

UNIVERSITY OF HELSINKI

Keywords in heavy metal lyrics

A Data-Driven Corpus Study into the Lyrics of Five
Heavy Metal Subgenres

Jesse Taina
Pro gradu
English Philology
Department of Modern Languages
University of Helsinki
April 2014

Tiedekunta/Osasto Fakultet/Sektion – Faculty		Laitos/Institution– Department	
Humanistinen tiedekunta		Nykykielten laitos	
Tekijä/Författare – Author			
Jesse Taina			
Työn nimi / Arbetets titel – Title			
Keywords in Heavy Metal Lyrics: A Data-Driven Corpus Study into the Lyrics of Five Heavy Metal Subgenres			
Oppiaine /Läroämne – Subject			
Englantilainen filologia			
Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages	
Pro Gradu - tutkielma	04 / 2014	89 s. + 2 liitettä	
Tiivistelmä/Referat – Abstract			
<p>Heavy metal -musiikkia on akateemisessa maailmassa tutkittu vain vähän. Tutkimuksen kohteena on yleensä ollut heavy metal genrenä tai kulttuurisena ilmiönä. Ne harvat tutkimukset, jotka ovat keskittyneet heavy metal -kappaleiden sanoituksiin, ovat useimmiten ottaneet lähtökohdakseen kvalitatiivisen metodin. Tämä tutkimus hyödyntää kvantitatiivisia menetelmiä.</p> <p>Tutkimuksen lähtökohdaksi on tutkia heavy metal -kappaleiden sanoituksia tilastollisesti korpuslingvistisiä metodeita hyödyntäen. Menetelmistä tärkein on avainsanametodi, jonka avulla teksteistä voi tunnistaa tilastollisesti usein esiintyviä sanoja. Tutkimusta varten koottiin pienikokoinen korpus (<i>METAL corpus</i>), joka koostuu 200 tekstistä. Korpus on jaettu viiteen 40 tekstistä koostuvaan osakorpukseseen, joista kukin edustaa yhtä tutkittavaa alagenrea: black metal, death metal, power metal, thrash metal ja traditional heavy metal. Kunkin genren avainsanat saatiin vertaamalla tutkittavana olevaa kohdekorpuksista neljän muun osakorpuksen kaikkiin teksteihin, joita käytettiin referenssikorpuksena. Tällä tavoin esiin nousivat ne sanat, jotka esiintyivät kussakin tiettyä alagenrea edustavassa osakorpuksessa tilastollisesti useammin kuin muissa osakorpuksissa.</p> <p>Tutkimuksen tavoitteena on löytää vastaukset seuraaviin kysymyksiin: mitä avainsana-analyysi kertoo alagenreiden kappaleiden sanoitusten välisistä eroista? Kuinka hyvin avainsana-analyysiä voi hyödyntää kappaleiden sanojen tutkimisessa? Kuinka hyvin automatisoitu sanaluokkien annotointi (POS-tagging) toteutuu, kun aineistona ovat sanoitukset?</p> <p>Avainsana-analyysin tulokset kertovat kunkin alagenren sanoitusten tyylisiä sekä antavat hieman osviittaa tyypillisistä teemoista. Avainsanoiksi nousseiden kielipiirien sanojen perusteella alagenret black ja death metal suosivat kirjallisuutta ilmaisua, kun taas power, thrash ja traditional heavy metal suosivat puhekielisempää ilmaisua. Tulosten perusteella black ja death metal -kappaleiden sanoituksissa käytetään <i>of</i>-preposition sisältäviä nominilausekkeita tilastollisesti enemmän kuin muissa alagenreissa. Preposition <i>of</i> lisäksi kirjalliselle ilmaisulle tyypilliset sanat <i>as</i>, <i>before</i>, <i>nor</i> ja <i>the</i> nousivat myös avainsanoiksi black metalissa sekä <i>beyond</i>, <i>upon</i>, <i>thee</i>, <i>thou</i> ja <i>thy</i> death metalissa. Sitä vastoin puhekielille ominaiset lyhennetyt muodot nousivat avainsanoiksi power metalissa ja traditional heavy metalissa. Thrash metal -sanoitukset muista alagenreista erottavaksi tekijäksi havaittiin kirosanat, jotka nousivat kyseisen alagenren avainsanoiksi.</p> <p>Sanoitusten sisällöstä kielivien avainsanojen analyysin tulokset tukevat aikaisempaa heavy metal sanoituksista tehtyä tutkimusta. Sosiologi Kahn-Harrisin extreme metal -genrejen (tässä tutkimuksessa black, death ja thrash metal) sanoituksista tehdyt havainnot täsmäsivät tässä tutkimuksessa tehtyihin havaintoihin. Näiden genrejen edustamien kappaleiden sanoitusten sisältö ja teemat liittyvät usein vakaviin ja synkkiin aiheisiin. Myös näiden alagenreiden diskursseille Kahn-Harrisin mukaan tyypillinen provokatiivisuus ja normien rikkominen näkyvät tämän tutkimuksen tuloksissa. Power metal -sanoitusten sisällön yhteydet fantasiakirjallisuuden elementteihin kuten soturikulttuuriin ja sankaruuteen olivat nähtävissä alagenren avainsanoissa. Traditional heavy metal -sanoitukset poikkesivat muiden alagenreiden sanoituksista siinä, että niiden ilmaisu on puhekielisintä ja näistä alagenreista lähimpänä rock- ja muuta populaarimusiikkia.</p> <p>Sanaluokkien annotointi toteutui odotettua paremmin, sillä onnistumisprosentti (oikein annotoitujen sanojen osuus) oli korkea (94,7%) ja verrattavissa korkeimpiin annotoinnissa saatuihin tuloksiin (95-98%). Annotoitua korpuksista hyödyntämällä oli mahdollista tutkia kielipiirillisiä ilmiöitä, kuten genitiivisten nominilausekkeiden suhteellista määrää eri alakorpuksissa.</p> <p>Tutkimuksessa käytetty korpus on varsin pieni, mikä vähentää tilastollisten laskelmien merkittävyyttä. Saatuja tuloksia voidaankin pitää kohdistamina, jotka osoittavat mihin suuntaan heavy metal -sanoitusten tutkimusta voi jatkaa.</p>			
Avainsanat – Nyckelord – Keywords			
heavy metal, lyrics, keyword analysis, corpus linguistics, black metal, death metal, power metal, thrash metal, POS-tagging			
Säilytyspaikka – Förvaringställe – Where deposited			
Muita tietoja – Övriga uppgifter – Additional information			

Table of Contents

1 Introduction	4
2 Heavy Metal	5
2.1 Genre	6
2.2 A Short History of Heavy Metal.....	10
2.3 The Heavy Metal Subgenres in this Study	10
2.4 Lyrics in Heavy Metal.....	13
3 Corpus Linguistic Theory	21
3.1 Corpus Linguistic Approaches	21
3.2 Earlier Corpus Linguistic Studies on Song Lyrics	22
3.3 The Keyword Approach	25
3.3.1 Earlier Studies Utilizing the Keyword Approach	31
4 Data & Methods.....	32
4.1 Data Collection	32
4.2 Data Pre-processing.....	36
4.3 The METAL Corpus	37
4.4 Methods of Analysis	40
5 Analysis	43
5.1 Keywords of the Subcorpora	47
5.2 Keykeywords of the Subcorpora	62
5.3 Keywords of Individual Song Texts.....	72
5.4 Utilizing the POS-Tagger and the Tagged METAL Corpus.....	75
5.4.1 <i>Of</i>	78
5.4.2 <i>-ing</i>	80
6 Discussion	80
7 Conclusions	85

Works Cited 88

Appendix 1

Appendix 2

1 Introduction

Heavy metal music has been researched to some extent as a genre and cultural phenomenon from a sociological point of view (e.g. Weinstein 1991, Kahn-Harris 2007) but there are only few studies that focus on the linguistic aspects of heavy metal. This study takes a data-driven corpus linguistic approach to heavy metal lyrics, attempting to look into some linguistic factors that make heavy metal subgenres lyrically distinct from each other. Previous studies on rock and heavy metal lyrics are scarce and the ones that I have encountered have usually taken a qualitative approach (e.g. Saarinen 2013 and Träff 2011). The few studies that have taken a quantitative approach on song lyrics are only MA-thesis level or lower (Falk 2013) or are focused on other genres besides heavy metal (e.g. Katznelson & al. 2010).

This study employs corpus linguistic methods to find out what kind of differences there are between the lyrics of five heavy metal subgenres by examining which keywords (statistically significant words) characterize those different subgenres (the subgenres represented in this study being 'traditional heavy metal', 'thrash metal', 'death metal', 'black metal' and 'power metal'). I am hoping that a quantitative method will yield salient results that either support or challenge not only stereotypes on metal lyrics but also claims made about the content of the lyrics in earlier (qualitative) research (e.g. Weinstein 1991, Kahn-Harris 2007). A second objective is to test how well corpus linguistic methods, such as keyword analysis and part-of-speech tagging, can be applied to studying song lyrics.

The research questions of the study are the following:

1. What does a keyword analysis reveal about the differences between the lyrics of different heavy metal subgenres in terms of lyrical content and style?
2. How well can a part-of-speech tagging procedure be applied to a corpus consisting of song lyrics and can the tagging be useful for further analysis of the song lyrics?

Heavy metal as a musical genre and subculture and lyrics in heavy metal are discussed in chapter 2. Chapter 3 presents earlier corpus linguistic research and theory relevant to this study, focusing on the keyword approach. In chapter 4, the data and methods of this study are explained. The fifth chapter presents the results of study, while in chapter 6 those results are discussed. Finally, conclusions are presented in chapter 7.

2 Heavy Metal

This chapter discusses heavy metal as a cultural phenomenon and musical genre. The notion of genre is discussed in 2.1. A brief historical perspective is presented in 2.2, after which the heavy metal subgenres to be researched in this study are presented in 2.3. Finally, in 2.4, the role of lyrics in metal music and earlier research into heavy metal lyrics is discussed, in addition to which I will present arguments for a quantitative approach into studying lyrics.

Sociologists (Kahn-Harris 2007, Weinstein 1991) and musicologists (Fabbri 1982, Walser 1993) alike agree that defining a musical genre is no simple task. There is much more to a musical genre than merely the auditory aspect of it. Besides the actual sound of music (which includes e.g. the instruments used, chord progressions etc.), which naturally plays a major part in what makes a genre, there are many social and cultural aspects to take into account. Indeed, Deena Weinstein (1991:5-6) describes heavy metal a bricolage, a loose collection of diverse cultural elements, with “the sonic, the visual, and the verbal dimensions all mak[ing] crucial contributions to the definitions of the genre” (Weinstein 1991:7). Moreover, Robert Walser (1993:xviii) considers popular music (in which heavy metal is included) to be such a complex phenomenon that only an interdisciplinary inquiry that takes into account arguments of sociologists, musicologists, rock critics and cultural theorists is an adequate approach to studying it.

Weinstein’s 1991 book, *Heavy Metal: A Cultural Sociology*, which investigates all aspects of metal, such as the music, lyrics, artists, fans, dress code and behavioral codes and attitudes toward the subculture, is one of the major works that this study draws from. Another important work is musicologist Robert Walser’s *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music* (1993), which, similarly to Weinstein (1991), takes a holistic approach to studying heavy metal music, but focuses more on the formal aspects of the music than Weinstein, while adding the role of genre in the discussion. Keith Kahn-Harris’ sociological inquiry, *Extreme Metal: Music and Culture on the Edge* (2007), provides a more recent perspective to heavy metal music and to extreme metal in particular. Franco Fabbri’s musicological paper “A Theory Of Musical Genres. Two Applications” (1982), a solid work despite its age, is used as the basis for defining genre here. Some of the arguments by the above mentioned researchers have been taken into account in the methodology of the present study, insofar as they are related (see Chapter 3).

2.1 Genre

What makes a musical genre? Although there is a copious number of definitions for 'genre' (see e.g. Kahn-Harris 2007:10-11, Fabbri 1982), in this study, I will be focusing on one, following the ideas on genre that are sketched in Fabbri (1982:1): "a musical genre is 'a set of musical events (real or possible) whose course is governed by a definite set of socially accepted rules'¹". It is the musical community that he considers to form the basis for the definition of a genre, with the notion of code playing an important part (The concept of code is also discussed in Kahn-Harris 2007 and throughout Weinstein 1991). According to Fabbri (1982), there are five types of rules that are codified in a genre, summarized below:

- I. Formal and technical rules, which play a major role in every genre, with all genres having their typical forms. According to Fabbri, it is these rules that musicologists prefer to study at the expense of others.² He points out, however, that "a record buying adolescent of today has clearer ideas on musical genres than the majority of musicologists who have made such a fuss about them." (1982:3)
- II. Semiotic rules, pertaining to codes that "create a relation between the expression of a musical event and its content" (1982:4). Fabbri concedes that all of the five rules can basically be seen as semiotic rules but he uses the term particularly to refer to communicative functions such as referential, emotional and metalinguistic functions. Their relative importance varies, as the rules of a genre determine which functions dominate the others. In addition to the verbal element, semiotic rules include other elements such as the distance between musicians and audience, the postures of the musicians on stage, and gestural-mimical codes.
- III. Behavioral rules, such as rules of conversation; which personality traits are valued within a genre; and rituals associated with being part of a genre.
- IV. Social and ideological rules, which are related to a genre's internal social structure and social functions.
- V. Economical and juridical rules.

¹ The quotation marks are present in the original text but they are not a citation. Fabbri uses them to delineate the definition of a musical genre.

² Although a sociologist, Weinstein (1991:6), too, sees that to be included in a genre, at its minimum, music has to meet certain sonic requirements, Weinstein's 'sonic' being practically synonymous with Fabbri's 'formal' in this case.

The scope of the present study excludes studying phenomena directly related to most of these rules. As the perspective here is linguistic and the study material a collection of heavy metal song lyric texts, out of Fabbri's five types it is the realm of semiotic rules that pertains to the topic matter. It should be kept in mind, however, that song lyrics are merely a small part of the semiotics of heavy metal. As Walser points out, the meaning of a song's lyrics does not equal to the meaning of the song, verbal meanings constituting only a small part of "whatever it is that makes musicians and fans respond to and care about popular music" (Walser 1993:26).

When it comes to meaning in heavy metal music in general, both Weinstein (e.g. 1991:34-35) and Walser (e.g. 1993:2-3) emphasize the role of power: the sonic power of the music itself, the powerful voice of the singer, the loudness of the music, and the formal choice of using power-chords frequently. Moreover, lyrics are often not seen as the most important thing to listen to by metal fans. In fact, Walser claims that "musicians and fans alike tend to respond primarily and most strongly to musical meanings" (1993:39) and backs this up with the results of his interviews with both musicians and fans: 48 percent of metal fans reported that they pay attention mostly to the music; 41 percent to both music and lyrics; and only 11 percent primarily to the lyrics (1993:187). Somewhat similarly, according to Weinstein, fans and critics alike "would agree that in heavy metal the lyrics are less relevant as words than as sound" (1991:25–26). Instead, she considers vocal power (and the power of the music) to be more important to the meaning of a song than "any delineated meaning presented in the lyrics." (1991:34). What I want to underline here is that I am studying a textual aspect of song lyrics, that is, keywords of song texts, and even though I am going to make interpretations on what the lyrical contents or themes be in particular subgenres, it is not possible to make claims on the full meaning of the songs – let alone heavy metal as a genre – by studying the lyrics only.

What is it that makes a heavy metal *subgenre* then? Based on my experiences, such as talking with metal fans, reading magazines, online articles and comments on online forums, it could be said that there is actually no consensus on exact definitions of heavy metal subgenres. For instance, a typical argument between metal fans – a kind of which I have witnessed and even participated in myself – is one that has to do with whether band A represents metal subgenre X or subgenre Y. Moreover, there is no unanimity in the metal scene³ on which band is considered metal and which is not. As Weinstein points out, "[c]onstant concern with drawing fine distinctions between who is

³ See Kahn-Harris 2007:13-15 on discussion about the concept of 'scene'.

within and who is without the pale of heavy metal is a staple of discourse among heavy metal fans" (1991:15). According to Walser, 'heavy metal' is a term that is constantly under debate while "genre boundaries are not solid or clear" but are instead "conceptual sites of struggles over the meanings and prestige of social signs" (1993:4). An example of a group that lies in the border of metal and rock is KISS. Depending on whom you ask people might characterize the group alternatively as heavy metal or hard rock (or glam rock/metal). Moreover, since metal has evolved to more extreme and diverse directions (both formally and semiotically) since its appearance, what was initially considered heavy metal might not seem so 'heavy' anymore to, say, listeners who prefer modern extreme metal with blast-beats⁴ and heavily distorted guitars. Now, with regard to KISS, Weinstein writes that KISS "were too melodic to be considered heavy metal in the 1970s" (1991:20). Curiously enough, the website *Metal Storm* characterizes their 1970s albums as hard rock, 1980s albums as hard rock/glam metal and 1990s albums as hard rock/heavy metal (the issues with *Metal Storm's* categorization are discussed in section 4.1). This raises the question whether these differences in categorization are due to changes in the band's style during their career or due to a change in overall perspective, that is, in what counts as metal and what does not, or whether it is simply a result of the myriad and sometimes rather arbitrary ways in which genre labels are assigned by different people (mostly journalists, sometimes musicians themselves). Or is this merely due to the dynamic nature of genres, as genres are "constantly developing and reconfiguring" (Kahn-Harris 2007:7) and "can never be static" (Kahn-Harris 2007:12)?

However it may seem genre labels are not random. It is rather so that genre definitions can be rather vague (and subjective) and that (heavy metal) music varies so much depending, on one hand, on the band, album and song which is about to be categorized, and on the other hand, on the authority behind assigning the label. That is, pigeonholing a piece of metal music into one strict category is surely not easy nor necessarily even feasible. According to Weinstein, the code that allows the classification of "a song, an album, a band or a performance" is "not systematic but it is sufficiently coherent to demarcate a core of music that is undeniably heavy metal" (1991:6). Moreover, Fabbri (1982:3) points out that of the five types of rule he establishes some are more essential than others. He adds that it is not atypical that only a few of the rules are very important in a genre while the rest of them are of lesser importance.

⁴ An intense, quick and aggressive drum-beat typically based on the rapid alternations of the bass drum and snare.

With regard to the variance of the importance of generic rules, let us consider 'war metal', which is a derivative genre of black metal. Derivative genre is a term used here to refer to further developments of subgenres, which gives us the hierarchy genre > subgenre > derivative genre. If I take the role of the authority behind assigning a genre label I could logically say that if a band does not adhere to a textual semiotic rule of having (most of their) song lyrics about war and related themes (an invented rule), the band would not be considered war metal. However, as discussed above, lyrics are not necessarily considered that important a factor for listeners, nor is their topic matter necessarily strictly dictated by textual semiotic rules. Indeed, Kim Kelly, a contributor to the blog on heavy metal music called *Invisible Oranges*, who discusses the genre of war metal in a blog post, seems to consider formal rules and semiotic rules having to do with image as more important than the rules pertaining to the topic matter of the lyrics, when it comes to defining the genre. She characterizes war metal as "a blasphemous, violent black/death metal hybrid so extremely fast, raw, and chaotic that it often borders upon grind – 'black/grind', even" (Kelly 2014). She considers esthetics to play an equally important role in the definition of the genre as the music, but she does not discuss the role of lyrics, not directly anyhow, which suggests to me that she considers textual semiotic rules as less important than formal rules and image-related semiotic rules. Also, the difficulty in establishing the bases on which a genre label is assigned is obviously present in Kelly's blog post, which can be seen in the introduction: "keeping the inherent absurdity of the concept of a 'microgenre' firmly in mind, allow me to draw your attention to one of the most Lilliputian semi-genres of all, both so ultra-specific in sound and with borders so hazily blurred that its very definition is a battleground" (Kelly 2014).

Based on my experience of the heavy metal scene, my own impression, which is rather similar to Weinstein's, is that subgenre categorizations in metal music are mainly based on musical features, that is, the formal and technical rules of the genre, more than, say, lyrical themes. However, in certain subgenres other genre rules are highlighted. For instance, in the definition of Christian metal, black metal and especially unblack metal, textual semiotic rules often play an imperative role. Unblack metal is a subgenre which is musically black metal but lyrically and ideologically completely opposite in that unblack metal promotes Christianity – similarly to Christian metal – whereas black metal lyrics are usually "openly Satanic" (e.g. Kahn-Harris 2007:37-38) and misanthropic. To summarize, there are no exact, unanimously agreed definitions for the genre of

heavy metal or its subgenres, but each subgenre has a sufficient amount of elements and codified rules that are adhered to for a certain core of music to be recognized as part of each subgenre.

2.2 A Short History of Heavy Metal

A very brief history of heavy metal from its birth to present is now presented, to give the study and its data a historical framework. There is no exact date when heavy metal as a musical genre was born but scholars (Kahn-Harris 2007:2, Walser 1993:3, Weinstein 1991:11) date its emergence in the late 1960s and/or early 1970s. Weinstein notes that there is no consensus on who the first metal bands or their influences were or which tracks were the first “full-fledged [heavy metal] songs” (14). Many fans seem to regard Black Sabbath as the first heavy metal band, although some consider Led Zeppelin and Deep Purple as equally important. At the time of its emergence, heavy metal was not yet considered a genre of its own, and the term itself was coined only until some years later. Indeed, at first, in the late 1960s, heavy metal was still considered a “‘harder’ sort of hard rock” (Walser 1993:3).

Heavy metal’s fragmentation into subgenres is dated in the early 1980s (Weinstein 1991:21, Kahn-Harris 2007:2). Weinstein divides heavy metal into two main subgenres, namely ‘speed/thrash metal’ and ‘lite metal’ (or ‘melodic metal’ or ‘pop-metal’) but that division is too simplified, or rather just obsolete, as her book is already over 20 years old and since then, countless new varieties of metal have emerged. According to the Wikipedia article *Heavy Metal Subgenres* (2014), at the time of writing this there were 26 heavy metal subgenres, several with two or more derivative genres, such as the subgenre doom metal with its derivatives ‘death/doom’, ‘drone’ and ‘sludge metal’. At the present day, new metal subgenres emerge rapidly, as many metal musicians create fusion genres that combine heavy metal esthetics with other musical styles. Indeed, the kind of traditional heavy metal popular in the 1970s is a relatively rare style to be played by contemporary metal bands.

2.3 The Heavy Metal Subgenres in this Study

Five heavy metal subgenres are represented in this study (the method for the selection of those genres is discussed in section 4.1): ‘traditional’ or ‘classic heavy metal’, ‘thrash metal’, ‘power metal’, ‘death metal’ and ‘black metal’. All five are established subgenres that have existed each at least over 20 years (see e.g. Kahn-Harris 2007:2-5;22). Out of the five, thrash metal, death metal and black metal can be collectively referred to as ‘extreme metal’, although other extreme metal

subgenres exist as well. 'Heavy metal' and 'metal' are used here to refer to all metal music as a whole. I will now briefly describe the subgenres in this study focusing mostly on the musical elements. The characterizations are, to a large extent, based on my own listening experiences and involvement in the scene for over 10 years and, to a lesser extent, on earlier research.

Traditional or classic heavy metal (from now on referred to as traditional heavy) can refer either to heavy metal music prior to its fragmentation into subgenres in the 1980s and/or to newer music that is similar to the heavy metal music of that initial period. It is a useful term, since 'heavy metal' can be understood as an umbrella term to include all subgenres of metal. Metal bands of all subgenres typically employ drums, bass guitar and at least one guitar, although usually two. Traditional heavy music is characterized by loudness, power, distorted guitars and usually high-pitched vocals sung at high volumes (e.g. Weinstein 1991: *passim*). Most of these characteristics, except for the type of vocals, apply to the other subgenres as well. The sound of traditional heavy is, however, more reminiscent of hard rock, rock 'n' roll and blues than the sounds of many of the other subgenres, which may have been stripped of these elements altogether. Examples of traditional heavy bands are Black Sabbath, Judas Priest and Diamond Head.

Thrash metal is an extreme metal genre influenced by (hardcore) punk music and New Wave of British Heavy Metal (a traditional heavy subgenre) and a band called Venom, and is typically fast-paced, aggressive and serious (Kahn-Harris, 2007:3). It tends to be less melodic in its instrumentations and vocal performances than traditional heavy. In fact, lyrics are often shouted or yelled, typically with a raspy, overdriven tone of voice. Weinstein considers thrash to be heavy metal's fundamentalist strain, where tempo and power is emphasized, while arty and fantastic elements as well as fancy stage wear is minimized (1991:48–50). In comparison to traditional heavy, the general feel of songs tends to be more serious and urgent. Notable thrash metal bands include Slayer, Testament and (early) Metallica. Speed metal, which can be musically quite similar to thrash but emphasizes speed and is not necessarily quite as aggressive, is sometimes included under the label of thrash metal and sometimes considered a separate subgenre. However, both Weinstein (1991) and Kahn-Harris (2007) use these labels synonymously.

Power metal is characterized by fast tempos, catchy melodies and often easy-to-approach chord progressions. Its overall sound is rather energetic and uplifting than aggressive – as opposed to extreme metal music – with vocals usually sung by a tenor with a clean tone of voice and a bright

timbre. It is not uncommon for a synthesizer playing a string ensemble or brass section sound to be included along with the more traditional band instruments. There tends to be a heroic and fantastic quality to the music – even a lightness that is often absent from other types of metal (apart from some traditional heavy music). Kahn-Harris considers power metal to be a hybrid of traditional heavy and thrash metal and situates power metal at the margins of the extreme metal scene (2007:22), although I find the subgenre musically too approachable to consider it very extreme, added that extreme metal fans often shun power metal according to my experience. Well-known power metal bands include Stratovarius, Helloween and Hammerfall. It should be mentioned that (partly due to history of local metal scenes) ‘power metal’ is sometimes used to refer to bands that I would characterize as thrash or speed metal (e.g. Weinstein (1991) considers ‘power metal’ a synonym for ‘thrash’).

