
							death_us_10-01-02	2. Pentagram	126	
							death_us_10-01-03	3. Burning in Hell	207	
							death_us_10-01-04	4. Evil Warriors	331	
							death_us_10-01-05	5. Seven Churches	118	
							death_us_10-01-06	6. Satan's Curse	149	

No	Band	Band code	Country	# of Full-length albums	Album	Album code	Song code	Song Title	# of Words	Total # of words
							death_us_10-01-07	7. Holy Hell	170	
							death_us_10-01-08	8. Twisted Minds	149	
							death_us_10-01-09	10. Death Metal	197	
					Revelations of Oblivion (2019)	death_us_10-02	death_us_10-02-01	2. No More Room in Hell	235	
	TOTAL				**35 albums**				**33,859**	**33,859**

G. REFERENCES TO THE BANDS IN MC

Subcorpus	No	Band	Country	References
HMSC	1	Angel Witch	UK	Christe (2003), Klepper et al. (2007), Phillips and Cogan (2009), Metal Evolution
HMSC	2	Black Sabbath	UK	Walser (1993), Arnett (1996), Weinstein (2000), Shuker (2001), Konow (2002), Christe (2003), Morrison (2006), Buts and Buelens (2008), Phillips and Cogan (2009), Phillipov (2012)
HMSC	3	Blaze Bayley	UK	None
HMSC	4	Fist	UK	Metal Evolution
HMSC	5	Girlschool	UK	Christe (2003), Metal Evolution
HMSC	6	Iron Maiden	UK	Walser (1993), Arnett (1996), Berger (1999), Weinstein (2000), Shuker (2001), Christe (2003), Klepper et al. (2007), Farley (2009), Phillips and Cogan (2009), Weinstein (2009), Davisson (2010), Tsatsishvili (2011), Phillipov (2012), Metal Evolution
HMSC	7	Judas Priest	UK	Walser (1993), Arnett (1996), Berger (1999), Weinstein (2000), Christe (2003), Klepper et al. (2007), Buts and Buelens (2008), Phillips and Cogan (2009), Weinstein (2009), Davisson (2010), Phillipov (2012)
HMSC	8	Saxon	UK	Christe (2003), Klepper et al. (2007), Farley (2009), Weinstein (2009), Tsatsishvili (2011), Metal Evolution
HMSC	9	Tank	UK	Metal Evolution
HMSC	10	Tygers of Pan Tang	UK	Klepper et al. (2007), Farley (2009), Metal Evolution
HMSC	11	Armored Saint	USA	Phillips and Cogan (2009)
HMSC	12	Dee Snider	USA	None
HMSC	13	Dokken	USA	Christe (2003)
HMSC	14	Fozzy	USA	None
HMSC	15	Impellitteri	USA	None
HMSC	16	Quiet Riot	USA	Christe (2003)
HMSC	17	Racer X	USA	None
HMSC	18	Sebastian Bach	USA	None
HMSC	19	Virgin Steele	USA	None
HMSC	20	W.A.S.P.	USA	Weinstein (2000)
TMSC	1	Annihilated	UK	None
TMSC	2	Evile	UK	None
TMSC	3	Lawnmower Deth	UK	None
TMSC	4	Onslaught	UK	Farley (2009)