Death metal is a subgenre one notch more extreme than thrash metal, from which it developed. The songwriting tends to be more complex, the guitarwork more technical and the sound more austere (Kahn-Harris 2007:3). Austerity is usually emphasized with low-tuned guitars (Weinstein 1991:51). Tempos vary from song to song and band to band but in general death metal songs are fast-paced, intense and heavy. According to my experiences, death metal music is often shunned by others than metal fans, for instance, because it is perceived to sound like noise or it is felt to evoke feelings of anxiety (this also applies to black metal). The vocals are growled or snarled rather than sung (Weinstein 1991:51). Examples of death metal bands are Cannibal Corpse, Morbid Angel and Vader.

Black metal is perhaps the most notorious of all heavy metal genres, with its overt connections to Satanism combined with an Anti-Christian sentiment, and the tendency of many groups to openly demonstrate their individualistic and even misanthropic (also Kahn-Harris 2007:40;42) views, although only a fraction of the groups involved in the scene have committed the kind of crimes the scene is known for (murders and church burnings). Typically the label refers to a style of black metal that developed in Norway in the early 1990 “characterized by screamed, high-pitched vocals, extremely rapid tempos, ‘tremolo’ riffs, a ‘trebly’ guitar sound, and simple production values.” (Kahn-Harris 2007:4). The atmosphere of black metal songs is often dark and otherworldly. Similarly to death metal, the overwhelming blast-beat is frequently used. Black metal is another extreme metal genre that, along with thrash and death metal, is markedly different from traditional heavy in that it is distinguished by certain musical radicalism and

transgressive behavior (discussed below in 2.4), being the most radical of the extreme metal genres in these two aspects (Kahn-Harris 2007:29-43). Black metal imagery draws from a Nordic pagan past and an ideology of romanticism (Kahn-Harris:41), and indeed, images of nature, especially the forest, are often found in black metal lyrics.

For the sake of simplicity, genre classifications subsume derivative genres in this study when it comes to discussing bands. That is, derivatives of traditional heavy, such as 'New Wave Of British Heavy Metal', for instance, will simply be called traditional heavy, derivatives of black metal, such as 'depressive black metal', will simply be called black metal and so on. This is due to the fact that the majority of the albums from which the songs in this study are selected are classified by *Metal Storm*⁵ additionally or alternatively as something else than a main subgenre, such as death metal. For instance, *Crimson* by Edge of Sanity is labeled there 'progressive death' and *Ashes Against The Grain* by Agalloch as both 'atmospheric black' and 'neofolk'.

2.4 Lyrics in Heavy Metal

People surely have all kinds of ideas what heavy metal lyrics are about but many of those ideas are based on misinformation, stereotypes and prejudice. Both Weinstein (1991) and Walser (1993) quote Jon Stuessy, a professor of music, who writes in his 1985 work that heavy metal's basic themes are "...extreme rebellion, extreme violence, substance abuse, sexual promiscuity/perversion (including homosexuality, bisexuality, sadomasochism, necrophilia etc.), Satanism" (Walser 1993:139). On the contrary, Walser considers these topics to be uncommon in heavy metal lyrics, a statement which he backs up with the results of his own study (discussed in more detail later in this section). Walser is very critical of the earlier research on heavy metal, and to a large extent rightly so, since, as he points out, much of it is subjective, poorly argued or based on a narrow point of view. In fact, he goes so far as to write that "...rock journalists have published relatively few insights about heavy metal. Academics have achieved much less." (1993:21). Indeed, he considers Weinstein's 1991 book, to be a subjective, "impassioned defense of heavy metal" (Walser 1991:23) and argues that Weinstein focuses too much on the lyrics and that she does not have "anything useful to say about the music of heavy metal" (1991:23). In this I disagree with Walser, since in my opinion Weinstein makes many useful observations on heavy metal as a phenomenon and even if she characterizes the music itself with mostly non-musicological terms, her observations are hardly useless. I do agree with Walser that Weinstein

⁵ The website used for choosing the albums in the study (discussed in 4.1),

does not always seem that objective in that some of her arguments are questionable and her writing style reveals an emotional affinity for the heavy metal scene.

According to Weinstein (1991), Walser (1993) and Kahn-Harris (2007), there are certain themes that are typical for heavy metal in general and other themes that are typical for certain subgenres in particular. First, I will discuss Weinstein's (1991) viewpoint on heavy metal lyrics and point out some advantages and disadvantages of her study. Weinstein reports that her analysis of lyrical themes is based on her having listened to approximately 4,000 songs; discussing lyrics with heavy metal fans, artists, and media personnel; and reading "about 150 heavy metal magazines and all the academic literature on the genre" (1991:34).

According to her work (1991), there are two major lyrical themes in heavy metal, Dionysian and Chaotic. She considers this thematic division a binary opposition. However, I feel that this 'opposition' is a rather strange one, since the traditional literary dichotomy based on Greek mythology is Dionysian and Apollonian, while chaos is an aspect of Dionysus⁶ (Thro 1996). Nevertheless – and although Weinstein somewhat contradictorily to her own statement sees Dionysus and Chaos joined together (rather than opposing each other as they were supposed to be binary opposites) in a fight against the respectable society – there is a certain rationale in dividing the themes into these two categories, at least in the way that Weinstein does it, for they do seem to form useful thematic categories. It should be noted that Weinstein's statements regarding these categories refer to heavy metal lyrics as a whole, that is, they apply to heavy metal in general, although she sees that lite metal (or glam/pop/melodic metal) is mostly concerned with Dionysian and thrash/speed metal with Chaotic imagery.

In the Dionysian category, an important subtheme is "love in its earthy sense of lust and sex" (Weinstein 1991:36). Weinstein sees that heavy metal lyrics emphasize lust and the carnality of sex rather than any spiritual element. The lyrics involve depictions of women as sexual creatures while sex is seen as something fun and without commitments. Weinstein also states that heavy metal took many lyrical influences from blues music and rock and roll, some of which, such as 'rock and roll' itself, were code words for sex, sex being a key topic in the genres mentioned (1991:36). She asserts that 'dance', which she considers the main code for sex in rock and roll, is not used in heavy metal (1991:36). I have to disagree with her because a counter-example comes

⁶ Dionysus is the Greek god of wine, who "induces madness, passion, and irrational frenzy" (Rieger 1994:2). Apollo is the god of sun and is associated with reason and order among other things.

from the Judeo-Christian tradition (e.g. The Book of Revelations, references to the devil and hell), paganism and secular entertainment (such as horror literature and horror movies). According to Weinstein, songwriters vary in their attitude towards images of Chaos: often the lyrics are just descriptive and the stance unclear. She argues that some songs react to Chaotic imagery with responses that can be “earnest pleas directed against the forces of destruction” or “expressions of playful delight in the imagery” (1991:41).

Studying attitudes towards themes in lyrics is beyond the scope of this study because to do that properly one should read the song texts more closely. Chaos-related vocabulary, on the other hand, is a possible topic of scrutiny, and indeed, some Chaotic vocabulary will be discussed in the analysis of the present study. In all, Weinstein’s division of heavy metal lyrics into these two thematic categories is quite well-founded, as they can be seen as two opposite ends of the pole, or as ‘nice’ and ‘nasty’, if you will, since Dionysian emphasizes the pleasures and hedonistic aspects of life, whereas Chaos focuses on the sinister side. In light of the traditional literary dichotomy, however, Weinstein’s both major lyrical themes fall into the Dionysian rather than the Apollonian category.

Remarking on metal lyrics in general, Weinstein states that optimism and hope are absent in them, especially hope to be able to change the world. She also claims that there is “nothing lighthearted” in heavy metal lyrics (1991:35). However, I can again easily think of examples that do not support her arguments. Black Sabbath have a song called *Children Of The Grave* (example 2a), released in 1971, that encourages one to take action for a better world and exhorts to love rather than to hate. In general, Black Sabbath’s lyrics do have a somber tone, so this song is an exception amongst their work. The progressive metal band Persefone also have a concept album called *Spiritual Migration* (2013), which is ideologically Buddhist to a large extent and contains several parts in its lyrics that encourage one to love and accept oneself⁷ as seen in example 2b below.

2a) They're⁸ tired of being pushed around
 And told just what to do
 They'll fight the world until they've won

⁷ With regard to the topic matter of their lyrics, it is interesting that Persefone are categorized as (progressive/melodic) *death* metal by *Metal Storm* and *Metal Archives*. This I see as another example that supports the idea that metal subgenre divisions are based on musical features rather than lyrical features.

⁸ ‘they’ refers to the ‘Children of the Grave’ here.

And love comes flowing through
 (Black Sabbath – Children Of The Grave)

2b) We are sons of a great and deep source
 Transmitting unconditional love
 Its power flows to each one of us
 Eternally sustaining ourselves
 (Persefone – Mind As Universe)

With regard to the asserted seriousness of the genre, it has to be mentioned that there are several bands that intentionally combine heavy metal and (to a varying extent) humor in their lyrics, such as Tenacious D, Steel Panther, Crotchduster and Cradle Of Filth. These four groups have all surfaced only after Weinstein's 1991 book was published, however, so perhaps metal music featuring lyrics with humor was less common at the time of her study. Nevertheless, I believe that the above examples, which run counter to Weinstein's statements, illustrate that when the evidence is based on such a subjective method as one's own listening experience, the results may be quite variant, adding to that that neither Weinstein nor I supported the observations discussed above with any quantitative data. This is what I think makes a quantitative approach to studying lyrics, such as the one in the present study, interesting and hopefully illuminating, although it is surely not without its limitations too. As Weinstein points out, heavy metal lyrics should be understood as loosely arranged fragmentary signifiers that should not be interpreted literally but rather more holistically with contextual and figurative aspects in mind, these two aspects unfortunately lying outside the scope of this study to a large extent.

Kahn-Harris (2007) touches the theme of lyrics in extreme metal from the perspective of discursive transgression. Transgression is a key concept for sociologists and can be defined as "conduct that breaks rules or exceeds boundaries" (Jenks 2006: back cover). Kahn-Harris' (2007) approach to studying extreme metal is a holistic one, with his own research based on immersing into the scene as an ethnographer (which includes conducting lots of interviews), collecting fanzines and CDs, as well as he himself writing for an extreme metal magazine called *Terrorizer*. Kahn-Harris considers the different types of transgressive behavior – those related to songwriting, lyrics, musical choices and stage performance – to be the factor that makes extreme metal extreme. An example that he gives are the lyrics (not to mention the title of the track!) of the Cannibal Corpse song 'Fucked with

a Knife', which he rightly points out to be such an explicit description of sexual violence that one scarcely ever finds suchlike in popular music or even in heavy metal. Kahn-Harris regards explicit violence a key aspect of extreme metal discourse in particular. Comparing extreme metal lyrics to traditional heavy texts, he claims that although dark themes may be discussed in traditional heavy metal, the texts tend to be more theatrical and fantastic while the attitude towards the topic matter can be ironical or satirical, as opposed to extreme metal whose lyrics are more sinister, vivid, detailed, unambiguous and serious. For instance, he points out that while the lyrics of the traditional heavy song 'Suicide Solution' by Ozzy Osbourne can be interpreted as contemplative and ironical (as Walser 1993:147-149 does) regarding its attitude towards taking one's own life, the lyrics of the death metal song 'Sacrificial Suicide' by Deicide "are explicitly Satanic and explicitly glorify suicide" (Kahn-Harris 2007:35).

While Kahn-Harris (2007) does not thoroughly discuss the lyrics of particular extreme metal genres he does mention that doom metal songs contain melancholic lyrics; black metal songs feature Satanistic and occult themes and death metal tracks feature (although not exclusively) themes of violence, war, the occult, death, killing and mutilation. These themes pertain to 'abjection' which Kahn-Harris considers a key theme in extreme metal lyrics (although my impression is that they are most common in death metal in particular). The abject he defines as "that which is formless, disgusting, terrifying and threatening" (2007:29) and is "associated with 'vile' bodily fluids, but may be displaced elsewhere: to women, to Jews, to animals etc. The abject has to be removed from orderly society and/or destroyed" (ibid.). It is commonly something that has to be controlled or subjugated. Kahn-Harris considers these kinds of fantasies of control a crucial trait of extreme metal discourse. If Stuessy's claim that most metal lyrics are about extreme violence and perversion etc. is reconsidered, we could say that based on Weinstein (1991), Walser (1993) and Kahn-Harris (2007), it is extreme metal lyrics that are associated with these themes of abjection mentioned by Stuessy, not heavy metal in general.

Of the literature on heavy metal lyrics that I have encountered, the only work that has any numbers to back up the claims made by the author is that by Walser (1993). He conducted an analysis of the topics of 88 heavy metal song lyrics, dividing the 88 songs into 6 categories and counting the instances of songs occurring in each category:

1. Assertion of or longing for intensity: 27
 2. Lust: 17
 3. Loneliness, victimization, self-pity: 17
 4. Love 14 (affirmation 8; regret or longing, 6)
 5. Anger, rebellion, madness: 8
 6. Didactic or critical (antidrug, anti-Devil, anti-TV evangelism, critique of the subversion of justice by wealth): 5
- (adapted from Walser 1993:139)

Walser's categorization, which he himself considers a rather crude one, seems not to contradict with that of Weinstein's but is arguably rather along the same lines, since topics 1, 2 and 4 can be classified as Dionysian, while 3, 5 and, to some extent, 6 pertain to Chaotic themes. However, Walser's (and Weinstein's) results are over 20 years old. In addition, Walser's sample is rather small and only focuses on the song lyrics of more mainstream and commercially successful heavy metal bands (that is, music that would mostly fall under the category of traditional heavy). Therefore it is not that surprising that themes pertaining to abjection are absent from Walser's list, since most extreme metal genres have properly solidified only after his work was published.

If we consider another rather typical theme for metal song lyrics, death penalty, a brief lyrical/musical analysis of a handful of songs suggests a somewhat similar difference in style, as Kahn-Harris' examples, that is, extreme metal lyrics being more sinister and traditional heavy lyrics more theatrical. The lyrics of the thrash metal track 'Death Penalty' by Fog of War describe the fear of approaching death, some details of the actual execution and the anguish that goes with it. There is also a political aspect to the song, as the song title refers to the type of sentence given and the policy is commented in two lines: "The prisons are crowded, not an empty cell left / No taxpayer dollars, you have been sentenced to death". The song is rather austere in its musical and lyrical expression, and is devoid of hopefulness, finishing with the grim matter-of-factly statement: "Death is affirmed, the coroner confirms / Buried in the mud, you're left to the worms". An even grimmer atmosphere is found in the death metal song 'Death Row (No Regrets)', an oppressive, heavy, slowly paced piece of music by Hypocrisy. The narrator of the song, empowered by hate and laughing at his fate, is not afraid nor does he regret his deeds, but instead seems to find comfort in that "all the pain must cease" when his life ends:

4) I am no savior and I ain't no saint
They will take me to the electric chair
Yeah, ha ha ha

No remorse and no regrets
for those I've sent straight to hell
(Hypocrisy – Death Row (No Regrets))

Now, let us compare these two extreme metal songs to traditional heavy tracks. 'Death Row' by Judas Priest, is musically more groovy and less aggressive than the songs by Hypocrisy and Fog of War, and has some spoken parts in it that can be interpreted as less serious or ironical, based on the tone of the singer's voice and the lyrical content: "Death row - What's that - 'No stay?' / Oh! Ok I'll tell him". The narrator does not admit to doing anything wrong, conceding to having done only one mistake that "ain't that bad". There is a sentiment of resistance in the song as the chorus goes:

5a) Oh! No! - I won't go
You'll never get me down to
Death row
Oh! No! - They're taking me slow
One way ticket to death row
(Judas Priest – Death Row)

The lyrics also appear open for an interpretation where they are seen as criticizing society as an oppressive force:

5b) We've all been detained
For the crimes we've committed
They stayed execution
But we'll never get acquitted
Waiting for the day when they strap us in
(Judas Priest – Death Row)

In that sense, the narrator's circumstances do not seem quite as helpless and gloomy as they do in the two extreme metal songs. A more hopeful sentiment is found in the traditional heavy song

'Hallowed Be Thy Name' by Iron Maiden. It is musically quite catchy and not very gloomy – at least not in comparison to the other songs analyzed here. The protagonist of the song contemplates upon his past and on the remainder of his life, finding it hard to believe that he is truly going to die. He does not want to die but tries to take comfort in his wavering belief that there will be life after death, although he expresses no regret: "As I walk all my life drifts before me / And though the end is near I'm not sorry / Catch my soul 'cause it's willing to fly away". Of the songs here, 'Hallowed Be Thy Name' is the only one with overt references to Christianity, as the song title is borrowed from the Lord's Prayer and as 'soul' and its prospective existence after death is mentioned twice in the lyrics. By the end of the song, the protagonist seems to find solace in his belief in afterlife. These observations support those of Kahn-Harris in that the way things are described in the lyrics is gloomier in extreme metal than in traditional heavy.

I must point out that it is not difficult to find instances like the ones discussed above when one knows what to look for (and if one already has preconceptions) and that a few selected examples hardly prove anything about the general patterns in metal lyrics. This is why I think a quantitative study into lyrics is justified and needed. Of course, a quantitative method alone is never sufficient in the study of language. Therefore the points made about a more holistic approach to studying music and lyrics should always be taken into account

3 Corpus Linguistic Theory

This chapter discusses corpus linguistic approaches and some studies utilizing corpus linguistic methods. Typical approaches to corpus linguistics are discussed in 3.1, some earlier research on song lyrics utilizing corpora is presented in 3.2 and the key word approach, which is the main method applied in this study, is then discussed in 3.3.

3.1 Corpus Linguistic Approaches

This study takes a data-driven approach to corpus linguistics, as opposed to a corpus-based or corpus-driven one. That is, the particular linguistic features to be examined will be based on information extracted from the data itself (Rayson 2008), in my case, the data from the METAL (Metal Lyrics) Corpus compiled for this study (the corpus and its compilation are discussed in 4). Rayson (2008) considers the data-driven approach (which he calls Type III) to combine both the corpus-based (Type I) and the corpus-driven (Type II) methodologies. These latter two approaches are also known as microscopic and macroscopic respectively, with Type I typically focusing on

studying particular linguistic features such as certain words or lemmas while Type II tends to examine how a particular feature or a set of features characterize whole texts or varieties of language (such as registers) (Rayson 2008). According to Storjohann (2005), the corpus-based approach takes the corpus to be a tool that can be used to support theories and intuitions while in the corpus-driven approach corpus data is studied without prior intuitions, with the corpus serving as an empirical basis for detecting linguistic phenomena.

The data-driven or Type III approach merges the corpus-based and corpus driven ones in the way that first, whole texts are examined and data is retrieved (c.f. corpus-driven approach), and then, based on that data, the particular linguistic features (c.f. corpus-based approach) that are to be studied further are chosen (Rayson 2008). In this study, the melding of these two approaches is seen in that statistical calculations using WordSmith Tools (described in 4.4) provide lists of keywords from the data after which the keywords are then studied in detail to find out how they characterize certain varieties of language, that is, heavy metal subgenres.

Examples of corpus-based studies are Simpson & Mendis (2003), who investigate which idioms occur in academic spoken language and in what kind of contexts they occur, and Xiao and McEnery (2004), who investigate the use of aspect in Mandarin Chinese using corpora to develop a corpus-based model of aspect in that language. So, in these both studies, the linguistic features to be studied had been chosen prior to the research. On the other hand, Granger & Paquot's (2010) study on academic vocabulary is a corpus-driven one, investigating which (groups of) lexical verbs characterize academic literature on business, linguistics and medicine. The work by Van Bael & al. (2004) has a data-driven approach whose method employs spoken language corpora to investigate speech style variation with the goal to identify phonetic regularities in the different styles.

3.2 Earlier Corpus Linguistic Studies on Song Lyrics

Corpus research on music, let alone heavy metal music, is scarce to non-existent. Schneider & Miethaner (2006) used a corpus of blues lyrics (BLUR) to study verb complementation patterns in Earlier African American English, such as "a syntactic structure in which verbs like *start* are followed by *to* and a verbal *-ing* form" (2006:233-234). They successfully gained some insight on the syntax of the dialect, concluding that blues lyrics can be a "valuable source of linguistic data" (2006:251). However, Schneider & Miethaner did not exactly investigate blues lyrics from a genre point of view, but rather as a source of linguistic information on a dialect.

Logan & al. (2004) performed a semantic analysis of a set of lyrics for 15,589 songs by 399 artists using corpora and a statistical method called Probabilistic Latent Semantic Analysis, an automated text analysis technique to study semantic similarity. Only quantitative methods were used. In their results they present the most frequent words (excluding stopwords⁹) in the subcorpora they consisting of lyrics collected from online, representing the genres country, newage, rap, reggae and rock (Table 1). Logan & al.'s research question is completely opposite to the one in this study, as they investigated lyrical and musical similarity, not what makes each genre individual or which words are key. In their paper no interpretation on the lyrics is made whatsoever. Nor do they present any statistical information on the corpus apart from reporting the number of songs and artist.

REGGAE	COUNTRY	NEWAGE	RAP	ROCK
GIRL	LOVE	ADIS	I'M	I'M
LOVER	I'M	GO	LIKE	LOVE
KNOW	JUST	SAY	GET	DON'T
LOVE	DON'T	DAY	GOT	KNOW
I'M	KNOW	NIGHT	DON'T	JUST
LET'S	LIKE	LOVE	N****	LIKE
MI	GOT	SKY	KNOW	GOT
SHOUT	TIME	SAYS	S***	YOU'RE
LIKE	HEART	ERGO	AIN'T	TIME
GAL	GO	HEART	YO	OH

Table 1 "The ten most frequent non-stop words for selected genres. Obscene words have been obscured." Adapted from Logan & al. (2004:2).

Katznelson & al. (2010) investigated frequent words in American song lyrics in their corpus-based research project. Their data consisted of the lyrics of 433 songs (178,982 words), representing four different genres and their top chart hits from every second year, spanning from 1989 to 2009.

According to Katznelson & al., frequency lists can be investigated in order to comment on stereotypes connected to a genre, and also to study cultural ideations (2010:2;14), the meaning of which is unclear as the concept not defined in their paper. They extracted the ten most common nouns (Table 2) from their subcorpora in the study (representing rock, pop, country and hip hop)

⁹ Stopwords are words that are filtered out of lists but Logan & al. do not specify which stop words are used, but merely report using a "standard list of stopwords" (Logan & al. 2004:2). Usually stopword lists consist of function words such as determiners, prepositions and pronouns. Indeed, no function words are visible in their list of frequent words for the five genres.

and calculated a normed frequency for all the nouns. The normed frequency was calculated by “multiplying the hits of a given word in a genre by the average words per song and then dividing the total by the number of words in the genre (e.g. ‘love’ has 93 hits in Rock: $93 \times 254 / 27,898 = .85$ normed frequency)” (Katznelson & al. 2010:8). This is a completely fallacious way of normalizing and provides the strangest results as these ‘normalized frequencies’ indicate for instance, that *love* is over three times more common in pop than in rock. But by using Katznelson & al.’s equation in reverse ($2.66 * 45,799 / 449 \approx 271$), we can see that *love* has 271 hits in the Pop subcorpus. When the percentages of *love* are calculated of all the words in their respective subcorpora, we get 0.33 percent for *love* in Rock and 0.59 percent in Pop, which means that *love* is, in fact, only a bit less than two times more common in the Pop subcorpus in comparison to the Rock subcorpus. In addition, percentage calculations revealed that based on Katznelson & al.’s data, *love* is actually most commonly used in the Country subcorpus (0.86% of all words), not in Pop. The numbers in Table 2 are thus useless in making comparisons between the genres. However, the listing of the most frequent lexical words in each genre and the relative frequency between the words inside a single subcorpus in comparison to each other is data that is somewhat usable for, e.g. comparing the results to other studies.

Rock	Pop	Country	Hip Hop
27,898 words	45,799 words	31,579 words	67,758 words
254 words/song	449 words/song	282 words/song	622 words/song
LOVE 0.85	LOVE 2.66	LOVE 2.42	LIFE 1.19
TIME 0.78	BABY 2.93	WAY 1.04	TIME 1.15
WAY 0.71	WAY 1.64	HEART 1.01	ASS 1.04
PAIN 0.61	GIRL 1.37	BABY 0.86	BABY 0.94
WORLD 0.54	BOOM 0.95	NIGHT 0.77	THING 0.82
LIFE 0.48	HEART 0.91	TIME 0.73	WAY 0.74
BABY 0.44	TIME 0.88	MAN 0.70	GIRL 0.73
EYES 0.44	LIFE 0.60	LIFE 0.64	MUSIC 0.67
HEAD 0.41	WORLD 0.59	DAY 0.62	BITCH 0.66
HEART 0.35	MAN 0.54	THING 0.59	MONEY 0.62

Table 2. The ten most frequent content words in Katznelson & al.’s data. The normed frequency is miscalculated so the numbers are not comparable between subcorpora. Adapted from Katznelson & al. (2010:13).

As can be seen from Table 2, according to Katznelson & al.'s data, *love* is the most frequent content word in Rock and Country (and Pop if *baby's* normed frequency is higher than that of *love* because of a mistake). Content words (also known as lexical words), such as nouns and adjectives, are words that have to do with what a text is about, while function words, such as auxiliary verbs and prepositions, "relate content words to each other" (Stubbs 2001:39-40). Based on Katznelson & al.'s data, *love* seems to be an important theme in rock, pop and country. Although *love* shares the same rank in each of the three genres, it does not automatically indicate that it carries the same importance in each subcorpus. Moreover, when we look at *way*, we can see that it is at the third rank in Rock and Pop, while in Hip Hop it lies three ranks lower. Because of this rank difference, one could interpret *way* to be less important in hip hop than in rock, for instance. But as Rayson points out, comparing word ranks and their relative positions in lists can lead to false interpretations (2008), as comparing ranks does not take into account the frequencies or percentages of occurrence.

Katznelson & al. make rather audacious interpretations based on their data. They claim that the list (Table 2) reveals cultural ideations and "shows the type of influence they [the genres] enact upon their audiences" (2010:14). In addition, they see the words in the Rock column in Table 2 to pertain to "the heartbreak type of love [they] hypothesized" (2012:14) as opposed to "the carefree love of pop" (2010:14). I feel that in order to make the results agree with their hypothesis and preconceptions Katznelson & al. are, perhaps, reading too much from the words only. They do not report having conducted any searching or concordancing (discussed in 4.4) in order to see in what kinds of contexts the top lexical words in their study occur. As is pointed out by Scott & Tribble about word lists, "... what at first appears simple often hides a mass of complex phenomena" (2006:11).

To sum up, the corpus linguistic studies into song lyrics that I have encountered are few and far between and their usefulness for genre studies is limited. The studies by Schneider & Miethaner (2006) and Logan & al. (2004) are not focused on investigating how lexical items makes genres distinct from other genres and the study by Katznelson & al. (2010) is simply of low quality.

3.3 The Keyword Approach

To define 'keyword', 'word' needs to be defined. What counts as a word and what does not is a major question in linguistics and corpus linguistics, and a common topic of debate. Stubbs divides

'word' into two distinct senses: 'word-form' and 'lemma'(2001:24). Word-forms are units that occur in actual texts (Stubbs 2001:25), whereas a 'lemma' can be seen as an abstract concept that is part of vocabulary (Fitschen & Gupta 2009:553, Stubbs 2001:26) or "a name for a class of morphologically related words" (Fitschen & Gupta 2009:553). For instance SEE can be designated as the lemma that represents multiple word forms: *see, sees, saw, seen*¹⁰ etc. In essence, this study investigates word forms and thus I will use 'word' to refer to word forms. Here, contractions such as *n't* in *don't* and *'re* in *you're* are seen as word forms as well. This is related to two things: a) the software used in the present study automatically counts these forms as separate words¹¹ b) separating these word forms as distinct units allows investigating style.

The term 'keyword' has been used in alternative ways in different fields, and as Bondi points out "[t]he notion ... has no well-defined meaning in language studies" (2010:2). Based on how different researchers have studied keywords, Stubbs (2010:23-32) divides the senses in which 'keyword' have been used into three categories, which he considers only loosely related:

- Sense 1: Words and culture
- Sense 2: Words and texts
- Sense 3: Phrases and schemas

The sense to be studied in this thesis is number 2, in which the definition of keyword is purely statistical. This is the sense given in Scott & Tribble (2006:55), where keywords are defined as "items of unusual frequency in comparison with a reference corpus of some suitable kind". For them, keyness – the quality of being statistically significant – is a textual quality, not a linguistic, cognitive or cultural quality (2006:56) (although statistical keyness may probably be used for interpretations on these latter three qualities in some cases). That is to say, words can be key in particular texts when compared to other texts, but this does not reflect the overall keyness of words in the English language as a whole (2006:55).

Senses 1 and 3 relate to culturally important words (sense 1) and phrases (sense 3), so they can be said to be related to sense 2 but they take a different viewpoint on what counts as a keyword

¹⁰ Lemmas will be written in SMALL CAPS while word forms –and keywords because they are word forms in this study– in *italics*.

¹¹ The CLAWS POS tagger (discussed in more detail in 4.2, 4.4 and 5.4) counted words in contracted forms as two separate words. For instance, the tagger automatically separated words such as *don't* and *I'm* into two tagged words; *don't* was divided into the finite form of *do* and the adverb *not* resulting in *do_VD0 n't_XX*. AntConc and Wordsmith Tools (discussed in 4.4) separated contracted forms into separate words similarly.

(Stubbs 2010:32). In sense 1, important cultural keywords are identified based on learned intuition (2010: 23-25). In sense 3, a corpus-driven approach is applied, the focus being on uncovering cultural schematas and lexico-grammatical patterns (2010: 28-32). Stubbs (2010:32) considers the conceptual relations of the three senses so loose that he is reluctant to compare results coming from such different scientific traditions. Therefore, what makes a word key is not necessarily obvious.

So what do statistically unusually frequent words tell us about texts? As Stubbs points out, keywords (in all three senses) are “pointers to complex lexical objects which represent the shared beliefs and values of a culture” (2010:23), which, in my opinion, also crystallizes why studying keywords is a good method. Keywords are a guide to elements to be studied further with the ability to reveal not only important topics, themes or styles of texts but also text functions (Bondi 2010:3, Wynn 2008:733). Scott & Tribble define keyness as “a quality words may have in a given text or set of texts, suggesting that they are important, they reflect what the text is really about, avoiding trivia and insignificant detail” (2006:56). In comparison to studying frequency lists, Scott & Tribble (2006:58) point out that studying keywords is more telling of the ‘aboutness’ of texts than simple frequency lists because the top words in frequency lists are function words (e.g. *the*) with little semantic content while the top lexical words (e.g. *time*, *know*) are semantically too general to be considered keywords. That is, high frequency of occurrence in itself is not enough to imply keyness or importance of words.

Not all keywords are indicators of aboutness, as function words can also be identified as statistically important keywords. Even though function words are the most frequent words in texts, they do not appear as keywords unless they occur significantly more often in the target corpus (the text(s) studied) in comparison to the reference corpus (the texts against which the target corpus texts are compared). When they do appear as keywords, they do not point to the content of texts but instead are indicators of style (e.g. Bondi 2010:6). In addition, Wordsmith Tools identifies ‘negative keywords’, which are words that occur significantly less in the target corpus versus the reference corpus. In this study, the goal is to identify and analyze all types of keyword, that is, both content words and function words, and both positive and negative keywords.

In studying a set of texts, a further method is to investigate 'keykeywords' in addition to keywords. While KWs (keywords) are identified as being key in what the software (Wordsmith Tools, discussed in 4.4) considers a single text, KKWs (keykeywords) are keywords across multiple texts (Scott & Tribble 2006:78). For instance, in this study, the KWs for black metal are those that are identified when all the 40 song lyrics texts in the black metal subcorpus are treated as a single text. In contrast, the KKWs are identified by first extracting the KWs of each individual text (each song) in the black metal subcorpus and then finding out which of those KWs occur in multiple texts of the subcorpus. Studying KKWs can be useful as according to Scott, "KKWs are informative of aboutness, and to a much lesser degree, of style" (2006:83). Scott implies that keykeywords are not necessarily more key than keywords but they can be when "the genre in question is quite specific" (2006:83).

Analyzing the words identified as key is not necessarily very simple for a multitude of reasons. If a corpus is not lemmatized, as it is not in this study, different word forms are counted as different words. So even if all the words forms of SEE together but not individually occur significantly frequently in a target corpus SEE might not be identified as key if the corpus is not lemmatized. Moreover, the different word forms of a lemma may have, and often do have, differing collocation patterns (Stubbs 2001:26-29). 'Collocation' refers to the frequent co-occurrence of words (2001:29), which makes 'collocates' "words which tend to occur frequently in the vicinity of the search term." (Wynne 2008:725). For instance, the different word forms of CONSUME take different collocates: *consuming* occurs in phrases like *consuming passion* and *time-consuming* while *consume* and *consumed* co-occur with different words, e.g. *calories*, *energy* and *oil* (Stubbs 2001:27).

Another issue is that words in text cannot be properly analyzed without regard to context. As Stubbs points out, "word and context are inseparable" (2001:100) because the meaning of a word in a text is not independent but dependent of the environment in which it occurs (2001:16). Similarly, Milizia writes that "the aboutness of a text depends on the context in which it is embedded" (2010:134). In this study, this pertains to how the KWs indicating aboutness are interpreted. For instance, it is reasonable to assume that a word like *ruins* can be related to different things in historical texts in comparison to heavy metal song lyrics.

Considered that most words are polysemous (Hoey 2009:973), it becomes inevitable to consult the context in which keywords occur to be able to take a stance on their semantics. For instance, *counter* and *fast* are words that can have very different meanings depending on the context and the part of speech they occur in: *She stood on the counter.* vs *Her argument ran counter to his claim.* In his case study of the word *dry*, Hoey discusses 21 uses of *dry*, in all of which the word has a more or less different context-dependent meaning, for instance, “He also had a dry wit...”, “The result was a cold, dry, emotionless dissertation...” and “Not a dry eye in the house, we wager”(2009:975-977). In determining the meaning of a word then, one needs to recognize “the word’s relationships with other words in the same semantic field” (2009:974). The concept of ‘semantic field’ pertains to words having “different relations to each other, sometimes logical relations of sameness, difference and entailment, and sometimes vaguer relations within a topic area or semantic field.” (Stubbs 2001:35) These relations can be of a different type, such as, synonymy, antonymy, hyponymy, superordinacy, etc. (Stubbs 2001:36-38, Hoey 2009:974-975). A method to take this matter into consideration in the analysis is concordancing, which is discussed in 4.4.

With regard to the semantics of keywords, it is also worth remembering that “...individual words are not always the units of meaning” (Stubbs 2001:32). As the examples in the paragraph above showed, the meaning of words is often realized in the context in which they occur. Some words regularly occur in constructions where they carry little or no meaning. This is called delexicalization (2001:32). It is seen in the usage of common verbs like TAKE and MAKE, for instance in *take a decision*, *make a mistake* (2001:32), where the verbs’ contribution to the overall meaning of the expression is small. Other constructions include expressions where an adjective shares part of the meaning with a noun but where the adjective is not really necessary to understand the noun, for instance *physical proximity*, *scientific analysis*, *general trend* (2001:33). These relations between words should be taken into account if possible when keywords are analyzed. In fact, Stubbs writes that “words should be studied, not in isolation, but in collocations” (2001:45).

In interpreting the results, it also needs to be kept in mind that the choice of reference corpus affects the resultant keywords in ways that are not yet completely understood (Scott 2010). An important point is made by Baker (2004) quoted by Rayson: “a key word analysis will focus only on lexical differences, not lexical similarities” (Rayson 2008:526). So if one of the subcorpora in this study were compared to, say, the whole *Contemporary Corpus of American English* (COCA),

the keywords identified, revealing lexical differences between the corpora, might actually indicate what is textually typical of song lyrics in general instead of indicating aboutness or style of the lyrics of particular heavy metal subgenres. This would be due to the text-types of the target corpus and reference corpus not being the same, as the text-types in the COCA are all other than song lyrics.

As identifying keywords is based on statistical methods, some keywords will receive a higher keyness value than others, that is, some words will be statistically more significant than others. However, the words with the highest keyness value are not necessarily the most important keywords as such, since when rarer words are key, they get assigned higher keyness values than more common words will (WORDSMITH 6.0 FAQs), because rare words are more unlikely to occur in any text in any case. This is relevant when function words occur as key, as these words are very common and frequent in all texts. To ascertain how common or uncommon the words identified as key in this study were, the website *Word Frequency* was utilized. *Word Frequency*, which is based on COCA, provides a freely available list consisting of 5,000 most common words (lemmas). Of course, it has to be borne in mind that the song texts in this study are written by people from bands all across the globe. Only some of the song lyrics are written by American authors so there are many different kinds of Englishes in the lyrics.

In applying the keyword method to heavy metal lyrics it is worth taking into account some of the research into heavy metal. According to Weinstein, “[i]mportant words or phrases are more clearly articulated by the singer than the rest of the lyrics” (1991:34). However, when it comes to extreme metal, Kahn-Harris points out that death metal lyrics are “decipherable only with the aid of a lyric sheet” (2007:3). This is something that a textual analysis based on written lyrics simply cannot take into consideration, unless the song lyrics were actually listened to and transcribed for enunciation, which certainly would not be a simple process. However, Weinstein adds that these important words and phrases are repeated in the chorus section of songs and that “[t]he key phrases enunciated by the singer do more to convey a song’s meaning than do the entire lyrics (1991:34)”. She argues that “[t]he words of a song function for listeners, in metal and in rock in general, more as isolated words and phrases than as integral poetic texts” (1991:125) and that ‘evocative words’, such as *evil*, *black*, *night* and *death* are relevant for interpreting the meaning of song lyrics. She goes on to argue that metal lyrics should be read through the metal code (c.f. Fabbris’ rules codified in a genre discussed in 2.1) and that certain words and passages in lyrics

should be paid more attention to while others are less relevant (ibid.). As identifying keywords is based on the repetition of words, and these repeated words are deemed to be important in light of the theory of the approach, a keyword analysis is well-suited for identifying these repeated words central to the meaning of songs.

3.3.1 Earlier Studies Utilizing the Keyword Approach

Keyword analysis has been used to study a wide range of topics, of which I will now present a few to illustrate the flexibility of the method and the possibilities it offers for a linguistic analysis of texts. Tribble applied the method to investigate a newspaper, *The Guardian*, to find out what its main themes were and how those themes were presented in it as well as how those themes changed between the years 1996-2001. In addition, he studied the newspaper's gender balance in terms of visibility of male versus female pronouns and titles etc. (Scott & Tribble 2006:161-177). Tribble's approach can be considered a data-driven one, as he had not decided on any particular hypothesis or what the particular words or phrases to study were. He identified particular political personae that were written about a lot during 1991-2001 as KWs, for instance, Clinton, Blair, Milosevic, Bush and Pinochet. By studying the collocates, he found out, among other things, that *Clinton* often co-occurred with *administration* and *presidency*, while Blair co-occurred with *government* and *Milosevic* with *regime*. He explains this difference to be due to differences in semantic prosody between the collocates: *regime* has negative connotations and suggests dictatorship and warmongering, while *government* is more neutral and can be used for referring to one's own to ally governments. With regard to gender balance, Tribble found out that the frequency of male titles and male pronouns in comparison to female ones in *The Guardian* was manifold.

Using keyword analysis and investigating collocational patterns of keywords Milizia compared Tony Blair's speeches to those of George W. Bush, arguing that phrases are usually better indicators of aboutness than single words (Milizia 2010). In particular, she studied two words identified as KWs in Blair's texts (*climate* and *change*) focusing on their collocational relation. She found out that in almost 90 percent of the instances of *climate*, the word co-occurred with *change*, either in the phrase *climate change* or in other n-grams¹² such as *the issue of climate*

¹² a habitual collocation comprising of "two or more words which are not necessarily felt to be idiomatic and which co-occur more often than chance would predict for no obvious reason other than habit" (Milizia 2010:128). Other terms for related but different types of lexical relations include: 'word cluster' (Scott & Tribble 2006:19;32), 'lexical bundle', 'extended unit of meaning', 'congram' and 'multi-word unit' (Milizia 2010:128).

change and *change in the climate*. The relation was one sided, however, as *change*, which occurred over twice as frequently as *climate*, was found to occur in other contexts two thirds of the instances, that is without co-occurring with *climate*. In this case, studying collocations (of *climate*) revealed that it is climate change in particular, which was a major topic in Blair's speeches.

Scott presents an example of the keyword approach applied to studying the words spoken by the characters in Shakespeare's play *Romeo & Juliet* using all of Shakespeare's other plays as a reference corpus (Scott & Tribble 2006:59-63). That is to say, words found especially frequently in *Romeo & Juliet* in comparison to other Shakespeare's plays were identified as key. Most of the 48 KWs were nouns indicating aboutness but 6 were items related to style: *O, Ah, thou, art, wilt* and *she*. As *thou* and *art* pertain to the 2nd person singular, Scott argues that their abundance reflects the "intimate nature of the theme" (2006:60) of the play. Examining the concordances revealed that while both *Ah* and *O* were used in mockery, *Ah* was more likely to occur with a negative sentiment, as in *Ah, what an unkind hour* (2006:61).

As we have seen, the keyword method can be used to study a variety of different linguistic features of text. Sometimes it can bring up surprising elements to study, such as the exclamations *O* and *Ah* and the forms of BE in *Romeo & Juliet*. Sometimes it points to cultural/stylistic phenomena as in the choice between *government* versus *regime* in describing a state's governing group of people. I believe that the keyword approach can be useful in genre studies in general and that it can reveal linguistic phenomena that might otherwise be left unnoticed.

4 Data & Methods

This section discusses the data collection procedure (4.1), the pre-processing of the data (4.2), the compilation and design of the METAL corpus (4.3) and the methods used in analyzing the data extracted from the corpus (4.4)

4.1 Data Collection

The multipartite data selection procedure of the study was somewhat complex as I was aiming for a maximally objective way of choosing the primary sources. First, the heavy metal subgenres to be studied were chosen. Then, of the selected subgenres, albums (and at the same time bands) representing those subgenres were picked, after which one song was elected out of each album.

Important subgenres of heavy metal were attempted to be identified by utilizing Google's search engine. There are a multitude of sites that list and/or describe different metal subgenres but none of them seems to have an identical list. Five lists were compiled by counting the subgenres mentioned in the lists or articles on heavy metal subgenres on five different webpages: "Heavy Metal artists, style and genres" in *About.com*, "Heavy Metal" in *AllMusic*, "Metal" in *Last.fm*, "Browse bands by genre" in *Metal-Archives* and "Heavy Metal Subgenres" in *Wikipedia*. The lists were compared for subgenres common to all sites, with the purpose of establishing the subgenres that are regarded the most important or common. When those subgenres that did not appear in all of the five lists were excluded, seven subgenres remained: black, death, doom, power, progressive, traditional heavy and thrash metal. Of the seven, five (black, death, power, thrash and traditional heavy) were chosen rather subjectively based on two points: Firstly, my hypothesis is that the five subgenres chosen would yield the most interesting or diverse results. I decided to exclude doom metal, anticipating that its lyrical content would not differ significantly from that of death and black metal. Secondly, I find progressive metal a very difficult genre to define formally, because the aspect of progressiveness is often combined with another style yielding for instance 'progressive death metal' or 'progressive power metal' (this is also readily visible in the genre labels assigned by *Metal Storm*, a website used for data collection in this study, described in the next paragraph). I consider progressiveness not so much a distinct style but rather an aspect that has to do with some kind of breaking of musical boundaries. That is, breaking away from musical conventions, such as typical song structures and the time signatures. The five subgenres in this study are described above in 2.3.

As I wanted to objectivize (or at least remove my own subjectivity from) the choice of the albums for this study, an outside 'authority' was utilized, namely, the website *Metal Storm*. *Metal Storm* is "an international metal website with over 2.5 million pageviews per month" (*Metal Storm*), with users from all across the world. At the time of obtaining the data (November 2013), the site had information (line-up, styles, releases, videos, trivia, upcoming concerts, news, reviews, interviews etc.) on almost 8,000 bands and more than 51,500 albums¹³. Registered users can submit and edit this information, which is subsequently checked by staff members. Users can also rate albums and conduct searches based on different parameters.

¹³ The types of album include studio, EP, live, DVD, compilation, single, demo, boxset and various artist, of which studio albums are approximately 13,200.

The *Metal Storm* search tool was utilized in the choice of which albums to use as sources of the lyrics. Consequently, the lyrics feature both native and non-native English as the bands listed in *Metal Storm* are from all across the globe. The search parameters were set in such a way that in each genre, studio albums with at least a 100 votes were searched for, with the search results listed according to the albums' ratings based on the average of the votes given to them. The lists created by the search engine were copied onto Excel files (one file for each subgenre). For the subcorpus of each genre, 40 albums were chosen, and, for the sake of representativeness, each song was picked from a different album and by a different band. The selection of albums began from the top of the list generated by the *Metal Storm* search tool, that is, from the highest ranking album. Then the following criteria were applied for excluding an album from the corpus:

- I. The album had an additional genre specification that overlapped with the other subgenres in the study. For instance, Death's album *Symbolic* was classified as both 'progressive thrash' and 'progressive death'. It was therefore excluded. Other albums with mergers such as 'blackened thrash' and 'death thrash' were also excluded).
- II. An album by the same band already appeared earlier (more highly ranked) in the list (e.g. even though Metallica had several albums with high ratings only the best ranked album by Metallica was chosen).
- III. The album had instrumental songs only. (e.g. *Reflections* by Apocalyptica)
- IV. An album already appeared in the list of another subgenre with a higher rating (e.g. Metallica's *Ride The Lightning* was found at rank 1 in the thrash metal list, while *St. Anger* was at rank 99 in the heavy metal list. *St. Anger* was therefore excluded).

In the case of the thrash metal albums chosen, for instance, the method of exclusion resulted in that the 40 thrash metal bands and their albums were found between ranks 1-169 in the *Metal Storm* search (out of a total of 279 albums that appeared in the search results). That is, to reach 40 albums, almost 130 albums were excluded, which was largely due to violations of criteria 1 and 2. In the subgenre of traditional heavy, the 40 albums were found between ranks 1-272 (out of 339), similarly mostly due to violations of criteria 1 and 2.

This method was chosen, because it was the most systematic one that I could think of and also independent of my subjective ideas on which genre a band should be placed in. As it is, I disagree with several of the genre labels assigned by *Metal Storm* (e.g. I would not consider the bands

Children of Bodom and Wintersun 'extreme power metal' although I can see where the classification comes from) and if you have listened to metal music, you will probably disagree with some of the classification as well. Unfortunately I could not find information on whose decision the labeling was based on, that is, whether it is a single *Metal Storm* staff member's decision or the collective opinion of the users of the site or something else. In the case that the labels are actually based on the decision of single individuals, the labeling is quite subjective. Be that as it may, from the perspective of this study, this was the most doable and objective way to pick the sources of data that I could think of with regard to the resources and time available.

Having selected the albums, a song from each album needed to be picked. For simplicity's sake, I chose the first song from each album, provided that the following criteria were fulfilled:

- I. The lyrics have to include at least 100 words
- II. The lyrics have to be in English (or at least 100 words have to be sung in English).
- III. The lyrics need to be found online.

Criterion I. was deemed necessary to obtaining enough material for analysis in addition to the requirement that there be enough repetition in song lyrics to enable the identification of keywords of a single song text. When the first song of an album did not meet the above mentioned criteria, the second song was selected, and if the second song did not fulfill the criteria, the third was selected and so on, until an appropriate song was reached. When none of the songs in an album filled the criteria, the album was excluded, although this happened only a few times.

Two websites, which provide song lyrics for educational purposes, were used to retrieve the lyrics: *A-Z Lyrics Universe* and *Dark Lyrics*, the latter of which – being specialized in heavy metal – provided most of the material. The sites were chosen because the song lyrics there were readily available and they enabled an easy process of copying the song lyrics into text files for corpus use. In addition, *Dark Lyrics* was the biggest database of heavy metal song lyrics known to the author at the time of writing, advertising themselves as “the largest metal lyrics archive on the Web” and including the lyrics for “13 800+ albums from 4500+ bands” in March 2014 (*Dark Lyrics*). A list of the songs chosen for this study can be found in Appendix 1.

4.2 Data Pre-processing

The following processes were executed to prepare the song lyrics for corpus research. The lyrics were copied from *A-Z Lyrics Universe* and *Dark Lyrics* to text files. Only the actual lyrics were included; song and album titles were excluded from the text files. The files were organized alphabetically and numerically (numbered according to alphabetization) and were of the form <Subgenre## BAND TITLE - Album Title - Song Title>, e.g. <Power03 AVANTASIA - The Metal Opera - Reach Out For The Light>. I will be referring to particular songs in this study similarly as to the file names here, that is, by writing them inside angle brackets.

Several of the lyrics included other text besides the actual lyrics sung (usually some deictic metadata) and there were some corrections that needed to be made. The following modifications were performed, where required:

- Section movement indicators in square brackets, such as [VERSE], [PRE-CHORUS], [BRIDGE], [SOLO], and [THEME] were removed.
- Singer or character role indicators such as [John] or [Gabriel] were removed.
- Indicators such as [REPEAT CHORUS] and other indicators signaling repetition were replaced with the actual words sung in the song. In unclear cases, the songs were listened to in order to find out which the actual lyrics to be repeated were.
- Obvious typos were corrected, (e.g. the 'hye' in *hey, hye, hey, hey, hey* in <Heavy02>)
- A space was added after a period, after a sequence of (usually three) periods or after a comma or other punctuation marks when it was required (for the software used in this study to be able to distinguish between words), e.g. a space was added in *You've come to inherit what's Yours...The Mansion* in <Heavy23> where 'Yours...The' forms a single word from the perspective of many word processors.
- Metatextual comments and written narration (usually in *italics*) not actually heard in the song were removed from the lyrics when words in the text were recognized as such, e.g. "*carrying on [A]nnihilator's fine tradition of covering weird, psychological illnesses '[P]alace' is narrated by the emotion of guilt...*" in <Thrash01>.

After the modifications had been made, the corpus files were tagged for part-of-speech using the Free CLAWS WWW Tagger (available at <http://ucrel.lancs.ac.uk/claws/trial.html>). Tagging is useful in that it allows searching for grammatical patterns. This tagger was chosen because it was

available and free of charge but was still capable of tagging the amount of text in this study (one can “enter up to 100,000 words of English running text” in the tagger (Free CLAWS WWW Tagger)). To tag texts with the CLAWS tagger, one chooses a tagset and simply inserts text into a text box, either by writing or pasting, and clicks the ‘tag text now’ button. This was done for each of the 200 text files.

Consequently, two separate corpora, a tagged and an untagged one, were created for different purposes: the untagged one for word lists and keywords, the tagged one for testing tagging accuracy of song lyrics and for grammatical analysis through searching and concordancing (discussed in 4.4). For simplicity’s sake, I will nevertheless refer to each of the corpora as the METAL corpus.

The accuracy of the CLAWS Tagger was investigated by manually checking the correctness of tags in 10 song lyric files; numbers 11 and 31 of each subcorpus, e.g. <Tagged Power11> and <Tagged Power31> in the PM subcorpus. Erroneous tags in these 10 files were corrected, although this has the effect that only 10 of 200 files were completely correctly tagged, as 190 tagged texts were left unchecked and uncorrected. The results of the tagging process will be discussed in 5.4.