TMSC	5	Psython	UK	None
TMSC	6	Re-Animator	UK	None
TMSC	7	Sabbat	UK	Phillips and Cogan (2009)
TMSC	8	Toranaga	UK	None
TMSC	9	Virus	UK	None
TMSC	10	Xentrix	UK	None
TMSC	11	Anthrax	USA	Arnett (1996), Weinstein (2000), Christe (2003), Piccoli (2003), Purcell (2003), Pillsbury (2006), Pieslak (2007), Phillips and Cogan (2009), Weinstein (2009), Davisson (2010), Phillipov (2012), Strother (2013), Metal Evolution
TMSC	12	Metallica	USA	Walser (1993), Arnett (1996), Fast (2001), Christe (2003), Piccoli (2003), Purcell (2003), Klepper et al. (2007), Pieslak (2007), Buts and Buelens (2008), Farley (2009), Phillips and Cogan (2009), Weinstein (2009), Davisson (2010), Hutcherson and Haenfler (2010), Phillipov (2012), Strother (2013), Metal Evolution
TMSC	13	Slayer	USA	Walser (1993), Arnett (1996), Weinstein (2000), Christe (2003), Piccoli (2003), Purcell (2003), Pillsbury (2006), Pieslak (2007), Phillips and Cogan (2009), Weinstein (2009), Davisson (2010), Tsatsishvili (2011), Phillipov (2012), Strother (2013), Kitteringham (2014) Metal Evolution
TMSC	14	Megadeth	USA	Walser (1993), Arnett (1996), Weinstein (2000), Fast (2001), Christe (2003), Piccoli (2003), Purcell (2003), Pillsbury (2006), Phillips and Cogan (2009), Weinstein (2009), Davisson (2010), Phillipov (2012), Strother (2013), Metal Evolution
TMSC	15	Exodus	USA	Walser (1993), Weinstein (2000), Christe (2003), Pillsbury (2006), Phillips and Cogan (2009), Davisson (2010), Metal Evolution,
TMSC	16	Nuclear Assault	USA	Weinstein (2000), Christe (2003)
TMSC	17	Death Angel	USA	Christe (2003), Metal Evolution
TMSC	18	Testament	USA	Walser (1993), Christe (2003), Farley (2009), Metal Evolution
TMSC	19	Flotsam and Jetsam	USA	Weinstein (2000), Phillips and Cogan (2009)
TMSC	20	Overkill	USA	Phillips and Cogan (2009), Metal Evolution,

DMSC	1	Amputated	UK	None
DMSC	2	Benediction	UK	Mudrian and Peel (2004)
DMSC	3	Bolt Thrower	UK	Purcell (2003), Phillips and Cogan (2009)
DMSC	4	Carcass	UK	Christe (2003), Purcell (2003), Mudrian and Peel (2004), Phillips and Cogan (2009), Davisson (2010), Phillipov (2012)
DMSC	5	Dead Beyond Buried	UK	None
DMSC	6	Desecration	UK	None
DMSC	7	Gorerotted	UK	None
DMSC	8	Napalm Death	UK	Christe (2003), Mudrian and Peel (2004), Phillips and Cogan (2009), Phillipov (2012)
DMSC	9	Necrosanct	UK	None
DMSC	10	Vallenfyre	UK	None
DMSC	11	Autopsy	USA	Metal Evolution
DMSC	12	Cannibal Corpse	USA	Weinstein (2000), Christe (2003), Piccoli (2003), Purcell (2003), Mudrian and Peel (2004), Morrison (2006), Phillips and Cogan (2009), Tsatshisvili (2011), Mishrell (2012), Phillipov (2012), Metal Evolution, Strother (2013),
DMSC	13	Death	USA	Weinstein (2000), Christe (2003), Piccoli (2003), Purcell (2003), Mudrian and Peel (2004), Morrison (2006), Phillips and Cogan (2009), Davisson (2010), Tsatshisvili (2011), Phillipov (2012), Strother (2013), Kitteringham (2014), Metal Evolution
DMSC	14	Deicide	USA	Weinstein (2000), Christe (2003), Purcell (2003), Mudrian and Peel (2004), Morrison (2006), Phillips and Cogan (2009), Tsatshisvili (2011), Mishrell (2012), Phillipov (2012), Metal Evolution
DMSC	15	Malevolent Creation	USA	Weinstein (2000), Christe (2003), Mudrian and Peel (2004)
DMSC	16	Massacre	USA	Mudrian and Peel (2004)
DMSC	17	Monstrosity	USA	Purcell (2003)
DMSC	18	Morbid Angel	USA	Weinstein (2000), Christe (2003), Piccoli (2003), Purcell (2003), Mudrian and Peel (2004), Morrison (2006), Phillips and Cogan (2009), Davisson (2010), Tsatshisvili (2011), Mishrell (2012), Phillipov (2012), Strother (2013), Kitteringham (2014), Metal Evolution

DMSC	19	Obituary	USA	Weinstein (2000), Piccoli (2003), Purcell (2003), Mudrian and Peel (2004), Morrison (2006), Davisson (2010), Tsatshisvili (2011), Mishrell (2012), Phillipov (2012), Strother (2013), Metal Evolution
DMSC	20	Possessed	USA	Purcell (2003), Mudrian and Peel (2004), Morrison (2006), Phillips and Cogan (2009), Davisson (2010), Strother (2013), Kitteringham (2014), Metal Evolution