4.3 The METAL Corpus

The corpus compiled for this study, the METAL corpus, consists of the lyrics of 200 songs. The corpus has five subcorpora, one for each metal subgenre in this study, each subcorpus consisting of the lyrics of 40 songs by 40 bands. Statistics on the corpus, extracted by using the software WordSmith Tools (discussed in 4.4 below), can be seen in table 3.

Corpus	# of words	Mean song length (words)	Type-token ratio	Mean word length (letters)	# of long words (8 or more letters)
METAL	40,915	202	12.73	4.14	3189 (7.8%)
Black metal (BM)	7,167	179	26.96	4.36	694 (9.7%)
Death metal (DM)	7,480	187	26.15	4.32	739 (9.9%)
Power metal (PM)	9,146	229	17.87	3.98	607 (6.6%)
Traditional heavy metal (TRAD)	7,852	196	18.81	3.86	374 (4.8%)
Thrash metal (TM)	8,550	214	21.76	4.24	775 (9.1%)

Table 3. METAL Corpus Statistics.

Some observations can be made from Table 3. In average, power metal songs are the longest in terms of the number of words per song, while black metal songs are the shortest. Here, 'words' refers to the number of word tokens¹⁴. A reason why the songs in the PM subcorpus are longer could be due to a more faithful transcription of the lyrics in the choruses of power metal songs in comparison to the songs in the other subcorpora. If the power metal song lyrics in this sample happen to have more repetitions of the chorus written in them – not because of actual repetition of sung lyrics but because of choices of inclusion versus omission by whoever is responsible for writing down the lyrics – then this results in a higher mean song length and smaller type-token ratio.

The type-token ratios of the subcorpora seem to suggest that death and black metal songs show the greatest lexical richness, that is, the largest vocabularies, as the figures indicate that there is less repetition of the same word types in their lyrics than in the songs in the rest of the subcorpora. Power metal seems to have the lowest richness. However, the subcorpora are not exactly of the same size, and—as can be seen by comparing the type-token ratios of the corpora to their size—the ratio tends to go down when the number of words in a corpus goes up. Because measures of lexical richness (such as type-token ratio) are dependent on the length of a text (e.g. Tweedie 1998), the type-token ratios here do not automatically indicate differences in lexical richness between the subcorpora.

When we consider the mean word length and the number of long words – words that have 8 or more letters¹⁵ – we can see from Table 3 that the subcorpora with the highest type-token ratios also have the highest values for mean word length and the highest percentage of long words: the words are shorter in the PM and TRAD (Traditional heavy) subcorpora than in the three extreme metal subcorpora. A reason for this could be the tendency of extreme metal songs' lyrics to have more specific topics or just more specific or richer vocabulary than what the non-extreme metal song lyrics have. If we agree with Kahn-Harris' claim that extreme metal lyrics describe phenomena in more detail (see 2.4), it would make sense for words in extreme metal to be longer as, based on my experience, words denoting more precise concepts (often latinate or French loans) usually have more letters. This could also be a stylistic choice that the songwriters have

¹⁴ 'Token' refers to instances of a word while 'type' refers to a particular word such as *the* (Scott & Tribble 2006:12-13). For example, the three instances of *the* in the second sentence in this paragraph are tokens of *the*.

¹⁵ This was chosen as a threshold because the number of words decreased rapidly in each subcorpus after the length of 8 letters.

made, perhaps going for a more literary style. In Chapter 5 we will see how this relates to keywords.

As Table 3 indicates, the corpus is very small, which is a methodological issue. Baroni & Evert point out that even a “sample size of 100 [texts] is small by the standards of statistical inference” (2009:783). The present sample of each subgenre is only two fifths of that. Moreover, with regard to the size of the reference corpus, Scott reports that according to Sardinha (2004) “a reference corpus should be about 5 times the size of the node text.” (Scott & Tribble 2006:65), the node text being the text studied. Lacking time and a more efficient data collection method, I decided to carry on with this tiny corpus, although any statistical inferences based on my results have to be regarded with certain skepticism.

When identifying keywords, a single subcorpus is compared to the four other subcorpora of METAL, which are used as a single reference corpus. Although this does not meet Sardinha’s demands of corpus size, Scott reports of a similar type of approach by Culpeper (2002), in which keywords of the main character of Shakespeare’s *Romeo & Juliet* were identified by comparing the 5,000 words spoken by Romeo to the 14,000 words spoken by the other characters in the play (Scott & Tribble 2006:63-64). So even though the reference corpus in Culpeper’s study was less than three times the size of the node text, keywords indicating what Romeo’s lines are about were successfully identified. I surmise this success has to do with the nature of keyword analysis, as the analysis reveals differences and not similarities between the corpora compared (e.g. Rayson 2008). Therefore, I believe the keyword method can be used successfully in this study as well as the target and reference corpora here are rather similar to those in Culpeper’s study in the sense that the reference corpora in this study are small and represent the same text type as the target corpora.

With regard to the representativeness of the METAL corpus, the issue is slightly different than with a more conventional corpus which aims to represent, say, all the language used by a particular population with all its different registers as, for instance, COCA does. The text type is the same in each subcorpus of the METAL corpus as each subcorpus consists of song lyrics. As the goal is to identify lexical differences between them, the similarity of text type between the corpora is an advantage, since features pertaining to song lyrics in general should not emerge in the keyword lists (similarly to Culpeper’s (2002) study where features of Shakespeare’s language use typical of

all Shakespeare's texts did not appear as KWs). With regard to the choice of reference corpus and text type, Milizia makes a similar observation: "as evidence in the previous findings shows, by comparing spoken with spoken we lose the 'speechiness' that emerges when comparing spoken with written" (2010:132).

However, an issue possibly questioning the representativeness of the present corpus is that the subcorpora subsume songs representing derivative genres, the proportion of which was ignored in the compilation process, because so many bands actually represent derivative genres. That is, only some of the songs in the corpus represent the main genres of heavy metal, such as death metal or thrash metal, at least according to the labeling by *Metal Storm*. More often than not, an album used as a source for lyrics in this study was characterized with terms further determining the genre of the album. For instance, *Metal Storm* labels <Thrash22 MESHUGGAH> as 'technical thrash', 'math' and 'progressive' and <Death20 GOJIRA> as 'progressive death'. Therefore, there is a chance, for instance, that the analysis of one of the subcorpora, say the PM subcorpus, will yield results that pertain to the label 'progressive' rather than 'power'. This could happen if a considerable proportion of the songs in the PM subcorpus were actually from albums labeled 'progressive power' by *Metal Storm*. The features found could thus be an indicator of something that is going on in 'progressive' lyrics in general, that is, in 'progressive power', 'progressive death', 'progressive thrash' and so on. However, as was mentioned earlier, I feel that the aspect of progressiveness in its elusiveness does not define style that much.

In terms of balance, the corpus is relatively satisfactory as the number of words does not disproportionately differ between the subcorpora. The greatest difference in the number of words is found between the black metal and the power metal subcorpora (just under 2 000 words), although in a small corpus like this, even this degree of difference has its repercussions, as was seen in the case of the type-token ratios.

4.4 Methods of Analysis

The software Wordsmith Tools (available at <http://www.lexically.net/wordsmith/>) developed by Mike Scott was utilized for compiling word lists for the whole corpus and each subcorpus separately. The software has a word list function that creates word lists by calculating the words in text files uploaded on the software. For reasons mentioned in 4.3, a reference corpus word list consisting of all the words in the other four subcorpora was created for each subcorpus. For

instance, the reference corpus word list for thrash metal consisted of the 160 song lyrics in the BM, DM, PM and TRAD subcorpora. Word lists were exploited to retrieve word frequencies and to calculate the keywords of each subcorpus. Wordsmith Tools has a keywords function that automatically calculates the keywords of a word list when the word list is compared to a reference corpus word list.

The keywords function in Wordsmith Tools employs statistical functions to identify keywords. A typical statistical function that is also used in this study is Log Likelihood which gives numerical data on the probability of outcomes. From a log-likelihood value, it is possible to calculate a p-value which then indicates the statistical validity of an outcome. For instance, a p-value of 0.01 suggests that there is a 1 percent chance that the result was due to chance; the lower the p-value, the more significant the result. Usually a p-value of 0.01 is a good indicator of statistical significance. In social sciences, even a value of 0.05 is usually acceptable (Wordsmith Tools manual). However the default p-value threshold in Wordsmith Tools—which dictates which words will be included and which will be excluded in the keyword lists—is lower (0.000001). Setting a lower p-value threshold than what is typical for statistical research in general is reasonable because studying words using statistical methods is not quite the same as studying other phenomena since it can be taken as a premise that words in language do not occur completely randomly in any case¹⁶. Because of the small size of the METAL corpus, however, a higher minimum p-value was chosen to ensure that there were enough keywords to analyze, although their status as KWs thus becomes more questionable.

The minimum distribution was set at 3, which means that for a word to be identified as key in a subcorpus, it had to occur at least in three texts, that is, in 7.5 percent of the texts. In Wordsmith Tools the default is 5 percent but as the subcorpora are very small, 5 percent would mean that it would be enough for a word to occur in only two texts. To me, occurrence in two texts only seems hardly representative of a genre.

For the keykeywords, a word list for each song in each subcorpus was created. The same reference corpora word lists mentioned above were utilized to establish the individual keywords of each song. Not all songs yielded keywords, as some of the song lyrics contained so little repetition that no words would rise as key. When a keyword had a distribution of 2 or higher (that is, was present

¹⁶ According to personal communication with Turo Hiltunen on 20 Mar. 2014.

in two or more song lyrics), it was included in the keykeyword list of a subcorpus. This choice is due to the fact that so few words occurred as keywords in multiple texts that the threshold had to be lowered from 3 to 2 in order to have something to analyze. In this small corpus, a distribution of 2 actually amounts to 5 percent of the number of songs in a subcorpus. This percentage is typically used as a minimum (e.g. Scott & Tribble 2006:78). Of course, anything definitive about the importance of a word in a subgenre cannot be claimed when the keykeyword identified is based on an occurrence of 2, since two songs hardly represent a whole subgenre. The results and analysis are presented in 5.1 and 5.2.

The software AntConc (available at <http://www.antlab.sci.waseda.ac.jp/software.html>) was utilized for concordancing and searching for collocates. Wynne defines concordance as “a listing of each occurrence of a word (or pattern) in a text or corpus, presented with the words surrounding it” (Wynne 2008:710). AntConc allows one to conduct searches using one or more words as a search term, listing all the hits and their nearby context (e.g. 25 words to each side of the word). Concordances are a valuable tool with which the textual context of words can be taken into account. As Wynne points out, “[a]utomatic extraction of word lists, collocate lists, etc. can lead the analyst to deal only with words abstracted from the texts where they occur, and taken away from the place where meaning is created” (2008:711). As discussed in 3.3, the processes contributing to word meaning are complex and dependent on the context. Therefore one should not deal with keywords merely as words abstracted from texts (as Katznelson & al. (2010) do) but study them in context. By reading concordances one can access the context of occurrence of a word and determine its meaning (Wynne 2008:711).

Concordancing was utilized to investigate the context of occurrence of keywords to establish in what kinds of meanings the keywords occur (as dictated by the context). As AntConc allows the sorting of results (e.g. alphabetically according to the first word right of the node word), different kinds of sorting was used in performing searches on the data to examine whether some kind of patterns emerge. More often than not the file in which a search word (a KW) occurred was examined further in context of the whole text when investigating the concordance line was not enough for deducing the contextual meaning of the search word. However, although attempted here, arriving at a singular meaning for a particular word in song lyrics texts is far from straightforward, as the meaning often has to do with the context of the whole song text, in addition to which many song texts are open for interpretation. Investigating the contextual

meaning of KWs was deemed necessary for the interpretation of the significance of the KWs. In analyzing the semantics of words, the *Free Merriam-Webster Dictionary* was often consulted, especially in the case of polysemous words.

To examine collocational patterns, the keywords of the Black Metal¹⁷ subcorpus were investigated further for collocations and keyword linkages. Keyword linkage refers to a collocation between keywords, the quality of having 'co-keyness' or "shared keyness in the same text" (Scott & Tribble 2006:73). Studying keyword linkages is a further tool for studying aboutness, as words that are key and also co-occur are likely to be important to a text.

Tagging the METAL corpus for part-of-speech has the advantage that it allowed for searches according to grammatical category (e.g. tense) and lexical category (e.g. noun, adjective). The tagged METAL corpus was used to investigate patterns of the word *of* in detail, as it was identified as a keyword in two subcorpora (see 5.4.1), in addition to which the keyness of the -ing participle is discussed in 5.4.2.

5 Analysis¹⁸

This chapter discusses the results and analysis of the study. First, word frequencies in the METAL corpus and the subcorpora are investigated and compared to song lyrics corpora discussed in 3.2. Then, the keywords identified are analyzed, which is followed by an analysis of the keykeywords. Keywords of individual song texts are also briefly discussed. Finally, the application of part-of-speech tagging to the song lyrics in the METAL corpus, and some investigation using the tagged corpus are discussed.

The most frequent words in the METAL corpus and in each subcorpus (Table 4) are very similar to each other. In all cases, the top 10 words consist of function words, such as articles, pronouns and forms of the verb BE. This tells us very little about the aboutness or style of the genres, as there are no lexical words in any of the top lists and no major differences between the lists. A similar listing can basically be seen in any English language corpus, as the top words always tend to be function words (see e.g. Scott & Tribble 2006:26, Wynne 2008) as in Table 4.

¹⁷ The name of a subgenre is capitalized when it refers to a subcorpus (Black Metal) and written in lower case when it refers to the actual genre (black metal).

¹⁸ The layout of this chapter is rather spacey because many lines had to be added in between several paragraphs in order to make the tables fit into a single page.

METAL	Black metal	Death metal	Power metal	Thrash metal	Traditional heavy metal
THE	THE	THE	THE	THE	THE
OF	OF	OF	I	TO	I
TO	AND	TO	TO	YOU	YOU
I	IN	I	OF	OF	TO
AND	TO	IN	AND	I	AND
YOU	A	AND	IN	AND	A
A	I	YOU	YOU	A	OF
IN	MY	MY	MY	IS	IT
MY	IS	A	A	IN	S
IS	YOU	YOUR	YOUR	ME	IN

Table 4. The 10 most frequent words in the corpora sorted according to rank.

When the ten most frequent lexical words are considered, the picture already looks more interesting (Table 5). As the lexical words do appear in the lists because of much repetition in the respective subcorpora, they tell us something about the typical words used in the subcorpora and, with certain caution, they can be interpreted to suggest something about the typical content in the song lyrics of the subgenres. However, as was noted in 3.3, the most frequent lexical words tend to be too general in content to be considered key. And as we can see, several of the top lexical words seem rather poor in semantic content as they are such common words in English in general. For instance, *time* (at rank 52 in Word Frequency), *see* (r:67), *way* (r:84), *take* (r:63) and *come* (r:70)¹⁹, which are all present in Table 5, are all among the hundred most common words and not very indicative of aboutness.

¹⁹ The ranks here are those of the most common POS that the words occur in, based on the list in Word Frequency. For instance, the ranking 84 for *way* is that of *way* occurring as a noun, as opposed to it occurring as adverb, in which case it is at rank 4090.

METAL	Black metal	Death metal	Power metal	Thrash metal	Traditional heavy metal
DEATH 0.35	BORN 0.43	SEE 0.41	WAY 0.57	DEATH 0.83	GOT 0.39
LIFE 0.33	DEATH 0.39	SOUL 0.39	TIME 0.40	TIME 0.53	KNOW 0.36
TIME 0.32	LIGHT 0.35	EYES 0.35	LIGHT 0.35	BRING 0.51	LIFE 0.36
SEE 0.31	DARK 0.32	LIFE 0.32	HEART 0.34	LIFE 0.41	WAY 0.36
WAY 0.27	WORLD 0.29	DEATH 0.28	LIFE 0.34	DIE 0.40	COME 0.34
WORLD 0.25	NIGHT 0.27	FEEL 0.27	FIRE 0.33	BLOOD 0.34	LOVE 0.32
EYES 0.22	TIME 0.25	PAIN 0.24	SEE 0.33	SEE 0.34	SEE 0.31
COME 0.22	BLACK 0.24	WORLD 0.24	MIND 0.30	FEAR 0.25	GO 0.27
LIGHT 0.22	COME 0.24	COME 0.21	ROAD 0.24	FIRE 0.25	WORLD 0.27
DIE 0.21	LIFE 0.24	GOD 0.21	NIGHT 0.23	WORLD 0.25	TAKE 0.25

Table 5. The 10 most frequent lexical words and their percentage of total words in the corpora sorted according to rank.

A striking observation can be made from the METAL column: the single most frequent lexical word in all the 200 songs together is *death*, which is followed by *life* at rank two. In addition, the verb *die* appears in the METAL column. Based on these words' appearance in the top list, one could interpret that life and death is a common topic to be sung about in heavy metal. Indeed, *life* occurs as the collocate of *death* 25 times (out of a total of 142 tokens of *death*) in the METAL corpus. *Death* also appears in the columns for Black, Death and Thrash metal but not in those for Power and Traditional heavy. This supports Kahn-Harris' observations, as he argues that lyrics in extreme metal – which the three subgenres in whose list the word is found represent – tend to focus on the more sinister aspects of life. On the other hand, *light* and *life*, words that themselves seem hardly sinister, are also frequent in the METAL corpus and more interestingly, in the BM subcorpus. By consulting concordances, however, we can see that *life*, too, often appears in rather gloomy contexts in the METAL corpus (Table 6).

y is guilty The crime is life The sentence is death DA	Thrash08
d we pray it would last Life burns! Life burns! Life	Heavy05
nging death where he met life Demons ride to reach the	Power23
t comes To take one more life that can be near In the	Black25
s, take away thee mortal life Demigod, Satan son, comm	Death15
ect Looking back upon my life At the waste it's been T	Thrash25
uicide, ritual to end my life Behemoth infest my fate	Death15
dered for the rest of my life Condemned in a jail cell	Thrash35
on and on I've lived my life , fulfilled my dreams, I'	Power39
And now I'm watching my life go My wit is falling ap	Heavy14
is determined to take my life Here comes the night! N	Black07
- Satan Suicide, end my life I must die - Satan Suic	Death15
ams to decay. All of my life I've been waiting for so	Power05
ourself Save yourself My life is fading All is lost Tu	Heavy26
ral I was buried, too My life - it ended with yours An	Heavy35
I live my life like there's no tomorrow	Heavy38
only way out Restart my life Or self destruction To c	Thrash35
Rip my heart out, rip my life out Rip my eyes for I'll	Death34
w as I'm sitting here My life passing by I recall the	Power39
o call my own I gave my life to reason And now I'm wa	Heavy14
ammed to tell, end of my life Wrath of God - Satan Sin	Death15
Transylvania All natural life has for a long time ago	Black25

Table 6. Some concordances of *life* in the METAL corpus, organized according to the first word to the left.

Life, which occurs 134 times in the METAL corpus, appears in expressions like *The crime is life / The sentence is death* <Thrash08>, *Bringing death where he met life* <Power39>, *Suicide, end my life / I must die - Satan* <Death15>. *Life* appears in more positive contexts a few times as well, as in *I've lived my life, fulfilled my dreams* <Power39>. The concordance searches indicated, however, that the most typical way for *life* to occur is in a context where the meaning has to do with the absence of life, life ending, life going to waste etc.

For somebody who is already acquainted with these subgenres, the slightly less common words in the columns of each subgenre seem at least not to contradict with my preconceptions of the lyrical styles or with the research discussed in section 2. For instance, in the BM subcorpus the frequently occurring words *born* (r:1092), *black* (r:254), *death* (r:418), *night* (r:209) fit the gloomy atmosphere of black metal music. The situation is similar with the two other extreme metal genres as *death* and *pain* (r:940) are found in the DM subcorpus while *death*, *die* (r:404), *blood* (r:693) and *fear* (r:988) are found amongst the frequent lexical words in the TM subcorpus. On the other hand, *death* is absent from the columns for PM and TRAD. Instead, *love*, is found in the TRAD list, supporting the observation that traditional heavy is more linked to rock 'n' roll music than the other genres (*love* is also found in the top 10 lexical word of Rock in Katznelson & al.'s study).

Interestingly, *death*, *night* and *black* are among the four words that Weinstein gives as examples in her discussion of evocative words important for obtaining the meaning of heavy metal lyrics (see 3.3).

Compared to the results of Logan & al. (2004) and Katznelson & al. (2010) (discussed in 3.2), the data here suggests that *death* is a word that is particularly typical of heavy metal lyrics, because it is absent from the most frequent lexical words of other genres, based on a comparison to Logan's and Katznelson & al.'s top 10 lists (in 3.2). There is a striking difference in the top lexical words between the METAL corpus and Katznelson & al.'s corpus, exemplified in Table 7: whereas in the Rock, Pop and Country subcorpora the most common lexical word is *love* and in Hip Hop *life*, in the METAL corpus it is *death*. *Love* is also present in Logan & al.'s (2004) data (see Table 1 in 3.2): it is among the top lexical words of their Reggae, Country, Newage and Rock subcorpora, being absent only in Rap (similarly, *love* is also absent from the top lexical words in the Hip Hop subcorpus in Katznelson & al. (2004)). On the other hand, *baby* is a frequent lexical word in Katznelson & al.'s data, present in all genres, but not in the METAL corpus word lists.

METAL	Rock	Pop	Country	Hip Hop
DEATH 0.35	LOVE 0.33	LOVE 0.59	LOVE 0.86	LIFE 0.19

Table 7. The lexical words at rank 1 and their percentages²⁰ of the total words in their respective subcorpora in the METAL corpus and in Katznelson & al.'s (2010) corpus.

To sum up, the most frequent lexical words in the BM, DM and TM subcorpora suggest a more sinister atmosphere than those in the PM and TRAD subcorpora. The data can be interpreted to suggest that Metal lyrics in general be focused on the more sinister aspects of life (or Chaotic themes in the Weinsteinian sense) than the lyrics of the genres in Katznelson & al.'s (2010) and Logan & al.'s (2004) studies.

5.1 Keywords of the Subcorpora

This section discusses the analysis of keywords identified for the subcorpora of the METAL corpus. I will present interpretations on the keywords to establish an overview and then analyze selected KWs in more detail. Some keywords will not be included in the discussion because in some cases either the frequency or distribution of the KW was very low and/or simply no patterns were found. Some of the KWs that occurred also as KKWs will be discussed in 5.2. The KWs of the Black Metal

²⁰ The percentages were arrived at by reversing the normed frequency equation used by Katznelson & al., which gave the number of occurrence, and then applying a standard percentage calculation.

subcorpus will be discussed more thoroughly to illustrate the advantages of a more meticulous investigation and also some methodological issues related to the data.

All the subcorpora yielded a different number of keywords (see Table 8). This can – at least partly – be explained by considering the differing song lengths and type-token ratios between the five subcorpora (see Table 3 in 4.3). The BM and DM subcorpora, which yielded the lowest numbers of KWs, have the smallest numbers of words per song and the highest type-token ratios, compared to the other subcorpora. It is fair to conclude that when there is less repetition in general, the result is fewer keywords. Since the PM subcorpus has the lowest type-token ratio but still has fewer KWs than the TM and TRAD subcorpora, the type-token ratio must not be the only factor affecting the number of KWs identified. My assumption is that in the TM and TRAD subcorpora it is particular word types that are repeated often and they therefore rise as key but in the PM subcorpus the repetition of words is more spread, that is, the number of word types repeated is higher while this repetition is not focused on particular word types as much as in the TM and TRAD subcorpora. Due to time limitations, this was not investigated further.

Let us consider the function words and contracted forms that emerged as KWs first, since I believe those are the most significant in the results of this study. The BM and DM subcorpora both have *of* as a KW while the BM subcorpus has *the* as well. This suggests that longer noun phrases are used more in death metal and black metal because the preposition *of* is commonly used as part of a genitive noun phrase. Indeed the preposition is used in this function in a majority of the instances in which it is found in the data here (as we will see in 5.4.1), such as in the phrase *Angel of Death* in <Thrash30>. Moreover, *of* and *the* appear as negative KWs (Table 9) in the TM and TRAD subcorpora, that is, they occur significantly less in those texts, which suggests that heavy (as in grammatically complex) noun phrases are not preferred in those subgenres. The fact that the contracted forms of BE, *re* (from *are*) and *m* (from *am*), as well as the contracted form of the adverb NOT, *t*, appear as negative keywords in the BM and DM subcorpora supports the interpretation that black metal and death metal lyrics avoid using a colloquial style preferring instead a more literary style with their heavier noun phrases.

Black Metal	Death Metal	Power Metal	Thrash Metal	Traditional Heavy
BORN	THY	WAY	CAN	RE
OF	THEE	M	BRING	GOTTA
THE	ART	LEADING	DEATH	GONNA
AS	CAST	STORM	DOWN	GOT
RUINS	PEACE	MAGIC	FUCK	YEAH
BEFORE	OF	ALL	NIGHTMARE	YOU
NOR	BEYOND	THERE	ME	T
FUNERAL	THEIR	GAME	BLOOD	ROCK
TREES	SOUL	LEAVING	WALL	AIN
MEN		HEART	SHIT	SHAKE
DARK		RIDE	TAKES	IT
RETURN		HAIL	HELL	BABY
CLOSER		LL	COST	LOVE
		I	YOU	IF
		GHOST	SHOW	GIRL
			MURDER	ROLL
			TRAPPED	TURNING
			DIE	KING
				DEVIL
				SAME
				ANGEL
				FEELING
				SWEAT

Table 8. Keywords of the subcorpora sorted according to keyness rank (the uppermost word of each column has the highest keyness value).

In the three other subgenres, power metal, thrash metal and traditional heavy, a more colloquial approach to writing lyrics seems to be favored. The PM subcorpus has the contracted forms *m* and *ll* (from *will*) as KWs, and the TRAD subcorpus has the contracted forms *re* (*are*) and *t* (*not*) as KWs. Moreover, *of* and *the* appear as negative keywords in both the TM and TRAD subcorpora. These findings suggests that power metal, thrash metal and traditional heavy employ a more colloquial style in their lyrics in comparison to death and black metal, as contracted forms are typical of spoken language and less formal registers, while *of* and *the* are more common in more formal

registers. A similar observation on the pertinence of the two function words to written language is made in Scott & Tribble (2006:107).

Black Metal	Death Metal	Power Metal	Thrash Metal	Traditional Heavy
I	M	DEATH	LIGHT	SOUL
OUT	RE		OF	BY
YOUR	T		THE	THE
RE				DEATH
T				OF
CAN				
M				
YOU				

Table 9. Negative Keywords of the five subgenres sorted according to keyness rank (the lowermost word of each column has the lowest keyness value, that is, is most significantly negatively key).

Let us reconsider *death*. *Death* is a word that is not only present in the top 10 lexical words of the BM, DM and TM subcorpora but it was also found as a KW in the TM subcorpus. In contrast, it was identified as a negative KW in the PM and TRAD subcorpora from whose top 10 lexical words lists it is also absent. This suggests that power metal and traditional heavy lyrics avoid using the word *death* (even if that topic were discussed in the lyrics) whereas extreme metal lyrics use the word more commonly, especially thrash metal. Curiously, it is not a KW in Death Metal, although the word is a part of the name of that genre.

It is interesting that two personal pronouns and a personal determiner (*I*, *you* and *your*) are found as negative KWs in the Black Metal subcorpus. This suggests that in black metal lyrics, the use of personal pronouns is avoided in comparison to the other subcorpora. This could be due to an avoidance of referring to characters, the narrator or the audience/listener of the song by using pronouns. Another explanation is that black metal lyrics perhaps avoid themes related to personal relations, which then removes the need to use personal pronouns. Indeed, all the other subcorpora have some kind of personal word forms in their keyword lists, except for Black Metal. This certainly does not contradict the view (e.g. by Kahn-Harris 2007) that misanthropy is a theme present in black metal lyrics, since the absence of personal reference can also be seen as a manifestation of misanthropic attitudes.

I will now discuss keywords of each subcorpus, starting with Death metal and finishing with Black metal. Only few (9) KWs were identified in the Death Metal subcorpus, and of those only 4 (*art*, *cast*, *peace* and *soul*) are content words. Accordingly, the DM keywords are not very indicative of the aboutness of death metal lyrics. Moreover, of the 4 (potential) content words one actually appears mostly as a function word: *art* is found mostly as a second person singular form of the verb BE, that is, a function word in 13 of the 15 occurrences. In the two cases in which the word appears as a noun the context is rather grim: *Inside the wound I wish to learn / The art of fucking you* <Death03>; *The art of veneficium...* [poison making] / *This we learned from you...* <Death12>. The keyness of pronouns *thy* and *thee* – forms that along with *art* as a pronoun pertain to archaic English – and of the relatively uncommon preposition *beyond* (r:747), along with the occurrence of the preposition *of* as a KW and of *m*, *re* and *t* as negative KWs, suggest a tendency for a more literary writing style.

According to the data here, *peace* is a relatively rare word in heavy metal, occurring 15 times in the METAL corpus, of which 9 instances are found in the DM subcorpus. Examining the concordances of the KW reveals that its context is not related to, say, promoting peace or aspiring towards peace but to more sinister content, for instance the concept of ‘peace in death’ (keywords in examples are written in bold):

- 6a) End of life converges
 A sweet release
 Away from turmoil
 To find peace
 <Death10>
- 6b) Father Satan, I'll find peace when I am God
 Suicide sacrifice
 Destruction of holy life
 Blood of unholy knife
 Satan I sacrifice
 TAKE ME!
 <Death15>

There is an ambiguous case in which the meaning was not straightforward to interpret but seemed to have a similar idea as in 6a) and b), 'peace in death': *I feel peace within as your soul meets a new world / A world where you may step up to a higher learning* <Death36>. Other instances in which *peace* occurs were related to the absence of peace or difficulty in finding peace, e.g. *You'll never rest in peace* <Death38>; *Peace of mind is hard to find* <Death34>.

The KW *soul* occurred most commonly (in 10 out of 13 song texts) in contexts that had to do with dying, decay or withering of the soul:

- 7a) To foresee the end of a soul...Epitaph. <Death27>
- 7b) Curtains fall on religion's role
The cancer of the soul <Death09>
- 7c) Murder is etched in the deepest chasms of the soul. <Death39>
- 7d) The rotting of the soul.
The pain denies the soul. <Death31>
- 7e) my feet were tired
and my soul was half a way
to a shadow <Death33>

The way *peace* and *soul* are used in Death Metal, the contexts in which the words appear, is really not obvious from the keyword list. Moreover, the connotations and expected contexts of occurrence for these two words in general is probably something else than 'peace in death' or 'decay of the soul'.

With regard to Power Metal, the lexical words in the PM subcorpus are not very surprising, although I was perhaps expecting there to be more words that have to do with a fantastic or imagined world. That is, words such as, 'dragon' or 'warrior' (which were absent from the data) as my impression is that many power metal songs' lyrics are about heroes or warriors, ancient or mystical kingdoms, magical things and fighting evil. The keyword that most evokes fantastic images is *magic* (frequency: 12). It is only present in three texts, of which in two it seems to refer to actual magic (as opposed to a magical atmosphere for instance): *You have read the magic lines / leading to the light in your mind...* <Power03>; *He's now coming from the middle lands / handling*

proud his magic sword <Power30>; *Magic and steelgods lead us to a new dawn...* <Power30>. In one song the word occurs in the chorus section, where its meaning is open for a metaphorical or non-metaphorical interpretation: *After the storm when the magic has gone / Drown in the tears of a mandrake* <Power09>.

Power Metal keywords that I see as relating to a warrior or hero culture – which according to my experience are thematic constants in power metal – are *leading, hail, ride* and *heart*. This is visible in the data, exemplified by the excerpts below:

- 8a) First in the line of fire, first into hostile land
Tanks leading the way, leading the way <Power31>
- 8b) A hail to the metal invasion
A heavenly kingdom on earth <Power15>
- 8c) Watch me ride on the thunder
Followed by the warriors of steel, hail!!! <Power22>
- 8d) While war is in my heart
death is by my side <Power06>
- 8e) We are the knights on our glory ride
Law defenders, raise your swords <Power15>
- 8f) handling proud his magic sword
Glory, pride and honor ride with him <Power30>

Leading occurred in contexts with connotations to war in most cases (5/9). *Hail* occurred 9 times, of which 6 in the meaning 'to greet', and 3 times as a precipitation metaphor. Of those 6 occurrences in the meaning 'to greet', half had connotations to war/warriors as in 8b) and 8c). *Heart* and *ride* had these connotations only in few instances, such as 8d)-8f). References to glory and heroism can be seen in 8b), 8e) and 8f). *Magic* also appears in 8f) so arguably all the five KWs in 8a)-8f) can have connotations to warrior culture.

Moving on to Thrash Metal, the lyrics of the subgenre seem to make frequent references to death, based on the appearance of *death, blood, murder* and *die* as keywords. The other words occurring in the keyword list have either rather negative (*nightmare, trapped, hell*) or neutral (e.g. *can,*

bring) connotations. It is also noteworthy that the TM subcorpus is the only one in which obscene words, such as *fuck* and *shit*, occur as keywords. Using profanity could thus be a stylistic feature characteristic of thrash metal.

Murder occurs in its literal sense in 7 out of 9 occurrences in 3 out of 4 song lyrics that the word is found in, for instance: *Commit cold blooded murder / like Nazis during World War Two* <Thrash02>; *He wreaks of murder / His mind filled with sin* <Thrash10>; *Murder and murder / Chaos, disorder* <Thrash32>. In <Thrash13>, it is found in a metaphorical sense, referring to the aggressive behavior of metal fans in the audience:

- 9) Murder in the front row
 Crowd begins to bang
 And there's blood upon the stage
 Bang your head against the stage
 <Thrash13>

Cost only has few (6) occurrences in the TM subcorpus, found in 4 songs. In two songs (3 occurrences) it is contextually related to calamity: *Wars result, is it really worth the cost? / So much pain, the suffering never ends* <Thrash17>; *Heat waves and blizzards / Global death's the cost* <Thrash18>. In the other two, the meaning has to do with personal sacrifice: *The cost to me has been my soul / For these sins I've paid* <Thrash 25>; *You ain't-never seen no one like me / Prevail-regardless what the cost might be* <Thrash34>. The data shows that Thrash Metal keywords tend to be related to rather sinister content or imagery (cf. Weinstein's Chaotic themes).

The Traditional Heavy keywords are the most connected to typical (and perhaps more mainstream) rock lyrics in comparison to the KWs of the other four subcorpora. These include *love, rock, yeah, baby, girl, feeling, sweat* and *shake*. A large number of the KWs can be seen as pertaining to a colloquial style of expression: *re* (contracted form of *are*), *gonna, gotta, got, yeah, t* (contracted form of *not*), *ain* (part of *ain't*) and *baby*. The relation of traditional heavy music to rock 'n' roll music can most clearly be seen in the appearance of *rock* and *roll* as KWs and in the contexts in which they occur:

- 10a) No doubt, you're stressin' out
 That ain't what rock n' roll's about <Heavy02>

- 10b) Rock! Rock! till you drop
I say Rock! Rock! to the top <Heavy11>
- 10c) Know your body's made to move, feel it in your guts
Rock 'n' roll ain't worth the name if it don't make you strut <Heavy27>
- 10d) What do you see in the center of the public eye
Rock stars on smack, and a serial killer fries
Radicals blame suicide and murder on our form of art
Brainwash the youth, you know they claim we all play a part <Heavy32>

In 10a), c) and d), the role of rock ('n' roll) is discussed: it is not about stressing (10a)), the music has to be able to stimulate an emotional response (10c)), while in 10d) the lyrics point what the narrator of those lyrics sees as misconceptions about the genre as "form of art". The idea of heavy metal as rock music and the praise of rock music, which are both mentioned by Weinstein as lyrical themes, is clearly present in these lyrics.

The fact that *girl* appears as a KW can be seen as another indicator of heavy metal's relation rock music. It occurs in 4 songs, all by bands – Alice Cooper, KISS, Deep Purple and Volbeat – that are classified also as hard rock by many instances (*Metal Storm*, *Wikipedia* and *Metal-Archives*). Taking this into consideration, *girl* might actually pertain more to hard rock than heavy metal lyrics.

The KW *love* occurs often (in 8 out of 11 texts) in contexts in which it relates to the difficulties of love or love lost, or the absence of love. For instance, *The more love I feel for you, the more you fade away* <Heavy04>; *I got no love, no love you'd call real* <Heavy38>; *I still remember when I saw your face across the room / Told me to take you but the price of love would seal my doom* <Heavy24>. A completely different sentiment is found in <Heavy12> where a *gift of love* is sent to metal fans. <Heavy11> celebrates non-serious love or sex (c.f. Weinstein's idea of sex in heavy metal being fun and without commitments): *Fast, free and easy, livin' for today / Gotta lip service, get it while you can / Hot, sweat 'n' nervous love on demand*. The lyrics in the TRAD subcorpus that have a direct proclamation of love is found in a song about loving a car: *I love it and I need it / I seed it / Eight cylinders all mine* <Heavy10>.

Based on the presence of *angel* and *devil* in the TRAD column, biblical imagery could be interpreted to be part of the style of traditional heavy lyrics. However, each of the words occurs

only in few texts, 3 and 4 respectively. They were probably identified as key because each of the words is found multiple times in the chorus of a single song. 14 of 18 occurrences of *angel* are found in <Heavy04> in the phrase *you're an angel witch*, while 10 of 12 instances of *devil* are similarly present in a single expression, *Runnin' with the devil* in <Heavy38>. In half of the instances, *angel* is used figuratively, as in *Yes I'm coming / Angel from the sky* in <Heavy12> (and the phrase in the chorus of <Heavy04>). One of the instances seems to refer to an actual angel: *Holy angel lift me from this burning hell / Resurrection make me whole* <Heavy18>, while in <Heavy16> the reference is actually to the devil: *Our fallen angel vexed / Was banished from the sky*. Two of the three occurrences of *devil* can be seen as direct references to the devil: *THE WHEEL IS SPINNING, SPAWNED IS THE SEED / OF THE DEVIL INCARNATE - EVIL EXTREME* <Heavy20> and *John, you have been found guilty of being in league with the devil and of having used heathen and forbidden rituals* <Heavy31>. The phrase in the chorus of <Heavy38> is more ambiguous and can more easily be interpreted as metaphorical. The fact that these KWs are mostly present in a single song in each case lowers their status as KWs of the subgenre. On the other hand, biblical imagery was found to be used in heavy metal lyrics by Weinstein, which in turn consolidates their status as KWs.

As mentioned above, Black Metal keywords will be discussed in more detail. As seen in Table 10, the frequency and distribution of the BM keywords varies to a large extent. The actual keyness of KWs like *funeral* and *closer* is somewhat questionable as their overall frequency in the METAL corpus is so small, and their distribution is rather low, three instances each. Although a distribution of 3 accounts for 7.5 percent of the texts in the BM subcorpus – a number that would be acceptable in a larger-scale study – the smallness of the corpus decreases the significance of this percentage. The two words are also slightly rarer, based on the list in *Word Frequency*, than the other BM keywords. As was noted in 3.3 rarer words tend to receive higher keyness values than more common ones, due to their lower frequencies in texts in general. Of course, *funeral* and *closer* are not rare words in English as such, but in comparison to the other BM keywords they are relatively rare. On the other hand, the KWs *born*, *of*, *the*, *as*, *before* and *dark* have relatively high frequencies and distributions, and they can therefore be judged as more likely to represent actual, generically important keywords.

Rank (f)	Keyword	Frequency	%	Distribution	RefCorp. Freq.	Keyness (LL)	Commonness Rank
1	BORN	31	0.43	7	20	46.54	1092
2	OF	278	3.88	39	787	46.28	4
3	THE	619	8.64	40	2148	39.58	1
4	AS	59	0.82	27	111	27.70	c:33/i:49/r:129
5	RUINS	10	0.14	3	3	21.63	-
6	BEFORE	19	0.27	9	19	20.33	i:220/c:368/r:706
7	NOR	9	0.13	3	3	18.73	1066
8	FUNERAL	7	0.10	3	1	18.51	2908
9	TREES	8	0.11	4	2	18.37	596
10	MEN	11	0.15	6	7	16.64	94
11	DARK	23	0.32	16	34	15.82	860
12	RETURN	10	0.14	5	6	15.68	v:473/n:1245
13	CLOSER	6	0.08	3	1	15.35	3030

Table 10. Black metal keywords and statistics. LL = Log Likelihood. Commonness rank is based on the 5000 word lemma list available at the website *Word Frequency*. The letters in the Commonness rank column stand for (as they do in the website *Word Frequency*) c: conjunction, i: preposition, r: adverb, v: verb, n: noun. *Ruins* (or *RUIN*) are not found amongst the 5000 most common lemmas as a noun, the part-of-speech in which it occurs in the Black Metal subcorpus.

The keyness of the function words *of*, *the*, *as*, *before*, and *nor* suggests a stylistic difference between black metal and the other subgenres since the other subcorpora do not have such words as KWs, with the exception of Death Metal which includes *of*. These are function words more typical of the written register, another observation supporting the interpretation that black metal lyrics strive for a literary expression.

Before is used in several meanings in the lyrics. Interestingly, the KW is most often found in the sense 'in front of' in Black Metal. For instance, *The face here before him, meant to allure him* <Black16>; *Tremble before us / Lords of the star-veiled red sepulchers* <Black10>. The 7 times (or 8 if one ambiguous case is counted in) *before* occurs as a preposition like this, the word following the KW is always a personal pronoun or determiner: *me*, *my* or *us*. Curiously the use of *before* in this sense is completely absent from the DM, PM and TRAD subcorpora and found only once in the TM subcorpus. That is, while *before* occurs 19 times in the BM corpus and 19 times in the reference corpus, it is used 7 (or 8) times in the sense 'in front of' in the BM subcorpus and only once in this sense in the reference corpus. In the other cases it is found as a temporal adverb or a conjunction, as in <Black22>: *Bring me gifts before you're dead*.

An analysis of the concordances of *as* revealed that about²¹ 15 of the 59 instances of the KW in the BM subcorpus are similes, such as in the following examples:

- 11a) Eternising the soul soaked estate
Solvable as fractional fragments <Black05>
- 11b) I am as a plague, born to the priestess
The secret amour of her archangelic rape <Black10>
- 11c) MY GRAVE ROSE TO THE WEST...
FOR CENTURIES LONG FORGOTTEN,
RELENTLESS AS THE HUNGRY GATES OF DAWN. <Black36>
- 11d) The wild cold deep black ocean's waves
As wide as sky above <Black04>

On the other hand, in the reference corpus *as* occurred as a simile only about 5 times out of a total of 111 instances, and was used in this sense in the lyrics of a mere two songs (once in the TM and TRAD subcorpora each): *Fast as a shark he'll cut out of the dark* <Heavy01>; *Bang your head as if up from the dead / Intense metal is all that you need* <Thrash13>. Since similes are commonly used literary devices, this observation supports the black-metal-as-literary-texts interpretation.

Although a rather subjective evaluation, the similes in 11a)-d) have more poetic air to them than the ones in <Heavy01> and <Thrash13>.

The KW *born* is a rather common word, at rank 1092 in American English, according to *Word frequency*. It is found across 7 song texts, although a closer look at the data revealed that 17 of the 31 instances (~55%) occurred in <Black34> wherein *born* can be seen as an evocative word:

- 12) Born again, born again!
I know I can't resist,
Born again, born again!
Yes, I'll be born again!
Born again, born again!
So take what you will!

²¹ I write 'about' because the analysis was rather cursory and superficial. Moreover, some of the cases were ambiguous.

Born again, born again! You know what you have to lose!

I know what you have to lose!

<Black34>

Scrutinizing the context of occurrence it can be seen that in 5 out of 7 song texts *born* co-occurs with rather 'dark' topic matter or imagery, as seen in example 13a)-e)

- 13a) It was calm before the storm
When the hate in me was born <Black08>
- 13b) We are as a flame born unto darkness
Desires burning in palatial glades <Black10>
- 13c) All Born To Conquer Born of Nordlights
Here's A World Of Winter Here's A World Of Ice <Black15>
- 13d) Flaming suns crested the horizon shadows born a cloudless day
believers of the rise and set of the moon darkened twilight into night... <Black21>
- 13e) *NEMESIS COME BORN*
COME BORN DARKNESS CHILDE <Black30>

In 13a) a negative emotion is what is born. 13b) is more poetic and open for interpretation as the meaning can be seen as either related to birth or the verb BEAR, as in 'to move while holding up and supporting (something)' (*Free Merriam-Webster Dictionary*). In either case, the imagery is rather gothic, evoking images of passion in the night. In 13c) the meaning has to do with power or warrior-likeness (*Born To Conquer*) of the narrator of the song and his comrades (this is revealed by the rest of the lyrics). *Born of Nordlights* is related to the romantic idealization of nature typical of black metal. 13d) is more ambiguous (as the lines are not exactly grammatical or are fragments) but the imagery is rather dark, invoked by the words such as *shadows*, *darkened*, *twilight* and *night*. In 13e) a child of darkness is born, while *NEMESIS* refers to opponent. If we take into account the larger context of the songs, that is, the song and album titles, we can see that the sentiment of hate in 13a) is emphasized as the album title of <Black08> is *Chaos Born* and the song title *Calm Before The Storm* can be interpreted so that *Storm* refers to hate about to be discharged.

If we take keywords to be “what the text boils down to” (Scott & Tribble 2006:56) – content words indicating aboutness – then *born* being a keyword would suggest that birth or something being born is a key concept in black metal based on this sample. This suggests that in general, something being born is a theme often sung about in black metal. Or, it could just be a genre-specific stylistic matter, that is, the usage of *born* could be due to typical word choice in black metal. Based on the above analysis it the word tends to occur with dark topic matter (although this is probably not very exceptional, taking into account black metal lyrics’ general tendencies towards grimness).

The KW *dark* is used in various contexts but the most commonly it is related to some kind of reference to nature (10 of 23 instances):

- 14a) Dark clouds enraptured by sinister mists
 All the bodies are hanged they no longer resist <Black09>
- 14b) Nor seek in mountains dark the landscapes of the men
 long lost to where no pathway goes <Black35>
- 14c) Every time this year
 This dark fog will appear <Black25>
- 14d) Far away in the dark glimpsing moonlight
 Sickening souls cry out in pain <Black17>

That *dark* is often related to nature in Black Metal as in 14a)-b) could be an indication of the romanticism present in black metal discourse. Let us consider the literary concept of the ‘sublime’, which can be seen as “a way of talking about the impact on us of certain dramatic or powerful manifestations of nature (towering mountains, storms, avalanches, etc) or supernature (demons, angels, ghosts)” (Voller 2008). It could be argued that black metal lyrics – with their romantic tendencies – make use of the sublime. However, this interpretation would need a more in-depth analysis of black metal lyrics for it to be validated. Another reason for the frequent occurrence of *dark* in black metal lyrics could be that it is part of the conventional vocabulary of the subgenre.

The linkages between Black Metal keywords were examined for patterns of co-keyness. Collocates were looked for within the distance of 5 words to each side of the search term, which provided the results presented in Table 11.

Keyword	Keyword linkages
BORN	AS, BEFORE, OF, THE
OF	AS, BEFORE, BORN, DARK, FUNERAL, MEN, RETURN, RUINS, THE
THE	AS, BEFORE, BORN, CLOSER, DARK, FUNERAL, MEN, NOR, OF, RETURN, RUINS TREES
AS	BEFORE, BORN, DARK, OF, RETURN, THE, TREES
RUINS	OF, THE
BEFORE	AS, BORN, OF, THE
NOR	DARK, THE
FUNERAL	OF, THE
TREES	AS, THE
MEN	DARK, OF, THE
DARK	AS, MEN, OF, THE
RETURN	AS, OF, THE
CLOSER	THE

Table 11. Keyword linkages of the Black Metal keywords.

As Table 11 shows, the keyword linkages are mainly between function words and content words, or function words co-occurring with other function words. That function words occur as collocates (in general) is hardly surprising considering the overall high frequency of function words in all texts. The only lexical keywords that collocate with each other are *men* and *dark*. However, this co-occurrence is only present in the lyrics of <Black35> in which the words are found in the chorus that repeats thrice in the lyrics:

- 15) I would not find the burning domes and sands
 Where reigns the sun, nor dare the deadly snows
 Nor seek in mountains dark the landscapes of the men
 Of long lost to where no pathway goes
 <Black35>

In fact, the two words would not have been recognized as collocates without the inversion of the typical adjective-noun word order in <Black35>. That is, were the expression *dark mountains* used instead of *mountains dark*, the distance between *men* and *dark* would be 6 words, which would be too far in terms of the search parameters used in the present study. There does not seem to be any kind of obvious connection between *dark* and *men* in 15). Of course, it is possible to concoct interpretations supporting the idea of a conceptual connection between the words. For instance, if

we agree that dark mountains are an uninviting and uninhabitable place, the lyrics can be seen to draw a contrast between harsh nature (the sublime) and the weakness, limitedness or mortality of man. Based on this data, studying keyword linkages is not very telling of aboutness of song lyrics.

5.2 Keykeywords of the Subcorpora

In this section, similarly to 5.1, first, general observations will be presented and then selected KKWs will be discussed in more detail.

Most keykeywords of a subcorpus occurred in as little as two different texts. A few of the texts (three in Black Metal and one in Black Metal and Thrash metal each) had so little repetition that no words at all were identified as key in them, e.g. <Black10>. A few KKWs occurred in three or four texts, with one, *I*, occurring in five texts in the TRAD subcorpus. In all subcorpora more KKWs were identified than KWs. Similarly to the keywords in each subcorpus, fewer keykeywords were identified for the BM and DM subcorpora than for the other subcorpora. The three extreme metal genres have fewer KKWs than Power Metal and Traditional Heavy, although Thrash Metal has more KWs than Power Metal does. The number of KKWs coincides fairly well with type-token ratios: the subcorpora with the lowest type-token ratios (PM and TRAD) yielded the highest number of KKWs. Conversely, the subcorpora with the highest type-token ratios (BM and DM) yielded the lowest number of KKWs. The TM subcorpus, whose type-token ratio is in between the subcorpora with high and low values, also yielded an intermediate number of KKWs. One reason for this could be differences in songwriting styles between the subgenres.

As was noted in 2.4, according to Kahn-Harris, extreme metal genres actively engage in transgressive practices, such as breaking away from typical song structures. Therefore it is reasonable to argue that Black and Death Metal having fewer KWs and KKWs has at least partly to do with the absence of choruses, which, in turn, is connected to breaking away from songwriting conventions pertaining to mainstream (non-transgressive) music. Indeed, a scrutiny of the whole texts of 6 songs with few (0 to 4) keywords, <Death18, 20 and 22> and <Black03, 10 and 40>, revealed that none of the songs have a recognizable chorus section in their lyrics.

Black metal	Death metal	Power metal	Thrash metal	Traditional heavy
<u>BEFORE</u> 3	<u>CAST</u> 2	AM 2	ALWAYS 2	A 3
<u>BORN</u> 2	HER 2	FACING 2	BONDED 2	<u>AIN</u> 3
<u>CLOSER</u> 2	HIS 2	FIRE 2	<u>CAN</u> 2	<u>BABY</u> 2
EVERY 2	OUR 2	FLY 3	CHANGING 2	BATTLE 2
<u>MEN</u> 2	<u>PEACE</u> 2	FORCE 2	<u>COST</u> 2	BLEED 2
MOTHER 2	RID 2	<u>GAME</u> 2	<u>DEATH</u> 4	COMING 2
NEMESIS 2	SHE 2	GLORY 2	FIGHT 2	CRIMSON 2
<u>OF</u> 3	SLAVES 2	GRIEF 2	<u>FUCK</u> 2	DON 2
RAIN 2	SUFFERING 2	<u>HAIL</u> 2	FUCKING 2	EVIL 2
<u>RETURN</u> 2	THOU 2	HAVE 2	HAND 2	FAST 2
SEALED 2	<u>THY</u> 2	<u>HEART</u> 2	HE 2	<u>GIRL</u> 2
<u>THE</u> 2	UPON 2	I 3	HIS 4	<u>GONNA</u> 3
<u>TREES</u> 2	WOUND 2	<u>LEADING</u> 2	IT 2	<u>GOTTA</u> 3
WHERE 2		<u>LEAVING</u> 2	<u>ME</u> 3	<u>GOT</u> 3
		<u>MAGIC</u> 2	<u>MURDER</u> 2	HANDS 2
		MIND 2	<u>NIGHTMARE</u> 2	HEY 2
		<u>M</u> 4	OUR 2	HILL 2
		MY 3	<u>SHIT</u> 2	HOLD 3
		NEVER 2	<u>SHOW</u> 2	I 5
		RAISE 2	<u>TAKES</u> 2	<u>IT</u> 4
		SAVE 2	THE 2	KNOW 2
		<u>STORM</u> 2	ULTIMATE 2	<u>LOVE</u> 3
		TODAY 2	<u>WALL</u> 2	M 4
		UNITE 2	<u>YOU</u> 3	ME 2
		US 2	YOUR 2	MY 3
		<u>WAY</u> 2		OUT 2
		WE 4		<u>RE</u> 2
		WIN 2		RED 2
		WILL 2		<u>ROLL</u> 2
		YEARS 2		<u>SHAKE</u> 2
		YOUR 2		<u>I</u> 4
				TRACK 2
				<u>TURNING</u> 2
				UP 2
				YOU 3
				WE 3
				WORLD 2
				<u>YEAH</u> 3

Table 12. Keykeywords of the subgenres sorted alphabetically. Keykeywords that are also keywords in the respective genre are underlined. The numbers indicate the instances of texts in which the words occur.

In each subcorpus, several of the KKWs were also identified as KWs (see Table 12). These KKWs could be considered to be particularly important, because of this 'double keyness'. However, this

does not seem like a satisfactory conclusion because there are other factors to be taken into consideration. For instance, *closer*, which is both a KW and a KKW in the BM subcorpus, only has a frequency of 6 and a distribution of 3. Three of the instances occur in one particular song in one single line where the phrase *come closer* is repeated thrice. Because a half of the instances of the word are found in a single song and its frequency is low, I would not judge this word to be that key in black metal – not based on this sample in any case.

Another word that was identified both as KW and KKW in Black Metal is *born*. As mentioned in 5.1, more than half of the instances of the word occur in the chorus sections of a single text. However, as it has a relatively high frequency in the subcorpus (f31) and it is found in fairly many texts (7 = 17.5%), its status can more safely be judged as key. Indeed, *born* has the highest keyness value of the Black Metal KWs whereas *closer* has the lowest.

A more uncertain word in its keyness is *trees*, which is found 8 times in 4 texts, with a slightly higher keyness value than *closer*. It occurs 4 times in <Black31> and is part of the evocative phrase *our trees will grow higher* in the chorus. Since it is a nature-related lexical word that can be seen as pertaining to romanticism – which according to Kahn-Harris is a typical feature of black metal discourse – it seems more tenable to consider *trees* as a keyword than *closer*, even if its frequency and distribution is rather low.

Interestingly, the Black Metal KKWs that refer to people, *men* (which is also a KW) and *mother*, also have references to nature in several songs. In 3 out of 4 songs in which *mother* occurs, the word is found in front of a word whose meaning has to do with nature: *Mother Nature* in <Black27>, *Mother winter* in <Black04> and *Mother north* in <Black32> (whose title is also *Mother North*). The KKW *men* is found in connection to nature in 3 out of 6 songs: *The city's walls rise before us men from the seas* <Black04>; *Men of the North we leave the shores in flames* <Black04>; *We're Men Of Iron We're Men Of Might / Behold The Power Against The World We Rise* <Black15>; *Nor seek in mountains dark the landscapes of the men* <Black35>. The singular form was also found in a similar environment in <Black15>: *I'm A Man Of Mountains I Stand My Side / A Man Of The Forests A Man Of Pride*.

When it comes to Death Metal only 6 content words are amongst the Death Metal KKWs. Of those 6, four are exclusively in the KKW list: *rid*, *slaves*, *suffering* and *wound*, all of which have a distribution of 2 and low frequencies ranging from 4 to 7. The first and fourth word can be seen to

pertain to themes of abjection, e.g. detailed descriptions of violence (KKWs in bold, KWs underlined):

- 16a) Blood gushes from the wound
The cut is wide and deep <Death01>
- 16b) To praise the night, to praise my unknown faith
Must I sure find a way? To infest the wound <Death03>
- 16c) Rid us of our human waste
Cleanse our earthly lives
- 16d) Tearing, rid you of your limbs.
Infection soon, infection soon sets in.
Peeling, rid you of your skin.

The KKW *suffering* is found in contexts in which it has to do with more abstract agony: *You are your own victim / Still suffering, an imperial pain is getting you* <Death08>; *In her sad face signs of suffering are drawn / Caused by tears she shed for so many nights* <Death34>. The KKW *slaves* is found in contexts where the narrator of the song asserts dominance over others: *let me reign as a god among slaves!* <Death06>; *Follow me so you can see visions from beyond the grave / Now become nocturnal slaves* <Death29>. Although these KKWs have low frequencies, their keyness is supported by their pertinence to themes already identified in the subgenre in Kahn-Harris' research, that is, serious and sinister content matter (*suffering*), the abject (*rid* and *wound*), and fantasies of control and subjugation (*slaves*).

Archaic English forms of BE are found most frequently in Death Metal. The KKW and KW *thy* is found 33 times in 4 texts, although 26 of those instances are present in a single text, <Death28>. Almost every line of the text has either *thee*, *thou* or *thy* in it, as seen in 17).

- 17) Blasphemer, Heretic, Defiler of the Sacred Ones.
Thou art Deprived of Your Limbs.
Thy Nose Shall be Split.
Thou art Cast Down and Overthrown.
Ra-Harmakhis Destroyeth Thee.
He Damneth Thee and Driveth Hooks into Thy Body.

Isis Sayeth in Mighty Voice,

"The Number of Thy Days are Cut Short.

Thy Bones are Broken to Splinters Thy Vertebrae are Severed."

<Death28>

There is a curious contrast between the detailed premonitions of brutal violence, and the archaic style – the third person verbal ending *-th* is also present – which almost evokes a poetic atmosphere. A somewhat similar style is found in <Death06> in which, like in <Death28>, references are made to Egyptian deities:

18) children ov²² Seth
blast wide the portals
unveil the raptures
ov Thy kingdom

<Death06>

These forms are always found in connection to some kind of higher powers in the present data. In <Death28> and <Death06> the references are to Egyptian deities, while in <Death18> there is an opposition between the individualistic philosophy of the narrator and the Christian tradition: *I am my own God / See the truth beyond / Through endless lies thy kingdom come*. This reference is not necessarily clear from these lines only but the title of the song *Left Hand Path*, and other lines in the lyrics support the presence of an anti-Christian sentiment: *I dip my forefinger in the watery blood / Of your impotent redeemer*. The lyrics of <Death32> are vaguer. There *thy* seems to refer to the universe as a deity that is beyond human faculties of comprehension: *Though none may hope to fathom thee / Still glorious are thy works of might*. The title of the album, *Cosmogogenesis*, and the numerous words related to nature present in the lyrics (e.g. *storms, land, sea, flames, thunders, ocean, cliff*) support this interpretation. In all the four songs in which *thy* is present, modern English pronominal forms (such as *you* and *your*) are consistently avoided. <Death28> does have two instances of *you* and *your* each but still, *thou, thee* and *thy* are used much more, as together they sum up to 58 instances in the lyrics.

The Power Metal keykeywords offer further support for the observation that power metal lyrics have an affinity for heroism and warrior culture. Amongst the KKWs are: *force, glory, save* and

²² Writing *ov* instead of *of* is a stylistic idiosyncrasy characteristic of this particular band.

unite. The KKW's *unite*, *we* and *us* also point towards a sentiment of 'we-ness' present in the lyrics. Consider the following examples, in many of which a sense of we-ness is combined with themes of heroism or war:

- 19a) Heeding The Call, one and for all
 never surrender, with glory we'll fall
 Brothers unite, let's stand up and fight
 fulfilling our fate, we are Heeding The Call <Power17>
- 19b) A higher empathy
 We are one
 We are free
 The cross is history <Power32>
- 19c) Call for vengeance, raise your steel
 We are the knights on our glory ride
 Law defenders, raise your swords
 Freedom for us all <Power15>
- 19d) In our hands we hold the future
 As we live so we will die
 Carry on to save mankind
 Back to back we stand as one <Power15>
- 19e) Raise your fist, evangelist!
 [...]
 Get down the roar we march into war
 We strike down the enemy <Power28>
- 19f) Five centuries have now gone by, for the love of our peers we will sing and rejoice
 <Power12>
- 19g) Pushing the frontline forth with a tremendous force
 (Far ahead, breaks resistance)
 Making the way for panzer corps
 (Shows no fear, self-subsistent)

Unlike in the other subcorpora, in the PM subcorpus there are several KKWs that have to do with time: *never*, *today* and *years*. However, no common patterns were recognized for these words. Instead, *never* seems to occur in similar kind of contexts as the KW *all*. They both often serve as kind of intensifiers, contributing to a dramatic or bombastic sentiment often present in the music. These findings prompted searching for a related word, *forever*, and it turned out that it is used more (f11) in the PM subcorpus than the other subcorpora (f17). This difference is not enough for *forever* to count as key in the PM subcorpus, as it received a p-value of 0.0444. Nevertheless, as the word seems related in function to *never* and *all*, a few instances of *forever* are also included in the following examples that illustrate the bombastic sentiment present in the lyrics:

- 20a) Come follow me to wonderland
 And see the tale that never ends <Power04>
- 20b) But the light will shine forever
 The scares²³ will never leave me <Power18>
- 20c) The breakdown of me
 Must never occur <Power34>
- 20d) The dawn will never rise again for my eyes
 And I will never sleep again <Power38>
- 20e) I'm all alone
 And all is gone I ever lived for <Power32>
- 20f) For all eternity, for all the world to see
 We're riding high across the wind
 And when the storm begins, to wash away all sins <Power16>
- 20g) All of my life I've been waiting for someone,
 To share all my kingdom with me. <Power05>
- 20h) On the ride I'll cross the line
 Forever to be free <Power16>

²³ This is probably a typing error. I think it should be *scars*.

20i) Through the Broken Spectre rose a luring Angel
Sister of Morgana, forever dressed in black <Power14>

As 20a)-i) illustrate, power metal lyrics have a tendency for certain melodramaticism. The lyrics make use of words (*never*, *all* and to some extent arguably *forever*) with extreme or all-encompassing meanings to achieve this effect.

When it comes to thrash metal, the data suggests that profanity is a distinctive feature of the subgenre since *fuck* and *shit* were found as TM keywords and *fuck* and *fucking* as keykeywords. Indeed, further searches using related word forms as search terms revealed that while *fuck* does not occur at all in the other corpora, *fuck*, *fuckin*, *fucking*, and *motherfucking* have a total of 28 instances in the TM subcorpus. In some instances, the use of profanity is found in connection to criticism or disillusionment towards society:

21a) Our generation
Can be the fucking one
That overcomes the greed
Of corrupt nations
[...]
There's something fucking wrong
When war takes sons and daughters
Our lambs misled to slaughter <Thrash20>

21b) All You Stupid Fucking People
With Your Useless Fucking Lives
And Your Bullshit Priorities
And Your Sick Fucking Wars
And Your Pathetic Religion
And Your Shit Music
And Your Sick Fucking Hate
Bullshit, Sitcom, North American Ignorance <Thrash33>

- 21c) Before you go taking a walk in my world,
 ...you better take a look at the real world
 Cause this ain't no Mr. Rogers Neighborhood
 [...]
 I ain't happy about it, but I'd rather feel like shit
 ...than be full of shit!
 [...]
 But here's my apology and one more thing...Fuck you! <Thrash34>
- 21d) Kept under the feet of tyrants
 [...]
 Treated like shit
 Pushed aside and expendable <Thrash06>

This critical attitude towards society coincides with Weinstein's Chaotic theme, which includes resistance, injustice and rebellion. As was pointed out in 2.4, she considers thrash metal lyrics to focus on Chaotic imagery. There are also instances of profanity in contexts that relate to falling out of society and just letting go: *Console yourself, you're better alone / Destroy yourself, see who gives a fuck / Absorb yourself, you're better alone* <Thrash19>; *One Month And I'm Gone Forever / Fuck This Once And For All...* <Thrash33>.

There are 71 instances of *death* in the TM subcorpus. It occurs in 17 texts, of which in four it occurs as KW. In 16 instances the meaning is metaphorical, e.g. *Metal takes hold death starts to unfold / It's loud like the worlds at an end* in <Thrash13>. In <Thrash39> the meaning is ambiguous: *Looking in from the outside and tightening your grip / Death sickening laughter that's beginning to rip*. In most cases (77%), *death* is used in its literal sense:

- 22a) Heat waves and blizzards
 Global death's the cost <Thrash18>
- 22b) Slow death, immense decay
 Showers that cleanse you of your life <Thrash30>
- 22c) Sentenced to death
 With no hope in sight <Thrash24>

- 22d) Live for death, kill for life
 Insanity brings him down <Thrash12>

The word forms *you* and *your* were identified as KKW in the TM subcorpus, while *you* is also a KW. This could be due to a coincidence in sampling, since in three songs, *you* gets repeated very much. That is, although *you* is present in 28 texts in Thrash Metal, the instances of *you* in three songs, <Thrash14, 25 and 34>, amount to 102, which is almost a half of the total of 205 hits. Had the sample not included these songs, *you* would have not risen as key. The word is obviously key in these three songs but is this indicative of the keyness of the word in the whole subgenre? For instance, the repetition present in the chorus of <Thrash25> seems more likely to be an idiosyncratic feature than a genre specific feature:

- 23) 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>

The KKW and KW *can* is also excessively repeated in <Thrash25> and <Thrash34>, with these two songs covering 66 of the total of 93 instances. Since the instances of the two KKW are so concentrated, their status as keywords of the subgenre is suspect.

The KKW of Traditional Heavy are quite indicative of style. 18 out of the 38 KKW are function words and exclamations. The KKW are also shorter than those of the other subgenres, which partly explains why the TRAD subcorpus has the lowest value in mean word length; short words are repeated a lot. Traditional Heavy is the only subcorpus in which exclamations are key. The word *yeah* (which is both KW and KKW) and the KKW *hey* are exclusively key in the TRAD subcorpus. This suggests that exclamations are a stylistic convention typical of the subgenre. In addition, the appearance of the contracted word forms *ain*, *don*, *gonna*, *gotta*, *m*, *re* and *t* – of which all except *don* and *m* are also KWs – as KKW supports the interpretation of Traditional Heavy lyrics being stylistically colloquial.

Several KWs related to the rock 'n' roll roots of heavy metal were also identified as KKWs: *baby, girl, love, roll* and *shake*. The (K)KW *shake* refers to moving the body to the beat of music in 3 out of 4 songs it appears in: *Hey, hey, mama, said the way you move, gonna make you sweat, gonna make you groove. / Oh, oh, child, way you shake that thing, gonna make you burn, gonna make you sting* <Heavy25>; *On your feet you feel the beat, it goes straight to your spine, / Shake your head you must be dead if it don't make you fly* <Heavy27>; *Pull your head back / Hold your hands high / Shake your body*.

5.3 Keywords of Individual Song Texts

Based on the data in the present study, the keywords of individual song texts are better indicators of aboutness than the keywords or keykeywords identified for whole subcorpora. This is hardly surprising as themes of individual songs vary a great deal. Most keywords of songs were lexical words as can be seen in Table 13, in which are listed the keywords of some Black Metal texts that received a high number of KWs. Most of the songs have a list completely or almost completely unlike the lists of the other songs in terms of particular word forms. The only word form present in more than one list in the sample in Table 13 is *closer*.

If we consider the aboutness indicated by the keywords, we can see that by looking at the keywords in Table 13, one can hardly form an idea of a coherent whole of any of the song texts based on the KWs alone. The KWs are decontextualized and their order does not represent song structure. What they do is that they show which words are used more frequently in these songs in comparison to the lyrics of the songs in the other subcorpora. That is, relevant words for the narrative or sentiment of a particular song are not necessarily present in its keyword list, as the lists only show the words that occur statistically frequently enough in these texts in comparison to the reference corpus consisting of the other subcorpora.

Black 20 GRAVEWORM - As The Angels Reach The Beauty - A <i>Dreaming Beauty</i>	Black21 IMMORTAL - At The Heart Of Winter- <i>Withstand The Fall of Time</i>	Black22 KEEP OF KALESSIN - Armada - <i>Crown Of The Kings</i>	Black24 MARDUK - Rom 5:12 - <i>The Levelling Dust</i>	Black26 MELECHESH - Emissaries - <i>Rebirth Of The Nemesis</i>
AWAITING	BITTERNESS	ARMADA	DARKEN	AQUA
BEAUTY	BLACKER	<i>BEFORE</i>	DROOP	ARROGANT
BLOODRED	CENTURIES	BRINGS	DWELLING	CONJECTURES
DEVOURING	<i>CLOSER</i>	CLAW	ECHO	CREATOR
ECHOS	COLDER	<i>CLOSER</i>	FATES	DEITY
GRASPING	DARKER	CROWN	LINGERS	ELISH
KISS	HARDENING	CROWNED	MOSSY	ENUMA
LAKE	LAY	GIVEN	OAKEN	ENUMMA
MELODIES	THUNDRA	KINGS	<i>OF</i>	LINKING
ROSES	UNTOUCHED	RACE	PLAGUE	LIQUID
SHADOWS		REPTILE	REMNANT	LLA
SINGING		SPOKEN	SCOURGE	<i>NEMESIS</i>
SPHERES		UNDEFEATABLE	STREAMS	POLARITIES
TORPID		VICTOR	SUNSET	PRIDE
		YE	<i>THE</i>	REBIRTH
			THICKENS	<i>RETURN</i>
			UNDER	RETURNING
				RE-WRITTEN
				<i>SEALED</i>
				STONE
				TCHAF
				TIAMAT

Table 13. The keyword lists of five black metal songs with many keywords. KKWs *italicized* and in bold.

When the keywords of songs in Table 13 are compared to song titles and album titles we can see a pattern of correspondence, which helps interpret the aboutness of the songs. In <Black20>, the KW *beauty* also occurs in the song and album title. In <Black22>, all the lexical words in the album title and song title are also found as KWs of the song: *armada*, *crown* and *kings*. In <Black26>, the lexical words of the song title, *rebirth* and *nemesis*, occur as KWs of the song. A semantic correspondence is also visible. For instance, <Black21> has the KWs *thundra* [sic], *colder*, *darker* and *blacker* which can all be seen as semantically connected to album title *Heart of Winter*, as the

title refers to a time period that is particularly cold and dark. Similarly, the KW *centuries* is obviously within the same semantic domain as the word *time*, which is present in the song title. Consulting the context of the song confirms that these words are indicative of aboutness in <Black21>: *Hardening claws of cold tell me we are closer to colder times / through the days of bitterness still the thundra lay untouched / In the final of that day and all centuries after.* Interestingly *darker* appears as a KW in the song text, while *dark* is a KW for the BM subcorpus. Because the METAL corpus has not been lemmatized, it cannot be investigated whether it is only the particular word form *dark* that is key in Black Metal, or whether the lemma DARK with its related forms, such as *darker* and *darkest*, is also key.

Several song texts in each subcorpus received only few keywords. Curiously enough, the KWs in texts that only have few of them can actually be quite indicative of the aboutness of the songs. Consider the KWs in Table 14.

Heavy18 HALFORD - Resurrection - <i>Resurrection</i>	Heavy19 HEAVEN & HELL - The Devil You Know - <i>Atom & Evil</i>	Heavy21 IRON MAIDEN - The Number Of The Beast - <i>Invaders</i>	Heavy22 JUDAS PRIEST - Painkiller - <i>Painkiller</i>	Heavy23 KING DIAMOND - Abigail - <i>Arrival</i>
BRING	ATOM	<i>BATTLE</i>	BOILING	# ^α
JUDAS	CRY	<i>COMING</i>	BULLET	COACH
LIFT	<i>EVIL</i>	FIGHT	FASTER	<i>HILL</i>
<i>ME</i>	SPIDER	INVADERS	PAINKILLER	HORSEMEN
RESURRECTION			THAN	JONATHAN
SAINTS			WHEELS	MANSION
WHOLE				

Table 14. The keyword lists of five traditional heavy songs with few keywords. KKWs *italicized* and in bold. α: # refers to numerals.

In 4 of 5 texts in Table 14, the lexical words in the title of the song occur as KWs. It is reasonable to assume that the title of a song has to do with the content of the lyrics of a song so these KWs can be seen as indicators of aboutness. Indeed in three of the songs, <Heavy19, 21 and 22> the song title is clearly used an evocative word in the chorus. Consider, for instance, the chorus of <Heavy22>: *He is the Painkiller / This is the Painkiller* or the first chorus of <Heavy21>: *Invaders ... Pillaging / Invaders ... Looting.* In <Heavy23>, all the keywords are relevant to the narrative of the song. *Jonathan* is the protagonist who is making his way in a *coach* towards a *mansion* on the *hill*, where *horsemen* wait for him.

The keywords of individual songs do not necessarily indicate words important for a whole genre but they can be useful in investigating a song's evocative words or important elements for the theme of a song.

5.4 Utilizing the POS-Tagger and the Tagged METAL Corpus

This section discusses the success rate of the tagging process as well as using the tagged METAL corpus for investigating a few grammatical features that were observed to occur in differing amounts between the subcorpora. The use of the tagged corpus remains somewhat limited in this study due to restrictions of space and time.

An issue that affected the tagging results is the format of the texts. The lyrics downloaded from the internet do not conform to a uniform formatting style, although one feature is strikingly common and present in almost all the lyrics: lines do not end in punctuation marks. With regard to tagging, this means that when no punctuation was present, all the lyrics of a song were treated as one long sentence by the tagger. This yielded tagged texts, which are quite difficult to read, as can be seen when some lines of <Black32> in 24a) are compared to their tagged equivalents in 24b):

24a) A Future benighted still they are blind
 Pigeonhearted beings of flesh and blood
 Keeps closing their eyes for the dangers that threat... ourselves and our nature
 And that is why
 They all enrage me

 Sometimes in the dead of the night I mesmerize my soul
 Sights and visions prophecies and horror
 They all come in one

 Mother north - united we stand (together we walk)
 Phantom north - I'll be there when you hunt them down

 <Black32>

24b) A_AT1 Future_JJ benighted_JJ still_RR they_PPHS2 are_VBR blind_JJ
 Pigeonhearted_JJ beings_NN2 of_IO flesh_NN1 and_CC blood_NN1
 Keeps_VVZ
 closing_VVG their_APPGE eyes_NN2 for_IF the_AT dangers_NN2 that_CST
 threat_NN1
 ..._... ourselves_PPX2 and_CC our_APPGE nature_NN1 And_CC that_DD1
 is_VBZ why_RRQ They_PPHS2 all_DB enrage_VV0 me_PPIO1 Sometimes_RT
 in_II the_AT dead_JJ of_IO the_AT night_NNT1 I_PPIS1 mesmerize_VV0
 my_APPGE soul_NN1 Sights_NN2 and_CC visions_NN2 prophecies_NN2 and_CC
 horror_NN1 They_PPHS2 all_DB come_VV0 in_II one_MC1 Mother_NN1
 north_ND1 -- united_VVD we_PPIS2 stand_VV0 (_(together_RL we_PPIS2
 walk_VV0)_) Phantom_JJ north_ND1 -- I_PPIS1 'll_VM be_VBI there_RL
 when_CS you_PPY hunt_VV0 them_PPHO2 down_RP
 <Tagged Black 32>

As can be seen in 24b), only one sentence break is recognized by the tagger, the three dots in between a verse. I suspect that in some cases this lack of sentence breaks increased the number of mistakes committed by the tagger as the software could not distinguish which words belong to which sentences. The mistakes present in 24b), I assume, are due to other reasons than lack of punctuation however. The reverse word order in the first line causes *Future* to be tagged as an adjective (as _JJ. See Appendix 2 for the tagset). When the first line is inserted in the tagger as *A benighted future* instead of *A Future benighted*, *Future* gets tagged correctly as a noun. That *dead* gets tagged as an adjective in the phrase *the dead of the night* is probably due to *dead* typically occurring as an adjective, not a noun as it does in this case. An interesting case is that *Mother* in *Mother north* is tagged as a noun but *Phantom* in *Phantom north* as an adjective. As these look like parallel structures it would seem arguable that the grammatical structure would similarly be interpreted as parallel. However this case would not straightforward even for a linguist manually doing the tagging, since the status of *Phantom* in the phrase *Phantom north* is somewhat ambiguous.

To check the accuracy of the tagging, a sample of ten tagged files was chosen from the subgenres (two files from each, numbers 11 and 31). The accuracy or the correctness of the tagging was higher than what I expected as the mean mistake percentage turned out to be 5.3 percent, with percentages ranging from 0.4 to 15.8 percent (Table 15), resulting in a mean accuracy of 94.7 percent.

File/Song	Mistakes per total # of words	Mistake %
<Black11>	8/196	4,1
<Black31>	7/194	4,6
<Death11>	19/120	15,8
<Death31>	6/110	5,5
<Power11>	4/136	2,9
<Power31>	1/256	0,4
<Heavy11>	9/176	5,1
<Heavy31>	3/129	2,3
<Thrash11>	27/245	11,0
<Thrash31>	7/160	4,4
All 10 files	91/1722	5,3

Table 15. The accuracy of the sample of POS-tagged texts checked for mistake percentage.

According to Helmut, “[t]he state-of-the-art accuracy in POS tagging is between 95% and 98%” (2008:541). In comparison to that, the accuracy of the tagging of the METAL corpus based on the 10 files is quite good. However, if there are proportionally more files with a high mistake percentage (as in <Death11> and <Thrash11>) in the whole corpus than in the sample, the automatic tagging is not as reliable as this small sample indicates.

<Death11 > – in the tagging of which most mistakes were made by the software – proved to be a challenge for both the tagger and the researcher. It is a prime example of the type of lyrics that I expected would render the accuracy of the tagging relatively low (and that I expected there would be more of in the corpus). The high error rate is probably due to there being no finite verbs whatsoever in the lyrics of <Death11>, in addition to which there is an atypically large number of rare words. An excerpt of the lyrics can be seen in 25):

25) Desensitized - to perspicuous horror
Dehumanized - fresh cannon fodder...

Meritorious horror
Perspicuous onslaught
Dehumanized - cannon fodder

Killing sanitized
Slaughter sanctified

Desensitized - to genocide
<Death11>

In most cases, words like *Desensitized* and *Dehumanized* were mistakenly tagged as adjectives, when they should have been tagged as past participles. Mostly for this reason, the mistake percentage that <Death11> received was so high (15.8%). In <Thrash11> the high mistake percentage is due to the tagger having analyzed the key phrase repeated in the chorus *Thrash 'Til Death* as a noun + noun + noun sequence, instead of verb + preposition + noun, which it actually is. Because there is no punctuation whatsoever in <Thrash11>, the tagger treats the text as a single sentence and therefore interprets *Thrash* to be a noun. When the phrase *Thrash 'Til Death* is inserted in the tagger separately, *Thrash* is correctly tagged as a verb. Apparently *'Til* is a foreign word to the tagger, which considers the word to be a noun, independent of whether it is written with a lower-case or upper-case t, or with or without the single quotation mark.

5.4.1 *Of*

The tagged METAL corpus was used to investigate the preposition *of* in detail. The appearance of *of* as a keyword in the BM and DM subcorpora and a negative keyword in the TM and TRAD subcorpora suggests to me that the former subcorpora employ a more literary style with lots of noun phrases as opposed to the latter ones. Statistics were consulted to confirm whether this interpretation has any validity (see Table 16).

Corpus	Size	<i>of</i> (f)	% of <i>of</i> of all word tokens in subcorpus	% of all tokens of <i>of</i> in the METAL corpus	noun + <i>of</i> + word (f)	% of noun + <i>of</i> + word of tokens of <i>of</i> in subcorpus	KW
METAL	40,915	1065	2.6	100	768	72.1	-
Black metal	7,167	278	3.9	26.1	216	77.7	YES
Death metal	7,480	253	3.4 ²⁴	23.8	181	71.5*	YES
Power metal	9,146	225	2.5	21.1	161	71.6	NO
Traditional heavy metal	7,852	137	1.7	12.9	94	68.6	NEG.
Thrash metal	8,550	172	2.0	16.2	116	67.4	NEG.

Table 16. Statistics of *of* as a word form and as a part of the sequence noun + *of* + word (noun phrase).

²⁴ After these calculations it was noted that *of* occurs in the form *ov* 14 times in <Death06> and of those 14 instances 5 occur in NPs. This would raise the percentage of the preposition to 3.6% in the DM subcorpus and would lower the percentage of *of* in NPs to 69.7%.

The word *of* counts for 2.6 percent of all the word tokens in the METAL corpus with its frequency of 1065. Black Metal and Death Metal, in which *of* is key, count for 278 + 253 = 531 tokens, that is, 49.9 percent of the tokens of *of* in the corpus. In these two subcorpora *of* is used more often in comparison to the other subcorpora, with *of* counting for 3.9 and 3.4 percent of all tokens respectively, while in the other subcorpora *of* only counts for 1.7-2.5 percent of the tokens. The preposition is used in the construction noun + *of* + word, that is, in a noun phrase, a bit over two times out of three in all the subcorpora, although in the DM and PM subcorpora it is used in that construction slightly more frequently than in the TRAD and TM subcorpora and in the BM subcorpus it is used in that way the most frequently, 77.7 percent of the time. Examples of contexts in which this construction occurs are given in 26), in which the noun phrases including this construction are underlined.

- 26a) Judgement comes, before us lie
The paths of war and you may die <Black19>
- 26b) These are the lands, the lands of my birth
 Soon to be ruins, the ruins of my past <Black29>
- 26c) Inside the wound I hope to find
The essence and presence of you <Death03>
- 26d) Dissection of the light
The rotting of the soul <Death31>
- 26e) By the secret of steel you are blessed <Power22>
- 26f) Infamous butcher
Angel of Death <Thrash30>
- 26g) The world lies in the hands of evil <Heavy05>

The significance of the frequency of the noun phrase for each subcorpus was tested by using the Log Likelihood calculator available at <http://ucrel.lancs.ac.uk/llwizard.html>, comparing the target corpus frequency against the reference corpus frequency. In Black Metal, the phrase received a p-value lower than 0.0001, so it is clearly significant. In Death Metal the p-value is slightly higher: 0.0002. For the keywords in this study such a value is considered too low, as it is below the set

threshold of 0.0001. However, as phrases consist of multiple words and strings of particular words occurring together can be considered as less likely components in a text than single words, this p-value can be deemed enough for arguing that the noun phrase is also significantly more frequent in the Death Metal subcorpus. Conversely, in the TRAD and TM subcorpora the phrase was identified as significantly more infrequent, each receiving a p value lower than 0.0001. In Power Metal the result (which points towards underuse) is not statistically significant. So, in the BM and DM subcorpora in which *of* was identified as a KW, the word string *word + of + noun* can also be said to occur as key. In the TM and TRAD subcorpora in which *of* was identified as a negative KW, the word string can be said to occur as negatively key. This suggests that the frequent use of the noun phrase is indeed characteristic of black and death metal lyrics.

5.4.2 *-ing*

The occurrence of *leading* and *leaving* as KWs and KKWs in the Power Metal subcorpus prompted me to investigate whether *-ing* participles are used significantly more frequently in that subgenre. Using a similar procedure as in the case of *of*, it was found out that not only is the *-ing* participle statistically more frequently used in Power Metal ($p < 0.0001$) it is also used relatively frequently in Thrash Metal ($p = 0.0011$), although the latter result is not as definitive. Conversely, the results tentatively suggest that *-ing* participles are used more infrequently in Black Metal ($p = 0.001$) and Death Metal ($p = 0.0005$). Why the participle is used to such an extent in power metal lyrics and what function it plays in the lyrics are questions that require further investigation to be answered.

6 Discussion

The keywords extracted from each subcorpus are mostly illustrative of stylistic choices in the subgenres. The keywords and keykeywords of each subgenre did not reveal a great deal about the exact themes of songs, as it seems that song themes are too varied: the keykeywords identified had a distribution of 2 in most cases, and the highest distribution of any KKW was 5. Content words identified as key seem to reflect stylistic choices in terms of word choice and connotations. Those words might be seen as important words by songwriters and listeners, that is, as stylistic devices that contribute to the overall atmosphere and/or message of songs. It can be assumed that the topics of the songs in the METAL corpus differ so much that the KWs identified for the subcorpora do not really indicate general aboutness of songs that well. Of course, this cannot be properly verified unless the song texts are more closely scrutinized and analyzed. But if that had to

be done, it would run counter to the usefulness of the keyword method, in my opinion, since the advantage of a method lies in its ability to investigate large amounts of texts without having to read all of it closely.

The keywords of individual songs, on the other hand, were rather good indicators of aboutness. This seems reasonable, since songs typically seem to have one major theme. However, trying to establish the meaning of the lyrics of individual songs by investigating their list of keywords only is conjecture at best. Instead the keywords point to stylistic features related to word choice and can be indicative of the general theme and atmosphere of a song but this has to be verified by consulting the whole song text.

In terms of the merits of the keywords method, I consider the findings related to function words as most important. This is because human readers going through texts do not necessarily pay much attention to the relative abundance or scarcity of function words. The function words identified as keywords reveal quite much about the styles of the subgenres. In the Black Metal subcorpus, several function words indicative of a literary songwriting style were identified as KWs (*as, before, nor, of, the*) and KKWs (*before, of, the*). A similar tendency is visible in the function words that occur as KWs (*art, beyond, of, thee, thy*) and KKWs (*upon, thou, thy*) in the Death Metal subcorpus. In addition, in each subcorpus, sequences of noun + *of* + word occur more commonly than in the other subcorpora. Since dense noun phrases are a typical feature of written texts, I interpret their usage to be linked to an attempt at a literary style.

However, a fellow student pointed out to me that as many black and death metal bands are from non-English speaking countries, especially Scandinavia, the native language of the songwriters could have an effect on the grammatical structures of the lyrics, and thus the prolific use of noun phrases with *of* could be related to that. Certainly the native language of songwriters writing in English has some kind of effect on the lyrics but investigating what those effects are is beyond the scope of this study. In fact, the proportion of song lyrics written by native vs. non-native songwriters was not paid attention to in the present study.

On the other hand, the overall usage of function words in the Black Metal corpus points towards tendencies for a literary style:

- *as* occurs in similes much more often than in the reference corpus, e.g. *We are as a flame born unto darkness* in <Black10>;
- *before* occurs most commonly as a preposition of spatial relationship denoting 'in front of', e.g. *And fade before my eyes* in <Black 40>;
- *nor* obviously pertains to written language, e.g. *the bread of bitterness that ignites neither devotion nor fervour* <Black14>
- *of* and *the* are used often in more complex (in comparison to the other subcorpora) noun phrases, e.g. *The cascading pallor of ghostless feather* in <Black01>;

These literary tendencies might have to do with certain discourse features in the black metal scene observed by Kahn-Harris, such as "misanthropy ... often related to a self-conscious elitism" (2007:40) and "an ideology of romanticism" (2007:41). Writing in a more eloquent style could thus be a conduit through which this elitism is expressed or a way in which the elitism is realized. And if the romanticism in black metal draws from the Romantic era in the 18th and 19th centuries, it would seem reasonable to find influences of the style of that era in black metal song lyrics.

It is more challenging to account for the reasons why death metal lyrics would have similar tendencies towards a literary expression like black metal. It could have to do with the explicit description of the abject that Kahn-Harris discusses, as more detail can be provided with heavier noun phrases, e.g. *An opprobious [sic] crucible of molten human waste* in <Death11>. Then again, there could be numerous other reasons: perhaps there is a kind of elitism present in extreme metal scenes in general, death metal included; perhaps the usage of literary language and contrasting it with explicitly violent imagery (as in example 17) in 5.2) in death metal is another manifestation of transgressive behavior in the scene.

That *death* is a KW in the TM but not in the DM subcorpus suggests to me that the concept of death is probably discussed in more general terms in thrash metal than death metal lyrics and that could be why the word is used there so frequently. Death might already 'come with the territory' in death metal, so maybe that is why the word itself is not used as much. Moreover, death metal lyrics make references to death by other means, using explicit expressions to refer to killing, mutilation and related phenomena so perhaps using the word *death* itself would be redundant.

To me it is somewhat surprising that the results do not support the presence of literary language in Power Metal. My impression of power metal lyrics has been that they often draw from imagery

pertaining to fantasy literature, whose style they attempt to emulate. For instance, the power metal band *Rhapsody of Fire* has written two fantasy stories, *The Emerald Sword Saga* and *The Dark Secret Saga*. Each of the stories is told in the lyrics of the band's albums, with five albums being dedicated to each story. However, this might be exceptional in power metal lyrics in general. Still, my expectation was that there would be more features reminiscent of fantasy literature in power metal lyrics. When it comes to function words and style, the results of the keyword analysis do not support a particularly literary approach to writing lyrics. Instead, some features pertaining to spoken language are found: amongst the KWs are the contracted forms *m* and *ll*. However, the content words identified as KWs suggest that there are at least some aspects in the lyrics that can be seen as having a relation to fantasy literature. Power metal KWs and KKWs such as *force*, *hail*, *heart*, and *ride* were frequently found to occur in contexts related to war, warriors or heroism – themes that are also often present in fantasy literature according to my experience.

In addition, the results support that a sentiment of 'we-ness' is present in power metal, as was seen by inspecting the contexts in which the KKWs *we*, *unite* and *us* appear. This is in accordance with Kahn-Harris' observation that power metal lyrics, which he places in between extreme metal and other metal scenes, "have a tradition of self-celebration that other extreme metal lyrics lack" (2007:122) adding that "[t]he lyrics, together with their triumphalist musical backing, celebrate the 'we' of the scene, all bound together by the 'magic' of metal" (ibid.). Pertaining to the latter quote, it was also found that power metal songs often employ a bombastic style in their lyrics, using words like *never*, *all* and, with a lesser certainty, *forever*, for a dramatic effect.

In Thrash Metal the most striking finding is the presence of curse words, such as *fuck*, *fucking* and *shit* among the KWs and KKWs. This seems to be a stylistic feature characteristic of this subgenre in particular. This probably has to do with the history of the subgenre, as it has taken influences from the punk music scene, in which profanity is frequently used in the lyrics. The attitude in thrash metal seems to be similar to that in punk lyrics: the profanity often occurs in contexts in which society is criticized, as was seen in examples 21a)-d). That thrash metal lyrics use – as suggested by the results – a less literary approach to songwriting, could also be due to this counter culture attitude present in the scene. If eloquently structured language is seen as pertaining to a structured society, perhaps breaking away from such style emphasizes breaking away from society or disagreeing with the norms of the society. Thrash metal is an extreme metal genre, in the end, and according to Kahn-Harris, all extreme metal genres engage in transgressive behavior.

Most KWs and KKWs found in the TM subcorpus pertain to rather sinister imagery: *blood, death, die, fight, hell, murder, nightmare, trapped*. This is in accordance with Weinstein's observations that thrash metal lyrics prefer discussing Chaotic themes rather than Dionysian themes. Indeed, as seen in 5.1 and 5.2, in several cases these words occur in contexts where society is criticized, e.g. *Heat waves and blizzards / Global death's the cost* in <Thrash18> and *Commit cold blooded murder / like Nazis during World War Two* in <Thrash02>.

Similarly to Thrash Metal, and even more so, the Traditional Heavy keywords strongly suggest that colloquial language is frequently used in the lyrics. Among these KWs and KKWs are the exclamations *hey* and *yeah* and the contracted forms *ain, don, gonna, gotta, m, re* and *t*. Since traditional heavy metal has its roots in rock n' roll music, for which rebelling against society was typical at least in the beginning stages of the scene, this kind of language use could be seen as part of this rebellious spirit. Of course, there could be several other factors involved, such as sociolinguistic ones, since many members of traditional heavy metal bands come from rather poor families, as in the case with the members of Black Sabbath and Iron Maiden, for instance.

In addition, several content keywords were identified that can be seen as related to the rock 'n' roll roots of traditional heavy. Among these are: *baby, girl, love, rock, roll* and *shake*. Analyzing the KWs suggests that Weinstein's and Walser's observations on how love and sex are discussed in metal are valid. One of Walser's thematic categories for heavy metal song lyrics is 'love' (see 2.4), which is further divided into 'affirmation' and 'longing'. In the results of the present study, *love* was indeed found in such contexts, e.g. in *I got no love, no love you'd call real* in <Heavy38>. In the few songs in the data that the topic of sex is more or less obviously hinted at, the sentiment is in accordance with Weinstein's findings: sex is something fun and free of commitments. In general, based on this data, traditional heavy lyrics are closer in content to that of rock music and other mainstream music in comparison to the other subgenres.

The POS-tagging conducted using the Free CLAWS WWW Tagger had a comparatively high success rate, as 94.7 percent of the data was tagged correctly based on the sample of 10 song texts. However, this high rate may have been due to chance, as, combined with the small size of the METAL corpus, the sample of the tagged lyrics that was checked was very small. Be that as it may, the tagged corpus proved very useful in investigating grammatical phenomena in the subcorpora.

It allowed studying the patterns of *of* and its occurrence in noun phrases, as well as studying -ing particles.

7 Conclusions

The objectives of this study were to find out how well corpus linguistic methods, the keyword method in particular, can be applied to studying song lyrics and how POS-tagging could be applied to studying song lyrics. Let us revisit the research questions:

1. What does a keyword analysis reveal about the differences between the lyrics of different heavy metal subgenres in terms of lyrical content and style?
2. How well can a part-of-speech tagging procedure be applied to a corpus consisting of song lyrics and can the tagging be useful for further analysis of the song lyrics?

The keyword method as well as the part-of-speech tagging both proved useful and applicable in studying song lyrics. The exception was studying keyword linkages, which did not yield useful information about the data. With regard to question 1, the keywords identified revealed mostly stylistic features characteristic of each subgenre, and to a lesser extent, something of the content or aboutness of the lyrics. The findings related to content were in accordance with earlier, qualitative research, on heavy metal lyrics. New information was provided by the results pertaining the usage of function words and grammatical features, which are phenomena not studied in heavy metal lyrics before. Therefore, a quantitative analysis of song texts using corpus methods can be quite useful, because it can reveal phenomena that would remain unnoticed otherwise, especially when it comes to function words.

With regard to question 2, we have seen that tagged song texts allow investigating stylistic factors of lyrics as using the tagged corpus enables searching for grammatical patterns. However, the format of song lyrics (the lack of punctuation) and the style in which they are written (lines often consist of fragments) posed a challenge for the tagging software. Therefore, for a reliable tagged song lyric corpus to be created, tagged texts should be corrected manually, although this process would be very time-consuming.

The small size of the METAL corpus and a lack of meticulous pre-processing of texts proved to be methodological shortcomings in the present study. Therefore, studying song lyrics using corpora in

the future should be done on a larger scale with a well-designed and accurately pre-processed corpus.

With regard to lexical variance in the subcorpora, we have seen that when Traditional Heavy is compared to Black and Death Metal – as all these three subcorpora are almost equally sized – the numbers suggest that black and death metal lyrics show greater lexical variance than traditional heavy lyrics. Not only are the type-token ratios of these two subcorpora greater, but also the mean word length and the number of long words. The same is also visible when the BM and DM subcorpora are compared to the Power Metal subcorpus and, to some extent, the Thrash Metal subcorpus, the two of which have lower type-token ratios. However, with the bigger size of the PM and TM subcorpora, the comparison of type-token ratios is not as feasible. Therefore, lexical variance should be investigated with a larger corpus whose subcorpora are of equal size in order for the numbers to have more validity.

With regard to the keyword analysis in general, the small size of the METAL corpus decreases the credibility of the results of the present study. This is so especially when the borderline cases where a keyword has a low frequency and distribution are considered. In addition, the subcorpora are of a different size and the correctness of the transcription of the lyrics was not checked so the faithfulness of the texts to the actual lyrics sung in the songs is somewhat questionable. Therefore, most of the interpretations and implications made in this and the previous chapter require further study to be validated. In fact, the present study is best seen as a kind of pilot study, pointing towards features that could be investigated and analyzed more thoroughly.

As for future corpus studies on song lyrics, the corpus used should be significantly bigger. Each subcorpus should consist of a larger sample, at least a few hundred song texts but rather 1,000 texts or more, to increase the validity of statistical calculations and the credibility of claims made based on the results. Moreover, a better, more solid way to choose the central texts representing (sub)genres is needed than the one in the present study, that is, relying on *Metal Storm* because a) the website's role as an authority is questionable (in terms of assigning genre labels for instance) and b) the site does not include enough albums to compile a bigger corpus in a similar manner as in this study.

With regard to pre-processing of the corpus texts, a more thorough inspection of the primary sources would yield more valid results. These processes include proof-reading of the song texts

and proof-listening of the songs with an exact transcription of all the words (also shouts and exclamations etc.), since all repetition of words is crucial for the results of keyword analyses. It seems that in several texts in the present study, the choruses were not written down each and every time, which affected the number of keywords identified. Other factors to take into account include orthographical normalization of words with variant spellings, such as *leavin'* vs *leaving* and idiosyncratic variations, such as *stoopid* for *stupid* and *ov* for *of*. Lemmatization could also provide interesting results.

Studying collocations could also be used to investigate the lyrics further, as this could reveal networks of lexical relations between words in general and between keywords too. A bigger corpus might also reveal patterns of co-keyness. Collocations could also provide pointers towards important semantic fields in a genre.

There are factors that the present study did not take into account. One is that the lyrics are written during a time span that covers decades: the oldest texts are from the early 70s, while the newest are from 2010s. Despite of the over 40 years span, the analysis here was a synchronic one. Taking into account the diachronic aspect could reveal interesting phenomena, such as possible changes in style throughout the years. Another factor is the native language of the songwriters. While many bands that are represented in the METAL corpus come from the UK and the USA, a considerable portion of them are from European countries, the songwriters of which use English as a second language. That is, differences pertaining to style and word choice could be studied from the perspective of comparing native vs. non-native lyrics. When it comes to metal lyrics, such an approach would help determine reasons for stylistic differences between subgenres, such as the usage of literary language in Black Metal and Death Metal. In addition to that, songwriters could be interviewed about their writing styles to find out how big of a role that plays compared to the effect of songwriters' native language.

Since there is such a plethora of aspects to take into account in studying song lyrics, combining corpus linguistic methods with methods of other branches (such as musicology and sociology) would be a preferable. This way, cultural and generic elements can be considered (as they should) in interpreting the results of linguistic analyses of lyrics.

To sum up, both the keyword approach and part-of-speech tagging can be successfully applied to studying song lyrics. For more definitive results, more studies – preferably larger-scale – are needed.

Works Cited

- A-Z Lyrics Universe. N.p., n.d. Web. 22 Nov. 2013.
- Anthony, Laurence. 2011. AntConc 3.2.4w.
- Baroni, Marco, and Stefan Evert. "36. Statistical Methods for corpus exploitation." *Corpus Linguistics: An International Handbook Volume 2*. Berlin, Boston: Mouton de Gruyter, 2009. 777-803. Web. 17 Sept. 2013.
- Baroni, Marco and Stefan Evert. "37. Distributions in text." *Corpus Linguistics: An International Handbook Volume 2*. Berlin, Boston: Mouton de Gruyter, 2009. 803-822. Web. 31 Oct. 2013.
- Bondi, Marina. "Perspectives on Keywords and Keyness: An Introduction." *Keyness in Texts*. Ed. Marina Bondi and Mike Scott. Amsterdam: John Benjamins Pub., 2010. 1-18. Print.
- Bowar, Chad. "Heavy Metal Artists, Styles and Genres." *About.com*. N.p., n.d. Web. 7 Jan. 2014.
- Dark Lyrics. N.p., n.d. Web. 22 Nov. 2013.
- Encyclopedia Metallum: The Metal Archives. N.p., n.d. Web. 22 Nov. 2013
- Transgression: Critical Concepts in Sociology*. Ed. Chris Jenks. Vol. 1. London: Routledge, 2006. Google Books. 288 (back cover). Web. 31 Jan. 2014.
- Fabbri, Franco. "A Theory Of Musical Genres. Two Applications." *Popular Music Perspectives*. Ed. David Horn, Philip Tagg, and Marian Green. Göteborg: International Association for the Study of Popular Music, 1982. Web. 9 Jan. 2014.
- Falk, Johanna. *We Will Rock You: A Diachronic Corpus-based Analysis of Linguistic Features in Rock Lyrics*. BA Thesis. Linnaeus University, 2013. *Diva Portal*. 25 Feb. 2013. Web. 11 Sept. 2013.
- Granger, Sylviane, & Pacquot, Magali. *In search of a general academic vocabulary: A corpus-driven study*. Manuscript in preparation. Université catholique de Louvain: Centre for English Corpus Linguistics, 2010. Web. 14 Febr, 2014.
- "Heavy Metal." *AllMusic*. N.p., n.d. Web. 07 Jan. 2014.
- "Heavy Metal Subgenres." *Wikipedia*. Wikimedia Foundation, 01 Jan. 2014. Web. 07 Jan. 2014.
- Hoey, Michael. "45. Corpus linguistics and word meaning." *Corpus Linguistics: An International Handbook Volume 2*. Berlin, Boston: Mouton de Gruyter, 2009. 972-987. Web. 31 Oct. 2013
- Hunston, Susan. "9. Collection strategies and design decisions." *Corpus Linguistics: An International Handbook Volume 1*. Berlin, Boston: Mouton de Gruyter, 2008. 154-168. Web. 31 Oct. 2013.
- Kahn-Harris, Keith. *Extreme Metal: Music and Culture on the Edge*. Oxford: Berg, 2007. Print.
- Katznelson, Noah, Joseph Gelman, Katrin Lindblom, and Marie Caput. *American Song Lyrics: A Corpus-Based Research Project Featuring Twenty Years in Rock, Pop, Country and Hip-Hop*. 2010. San Francisco State University. Web. 11 Sept. 2013.

- Kelly, Kim. *Microgenres: War Metal – A Primer*. Web. Feb. 21, 2012. Jan. 14, 2014.
- "Metal." *Last.fm*. N.p., 26 Mar. 11. Web. 07 Jan. 2014.
- Metal Storm*. N.p., n.d. Web. 22 Nov. 2013. <<http://metalstorm.net/>>.
- Milizia, Denise, and Mike Scott. "Keywords and Phrases in Political Speeches." *Keyness in Texts*. Ed. Marina Bondi. Amsterdam: John Benjamins Pub., 2010. 127-45. Print.
- Rayson, Paul. "From Key Words to Key Semantic Domains." *International Journal of Corpus Linguistics* 13.4 (2008): 519-49. *Ingentaconnect*. Web. 03 Oct. 2013.
- Rieger, Branimir M. Introduction. *Dionysus in Literature: Essays on Literary Madness*. Bowling Green, OH: Bowling Green State University Popular, 1994. 1-5. *Google Books*. Web. 22 Jan. 2014.
- Saarinen, Eeva. "From War Pigs to Unsung Heroes: The Criticism and Justification of War in Metal Lyrics." MA Thesis. University of Turku, 2013. *Doria*. University of Turku, 12 June 2013. Web. 9 Sept. 2013.
- Scott, Mike. 2014. WordSmith Tools version 6. Liverpool: Lexical Analysis Software.
- Scott, Mike. "WORDSMITH 6.0 FAQs." N.p., n.d. Web. 7 Mar. 2014.
- Scott, Mike, and Christopher Tribble. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: Benjamins, 2006. Print.
- Simpson, Rita, and Dushyanthi Mendis. "A Corpus-Based Study of Idioms in Academic Speech." *Tesol Quarterly* 37.3 (2003): 419-41. *Wiley Online Library*. 4 Jan. 2012. Web. 11 Dec. 2013.
- Storjohann, Petra. "Corpus-driven vs. Corpus-based Approach to the Study of Relational Patterns." *Proceedings from the Corpus Linguistics Conference Series*. United Kingdom, Birmingham. 1st ed. Vol. 1. *The Corpus Linguistics 2005 Conference*. University of Birmingham. Web. 9 Dec. 2013.
- Stubbs, Michael. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford [England]: Blackwell, 2001. Print.
- Träff, Ville. "A Hammer Smashed Face: Violence in Extreme Metal Music, Lyrics and Performance." MA Thesis. University of Helsinki, 2011. Print.
- Tweedie, Fiona J., and R. Harald Baayen. "Measures of Lexical Richness in Perspective." *Computers and Humanities* 32 (1998): 323-51. Web. 11 Feb. 2014.
- Van Bael, Cristophe, Henk Van Den Heuvel, and Helmer Strik. "Investigating Speech Style Specific Pronunciation Variation in Large Spoken Language Corpora." *Proceedings of Interspeech (ICSLP) 2004, Jeju, Korea*. 14 Feb. 2014.
- Voller, Jack G. "The Sublime and the Gothic." *The Literary Gothic*. 18 Jan. 2008. Web. 21 Mar. 2014.
- Walser, Robert. *Running with the Devil: Power, Gender, and Madness in Heavy Metal Music*. Hanover, NH: University of New England, 1993. Print.
- Weinstein, Deena. *Heavy Metal: A Cultural Sociology*. New York: Lexington, 1991. Print.
- Wynne, Martin. "33. Searching and concordancing." *Corpus Linguistics: An International Handbook Volume 1*. Berlin, Boston: Mouton de Gruyter, 2008. 706-737. Web. 31 Oct. 2013.
- Xiao, Richard, and Tony McEnery. Preface. *Aspect in Mandarin Chinese: A Corpus-based Study*. Amsterdam: J. Benjamins Pub., 2004. Vii-iii. Web. 11 Dec. 2013.

Appendix 1

A list of the material used for the corpus in this study. Band names in ALLCAPS, album titles in standard formatting, song titles in *italics*.

Black Metal

- Black01 AGALLOCH - Ashes Against The Grain - *Falling Snow*
- Black02 ANOREXIA NERVOSA - New Obscurantis Order - *Mother Anorexia*
- Black03 ARCTURUS - Aspera Heims Symfonia - *To Thou Who Dwellest In The Night*
- Black04 BATHORY - Hammerheart - *Shores in Flames*
- Black05 BORKNAGAR - Empiricism - *The Genuine Pulse*
- Black06 CALADRAN BROOD - Echoes Of Battle - *City Of Azure Fire*
- Black07 CARACH ANGREN - Lammendam - *A Strange Presence Near The Woods*
- Black08 CATAMENIA - Chaos Born - *Calm Before The Storm*
- Black09 CHTHONIC - Takasago Army - *Legacy Of The Seediq*
- Black10 CRADLE OF FILTH - Dusk And Her Embrace - *Heaven Torn Asunder*
- Black11 DARK FORTRESS - Ylem - *Ylem*
- Black12 DARKTHRONE - A Blaze In The Northern Sky - *Kathaarian Life Code*
- Black13 DEAFHEAVEN - Roads To Judah - *Violet*
- Black14 DEATHSPELL OMEGA - *The Shrine Of Mad Laughter*
- Black15 DEMONAZ - March Of The Norse - *March Of The Norse*
- Black16 DIABOLICAL MASQUERADE - Nightwork - *Rider On The Bonez*
- Black17 DIMMU BORGIR - Enthroned Darkness Triumphant - *Mourning Palace*
- Black18 EMPEROR - Anthems To The Welkin At Dusk - *Alsvartr (The Oath)*
- Black19 ENSLAVED - Axioma Ethica Odini - *Ethica Odini*
- Black20 GRAVEWORM - As The Angels Reach The Beauty - *A Dreaming Beauty*
- Black21 IMMORTAL - At The Heart Of Winter - *Withstand The Fall Of Time*
- Black22 KEEP OF KALESSIN - Armada - *Crown Of The Kings*
- Black23 LIMBONIC ART - In Abhorrence Dementia - *In Abhorrence Dementia*
- Black24 MARDUK - Rom 5:12 - *The Levelling Dust*
- Black25 MAYHEM - De Mysteriis Dom Sathanas - *Funeral Fog*
- Black26 MELECHESH - Emissaries - *Rebirth Of The Nemesis*
- Black27 MOONSPELL - Wolfheart - *Wolfshade (A Werewolf Masquerade)*
- Black28 OLD MAN'S CHILD - In Defiance Of Existence - *Felonies Of The Christian Art*
- Black29 PRIMORDIAL - To The Nameless Dead - *Empire Falls*
- Black30 ROTTING CHRIST - Theogonia - *Nemecic*
- Black31 SAMAEL - Passage - *Rain*
- Black32 SATYRICON - Nemesis Divina - *Mother North*
- Black33 SHADE EMPIRE - Omega Arcane - *Ruins*
- Black34 SIGH - Hangman's Hymn: Musikalische Exequien - *Introitus/Kyrie*
- Black35 SUMMONING - Stronghold - *Long Lost To Where No Pathway Goes*
- Black36 THE KOVENANT - Nexus Polaris - *The Sulphur Feast*
- Black37 WATAIN - Sworn To The Dark - *Legions Of The Black Light*
- Black38 WINDIR - 1184 - *Todeswalzer*
- Black39 WOLVES IN THE THRONE ROOM - Two Hunters - *Behold The Vastness And Sorrow*
- Black40 WOODS OF DESOLATION - Torn Beyond Reason - *The Inevitable End*

Death Metal

- Death01 AMON AMARTH - With Oden On Our Side - *Valhall Awaits Me*
- Death02 AMORPHIS - Tales From The Thousand Lakes - *Black Winter Day*
- Death03 ARSIS - A Celebration Of Guilt - *The Face Of My Innocence*
- Death04 ASPHYX - Last One On Earth - *M.S. Bismark*
- Death05 ATHEIST - Unquestionable Presence - *Mother Man*
- Death06 BEHEMOTH - Demigod - *Sculpting The Throne Of Seth*
- Death07 BE'LAKOR - Stone's Reach - *Venator*
- Death08 BEYOND CREATION - The Aura - *No Request For The Corrupted*
- Death09 BLOODBATH - Nightmares Made Flesh - *Cancer Of The Soul*
- Death10 BOLT THROWER - ...For Victory - *When Glory Beckons*
- Death11 CARCASS - Heartwork - *Carnal Forge*
- Death12 CRYPTOPSY - None So Vile - *Crown Of Horns*
- Death13 CYNIC - Focus - *Veil Of Maya*
- Death14 DECAPITATED - Winds Of Creation - *Winds Of Creation*
- Death15 DEICIDE - Deicide - *Sacrificial Suicide*
- Death16 DISMEMBER - Like An Ever Flowing Stream - *Override Of The Overture*
- Death17 EDGE OF SANITY - Crimson - *Prologue*
- Death18 ENTOMBED - Left Hand Path - *Left Hand Path*
- Death19 EVOKEN - Antithesis Of Light - *Accursed Premonition*
- Death20 GOJIRA - From Mars To Sirius - *Ocean Planet*
- Death21 HOLLENTON - With Vilest Of Worms To Dwell - *Y Draig Goch*
- Death22 IMMOLATION - Close To A World Below - *Higher Coward*
- Death23 IN MOURNING - Shrouded Divine - *The Shrouded Divine*
- Death24 INSOMNIUM - Above The Weeping World - *Mortal Share*
- Death25 KALISIA - Cybion - *Arken Bringer (Union/Construction)*
- Death26 MORBID ANGEL - Altars of Madness - *Immortal Rites*
- Death27 NECROPHAGIST - Epitaph - *Epitaph*
- Death28 NILE - Annihilation Of The Wicked - *Cast Down The Heretic*
- Death29 NOCTURNUS - The Key - *Visions From Beyond The Grave*
- Death30 NOUMENA - Absence - *The End Of The Century*
- Death31 OBITUARY - Cause Of Death - *Infected*
- Death32 OBSCURA - Cosmogogenesis - *Anticosmic Overload*
- Death33 OMNIUM GATHERUM - New World Shadows - *Everfields*
- Death34 PERSEFONE - Core - *Sanctuary: Light And Grief*
- Death35 PESTILENCE - Testimony Of The Ancients - *The Secrecies of Horror*
- Death36 QUO VADIS - Day Into Night - *Absolution (Element Of The Ensemble III)*
- Death37 RAPTURE - Songs For The Withering - *Nameless*
- Death38 SEPTIC FLESH - Communion - *Lovecraft's Death*
- Death39 SUFFOCATION - Pierced From Within - *Pierced From Within*
- Death40 VADER - De Profundis - *Silent Empire*

Power Metal

- Power01 ADAGIO - Underworld - *Next Profundis*
- Power02 ANGRA - Temple Of Shadows - *Spread Your Fire*
- Power03 AVANTASIA - The Metal Opera - *Reach Out For The Light*

Power04 BLIND GUARDIAN - Imaginations From The Other Side - *Imaginations From The Other Side*
Power05 CAIN'S OFFERING - Gather The Faithful - *My Queen Of Winter*
Power06 CHILDREN OF BODOM - Hatebreeder - *Warheart*
Power07 DARK MOOR - The Gates Of Oblivion - *In The Heart Of Stone*
Power08 DRAGONLAND - Astronomy - *Supernova*
Power09 EDGUY - Mandrake - *Tears Of A Mandrake*
Power10 ETERNAL TEARS OF SORROW - Children Of The Dark Waters - *Angel Heart, Raven Heart (Act II: Children Of The Dark Waters)*
Power11 EVERGREY - Recreation Day - *The Great Deceiver*
Power12 FAIRYLAND - Score To A New Beginning - *Across The Endless Sea Part II*
Power13 FALCONER - Falconer - *Upon The Grave Of Guilt*
Power14 FATES WARNING - Awaken The Guardian - *The Sorceress*
Power15 FREEDOM CALL - Eternity - *Metal Invasion*
Power16 GAMMA RAY - Land Of The Free - *Rebellion In A Dreamland*
Power17 HAMMERFALL - Legacy Of Kings - *Heeding The Call*
Power18 HEAVENLY - Virus - *The Dark Memories*
Power19 HELLOWEEN - Keeper Of The Seven Keys Part II - *Eagle Fly Free*
Power20 KAMELOT - The Black Halo - *March Of Mephisto*
Power21 LABYRINTH - Return To Heaven Denied - *Moonlight*
Power22 LOST HORIZON - Awakening The World - *Heart Of Storm*
Power23 LUCA TURILLI - Kings Of The Nordic Twilight - *Black Dragon*
Power24 MASTERPLAN - Masterplan - *Spirit Never Die*
Power25 NIGHTWISH - Oceanborn - *Stargazers*
Power26 NORTHER - Mirror Of Madness - *Blackhearted*
Power27 PAGAN'S MIND - Enigmatic : Calling - *The Celestine Prophecy*
Power28 POWERWOLF - Bible Of The Beast - *Raise Your Fist, Evangelist*
Power29 PRIMAL FEAR - Seven Seals - *Demons and Angels*
Power30 RHAPSODY OF FIRE - Dawn Of Victory - *Lux Triumphans*
Power31 SABATON - The Art Of War - *Ghost Division*
Power32 SERENITY - Fallen Sanctuary - *All Lights Reversed*
Power33 SHAMAN - Ritual - *Here I am*
Power34 SKYFIRE - Esoteric - *Esoteric*
Power35 SONATA ARCTICA - Silence - *Weballergy*
Power36 STRATOVARIUS - Episode - *Father Time*
Power37 SYMPHONY X - Divine Wings Of Tragedy - *Of Sins And Shadows*
Power38 WINTERSUN - Wintersun - *Beyond The Dark Sun*
Power39 WUTHERING HEIGHTS - Far From The Madding Crowd - *The Road Ever Goes On*
Power40 YNGWIE MALMSTEEN - Rising Force - *Now Your Ships Are Burned*

Thrash Metal

Thrash01 ANNIHILATOR - Never Neverland - *The Fun Palace*
Thrash02 ANTHRAX - Among The Living - *Among The Living*
Thrash03 ARTILLERY - By Inheritance - *Khomaniak*
Thrash04 CAVALERA CONSPIRACY - Inflikted - *Inflikted*
Thrash05 CELTIC FROST - To Mega Therion - *Jewel Throne*
Thrash06 CHIMAIRA - Resurrection - *Resurrection*

Thrash07 CORONER - No More Color - *Die By My Hand*
Thrash08 DARK ANGEL - Darkness Descends - *Darkness Descends*
Thrash09 DEATH - The Sound of Perseverance - *Scavenger of Human Sorrow*
Thrash10 DEATH ANGEL - The Ultra-Violence - *Evil Priest*
Thrash11 DESTRUCTION - The Antichrist - *Thrash 'til Death*
Thrash12 EVILE - Enter The Grave - *Enter The Grave*
Thrash13 EXODUS - Bonded By Blood - *Bonded By Blood*
Thrash14 FLOTSAM & JETSAM - Doomsday For The Deceiver - *Hammerhead*
Thrash15 FORBIDDEN - Twisted Into Form - *Infinite*
Thrash16 HAVOK - Burn - *The Root Of Evil*
Thrash17 HEATHEN - The Evolution Of Chaos - *Dying Season*
Thrash18 KREATOR - Coma Of Souls - *When The Sun Burns Red*
Thrash19 LAMB OF GOD - Ashes Of The Wake - *Laid To Rest*
Thrash20 MACHINE HEAD - The Blackening - *Clenching The Fists Of Dissent*
Thrash21 MEGADETH - Rust In Peace - *Holy Wars... The Punishment Due*
Thrash22 MESHUGGAH - Destroy Erase Improve - *Future Breed Machine*
Thrash23 METALLICA - Ride The Lightning - *Fight Fire With Fire*
Thrash24 MUNICIPAL WASTE - Hazardous Mutation - *Deathripper*
Thrash25 NUCLEAR ASSAULT - Game Over - *Sin*
Thrash26 OVERKILL - The Years Of Deacy - *Time To Kill*
Thrash27 PANTERA - Vulgar Display Of Power - *Mouth For War*
Thrash28 SEPULTURA - Chaos A.D. - *Refuse/Resist*
Thrash29 SHADOWS FALL - Retribution - *My Demise*
Thrash30 SLAYER - Reign In Blood - *Angel Of Death*
Thrash31 SODOM - Agent Orange - *Agent Orange*
Thrash32 SOULFLY - Dark Ages - *Babylon*
Thrash33 STRAPPING YOUNG LAD - City - *Velvet Kevorkian*
Thrash34 SUICIDAL TENDENCIES - Lights Camera Revolution - *You Can't Bring Me Down*
Thrash35 TESTAMENT - The Legacy - *Over The Wall*
Thrash36 WARBRINGER - Waking Into Nightmares - *Jackal*
Thrash37 VEKTOR - Black Future - *Black Future*
Thrash38 WHITE ZOMBIE - Astro Creep: 2000 – Songs of Love, Destruction And Other Synthetic Delusions - *Electric Head Pt. 1 (The Agony)*
Thrash39 VIO-LENCE - Eternal Nightmare - *Eternal Nightmare*
Thrash40 VOIVOD - Dimension Hatröss - *Experiment*

Traditional Heavy Metal

Heavy01 ACCEPT - Restless And Wild - *Fast As A Shark*
Heavy02 ALICE COOPER - Hey Stoopid - *Hey Stoopid*
Heavy03 ALICE IN CHAINS - Dirt - *Them Bones*
Heavy04 ANGEL WITCH - Angel Witch - *Angel Witch*
Heavy05 APOCALYPTICA - Apocalyptica - *Life Burns*
Heavy06 BLACK LABEL SOCIETY - The Blessed Hellride - *Stoned And Drunk*
Heavy07 BLACK SABBATH - Paranoid - *War Pigs*
Heavy08 BRUCE DICKINSON - The Chemical Wedding - *King in Crimson*
Heavy09 CRIMSON GLORY - Transcendence - *Lady of Winter*
Heavy10 DEEP PURPLE - Machine Head - *Highway Star*

Heavy11 DEF LEPPARD - Pyromania - *Rock! Rock! Till You Drop*
Heavy12 DIAMOND HEAD - Lightning to the Nations - *Lightning to the Nations*
Heavy13 DIO - Holy Diver - *Stand up and Shout*
Heavy14 DOWN - Nola - *Temptation's Wings*
Heavy15 EUROPE - The Final Countdown - *The Final Countdown*
Heavy16 GHOST - Opus Eponymous - *Ritual*
Heavy17 GRAVE DIGGER - Tunes of War - *Scotland United*
Heavy18 HALFORD - Resurrection - *Resurrection*
Heavy19 HEAVEN & HELL - The Devil You Know - *Atom & Evil*
Heavy20 HELL - Human Remains - *On Earth As It Is In Hell*
Heavy21 IRON MAIDEN - The Number of The Beast - *Invaders*
Heavy22 JUDAS PRIEST - Painkiller - *Painkiller*
Heavy23 KING DIAMOND - Abigail - *Arrival*
Heavy24 KISS - Sonic Boom - *Modern Day Delilah*
Heavy25 LED ZEPPELIN - Led Zeppelin IV - *Black Dog*
Heavy26 MERCENARY - Architect Of Lies - *New Desire*
Heavy27 MOTÖRHEAD - Overkill - *Overkill*
Heavy28 OZZY OSBOURNE - Blizzard Of Ozz - *I Don't Know*
Heavy29 POISONBLACK - Lust Stained Despair - *Nothing Else Remains*
Heavy30 QUEENSRÿCHE - Operation Mindcrime - *Revolution Calling*
Heavy31 RUNNING WILD - Black Hand Inn - *The Curse*
Heavy32 SANCTUARY - Into The Mirror Black - *Future Tense*
Heavy33 SAVATAGE - Hall Of The Mountain King - *24 Hrs. Ago*
Heavy34 SAXON - Strong Arm Of The Law - *Heavy Metal Thunder*
Heavy35 SENTENCED - The Cold White Light - *Cross My Heart And Hope To Die*
Heavy36 SKID ROW - Slave To The Grind - *Monkey Business*
Heavy37 W.A.S.P. - The Crimson Idol - *The Invisible Boy*
Heavy38 VAN HALEN - Van Halen - *Runnin' With The Devil*
Heavy39 VIRGIN STEELE - Invictus - *The Blood of Vengeance*
Heavy40 VOLBEAT - Rock The Rebel/Metal The Devil - *The Human Instrument*

Appendix 2

UCREL CLAWS7 Tagset

Available at: <http://ucrel.lancs.ac.uk/claws7tags.html>

APPGE	possessive pronoun, pre-nominal (e.g. my, your, our)
AT	article (e.g. the, no)
AT1	singular article (e.g. a, an, every)
BCL	before-clause marker (e.g. in order (that), in order (to))
CC	coordinating conjunction (e.g. and, or)
CCB	adversative coordinating conjunction (but)
CS	subordinating conjunction (e.g. if, because, unless, so, for)
CSA	as (as conjunction)
CSN	than (as conjunction)
CST	that (as conjunction)
CSW	whether (as conjunction)
DA	after-determiner or post-determiner capable of pronominal function (e.g. such, former, same)
DA1	singular after-determiner (e.g. little, much)
DA2	plural after-determiner (e.g. few, several, many)
DAR	comparative after-determiner (e.g. more, less, fewer)
DAT	superlative after-determiner (e.g. most, least, fewest)
DB	before determiner or pre-determiner capable of pronominal function (all, half)
DB2	plural before-determiner (both)
DD	determiner (capable of pronominal function) (e.g. any, some)
DD1	singular determiner (e.g. this, that, another)
DD2	plural determiner (these, those)
DDQ	wh-determiner (which, what)
DDQGE	wh-determiner, genitive (whose)
DDQV	wh-ever determiner, (whichever, whatever)
EX	existential there
FO	formula

FU	unclassified word
FW	foreign word
GE	germanic genitive marker - (' or's)
IF	for (as preposition)
II	general preposition
IO	of (as preposition)
IW	with, without (as prepositions)
JJ	general adjective
JJR	general comparative adjective (e.g. older, better, stronger)
JJT	general superlative adjective (e.g. oldest, best, strongest)
JK	catenative adjective (able in be able to, willing in be willing to)
MC	cardinal number, neutral for number (two, three..)
MC1	singular cardinal number (one)
MC2	plural cardinal number (e.g. sixes, sevens)
MCGE	genitive cardinal number, neutral for number (two's, 100's)
MCMC	hyphenated number (40-50, 1770-1827)
MD	ordinal number (e.g. first, second, next, last)
MF	fraction, neutral for number (e.g. quarters, two-thirds)
ND1	singular noun of direction (e.g. north, southeast)
NN	common noun, neutral for number (e.g. sheep, cod, headquarters)
NN1	singular common noun (e.g. book, girl)
NN2	plural common noun (e.g. books, girls)
NNA	following noun of title (e.g. M.A.)
NNB	preceding noun of title (e.g. Mr., Prof.)
NNL1	singular locative noun (e.g. Island, Street)
NNL2	plural locative noun (e.g. Islands, Streets)
NNO	numeral noun, neutral for number (e.g. dozen, hundred)
NNO2	numeral noun, plural (e.g. hundreds, thousands)
NNT1	temporal noun, singular (e.g. day, week, year)
NNT2	temporal noun, plural (e.g. days, weeks, years)
NNU	unit of measurement, neutral for number (e.g. in, cc)

NNU1	singular unit of measurement (e.g. inch, centimetre)
NNU2	plural unit of measurement (e.g. ins., feet)
NP	proper noun, neutral for number (e.g. IBM, Andes)
NP1	singular proper noun (e.g. London, Jane, Frederick)
NP2	plural proper noun (e.g. Browns, Reagans, Koreas)
NPD1	singular weekday noun (e.g. Sunday)
NPD2	plural weekday noun (e.g. Sundays)
NPM1	singular month noun (e.g. October)
NPM2	plural month noun (e.g. Octobers)
PN	indefinite pronoun, neutral for number (none)
PN1	indefinite pronoun, singular (e.g. anyone, everything, nobody, one)
PNQO	objective wh-pronoun (whom)
PNQS	subjective wh-pronoun (who)
PNQV	wh-ever pronoun (whoever)
PNX1	reflexive indefinite pronoun (oneself)
PPGE	nominal possessive personal pronoun (e.g. mine, yours)
PPH1	3rd person sing. neuter personal pronoun (it)
PPHO1	3rd person sing. objective personal pronoun (him, her)
PPHO2	3rd person plural objective personal pronoun (them)
PPHS1	3rd person sing. subjective personal pronoun (he, she)
PPHS2	3rd person plural subjective personal pronoun (they)
PPIO1	1st person sing. objective personal pronoun (me)
PPIO2	1st person plural objective personal pronoun (us)
PPIS1	1st person sing. subjective personal pronoun (I)
PPIS2	1st person plural subjective personal pronoun (we)
PPX1	singular reflexive personal pronoun (e.g. yourself, itself)
PPX2	plural reflexive personal pronoun (e.g. yourselves, themselves)
PPY	2nd person personal pronoun (you)
RA	adverb, after nominal head (e.g. else, galore)
REX	adverb introducing appositional constructions (namely, e.g.)
RG	degree adverb (very, so, too)

RGQ	wh- degree adverb (how)
RGQV	wh-ever degree adverb (however)
RGR	comparative degree adverb (more, less)
RGT	superlative degree adverb (most, least)
RL	locative adverb (e.g. alongside, forward)
RP	prep. adverb, particle (e.g about, in)
RPK	prep. adv., catenative (about in be about to)
RR	general adverb
RRQ	wh- general adverb (where, when, why, how)
RRQV	wh-ever general adverb (wherever, whenever)
RRR	comparative general adverb (e.g. better, longer)
RRT	superlative general adverb (e.g. best, longest)
RT	quasi-nominal adverb of time (e.g. now, tomorrow)
TO	infinitive marker (to)
UH	interjection (e.g. oh, yes, um)
VB0	be, base form (finite i.e. imperative, subjunctive)
VBDR	were
VBDZ	was
VBG	being
VBI	be, infinitive (To be or not... It will be ..)
VBM	am
VBN	been
VBR	are
VBZ	is
VD0	do, base form (finite)
VDD	did
VDG	doing
VDI	do, infinitive (I may do... To do...)
VDN	done
VDZ	does
VH0	have, base form (finite)

VHD	had (past tense)
VHG	having
VHI	have, infinitive
VHN	had (past participle)
VHZ	has
VM	modal auxiliary (can, will, would, etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (e.g. give, work)
VVD	past tense of lexical verb (e.g. gave, worked)
VVG	-ing participle of lexical verb (e.g. giving, working)
VVGK	-ing participle catenative (going in be going to)
VVI	infinitive (e.g. to give... It will work...)
VVN	past participle of lexical verb (e.g. given, worked)
VVNK	past participle catenative (e.g. bound in be bound to)
VVZ	-s form of lexical verb (e.g. gives, works)
XX	not, n't
ZZ1	singular letter of the alphabet (e.g. A,b)
ZZ2	plural letter of the alphabet (e.g. A's, b's)