Barbara Berti

QUANTITATIVE AND QUALITATIVE ANALYSIS OF A MULTI-*GENRE* HEAVY METAL CORPUS: AN NLP-BASED APPROACH[1]

ABSTRACT. Heavy metal is a music *genre* that originated at the end of the 1960s, rose to prominence in the 1980s, and has evolved into numerous *subgenres* ever since. Its prolific nature has made it so diverse in terms of musical discourses, social practices, and cultural meanings that it would be more appropriate to define it as a meta-*genre*. Indeed, owing to its richness and diversity, heavy metal lends itself to being the object of linguistic analysis. The present work aims to investigate the lyrics of 10 *subgenres* of heavy metal using the tools of Natural Language Processings (NLP) applied to an *ad hoc* corpus consisting of 1,091,054 tokens. The results confirm that the umbrella term *heavy metal* encompasses very different styles in terms of lexical richness, themes, and sentiments. It embraces lexically sophisticated *subgenres* such as black metal and death metal, characterised by extensive use of rare words, as well as mainstream *subgenres* such as glam metal, metalcore, and nu metal, based upon simpler words and less diverse vocabularies. The themes cover the whole spectrum of human existence, from life and death to love and pain, and are conveyed through a wealth of different narrative styles, ranging from grindcore's fierce anger to glam metal's glossy tones.
KEYWORDS: Heavy metal. NLP. Sentiment Analysis. LDA.

Introduction

According to the OED, the first music-related attested use of the bigram *heavy metal* dates back to 1964 and was coined by William S. Burroughs. In his novel "Nova Express", Burroughs created a character named Uranian Willy The Heavy Metal Kid. However, *heavy metal* was not used in reference to music. Rather, it was a metaphor for a drug.

A subsequent emblematic usage can be found in the 1968 hit song "Born to be wild" by the Canadian band Steppenwolf: "I like smoke and lightning, heavy metal thunder". Once again, the bigram is used in a non-musical sense, this time as a metaphor for the roar of motorcycles. Yet Steppenwolf's song represents the first occurrence of the term *heavy metal* in music.

One of the first usages of *heavy metal* with reference to a music *genre* is found in a 1973 article from "Crawdaddy" magazine: "They find no comfort in glitter or Heavy Metal – Black Sabbath, Black Oak Arkansas and their ilk". The journalist's confident use of *heavy metal* suggests that the term was somehow familiar to the readership since no explanation is provided.

Indeed, a search in the Corpus of Historical American English (COHA) returns an interesting coeval occurrence of the term:

> "She turns on her boom box so it plays an amateur heavy metal tape, the lyrics in French, and she sings along as she sets out two paper plates, two plastic forks, two plastic knives and so on."

The excerpt comes from a play written by the Canadian playwright Carol Bolt entitled "Rosie learns French". Reference to an "amateur heavy metal tape" suggests that the *genre* had been around for long enough for enthusiasts to play it. This is consonant with Weinstein's findings that, despite few official occurrences in the press, the label *heavy metal* was in the air at the time and: "had been floating around the culturescape, ready to be captured and made a name" (Weinstein, 2014: 48).

In the COHA the collocates of *heavy metal* reflect both senses of the term. Amongst the 10 most frequent, four are chemistry-related (*door* [11 occurrences], *box* [7], *doors* [5], *open* [5], *contamination* [3], and *objects* [3]), whilst only four are music-related (*rock* [6], *music* [5], *kids* [4], and *punk* [3]).

However, if we turn to the Corpus of Contemporary American English (COCA), a contemporary and open corpus comprising documents from 1990 to the

present day, distribution of the two senses is reversed. The retrieved collocates are:

music (164), *band* (76), *rock* (55), *rap* (55), *door* (34), *poisoning* (26), *bands* (22),

drummer (21), and *contamination* (19). Only three refer to chemistry.

In spite of this, the predominant sense of *heavy metal* as a music *genre* is not

reflected in the lexicographic choices of some authoritative English dictionaries,

which tend to provide the literal definition first, as can be seen in Table 1, in which

the figure occurring before the definition indicates the sense number.

Cambridge Dictionary[2]	2. [U] a style of rock music with a strong beat, played very loudly using electric guitars
Collins English Dictionary[3]	1. UNCOUNTABLE NOUN is a type of hard rock characterized by violent, shouted lyrics.
Merriam Webster Dictionary[4]	2. *noun* energetic and highly amplified electronic rock music having a hard beat
Oxford English Dictionary[5]	3. *n.* a type of loud, vigorous rock music characterized by the use of electronically amplified instruments (typically guitar, bass, and drums), a heavy (usually fast) beat, intense or spectacular performance, and often a clashing, harsh musical style. Frequently *attributive* or as *adj*.

Table 1. Definitions of *heavy metal*.

2 https://dictionary.cambridge.org/dictionary/english/heavy-metal
3 https://www.collinsdictionary.com/us/dictionary/english/heavy-metal
4 https://www.merriam-webster.com/dictionary/heavy%20metal
5 https://www.oed.com

The Collins English Dictionary is the only resource that gives prominence to the music sense. It is also the only resource that makes reference to heavy metal lyrics, defining them as "violent" and "shouted", whilst the other resources limit their descriptions to strictly music-related features. The music of heavy metal is described as vigorous, loud, energetic and having a strong beat but not much more is added.

What these definitions fail to capture is the multifaceted nature of heavy metal, which Walser (1993: 2) condenses as follows:

> *"Heavy metal* now denotes a variety of musical discourses, social practices, and cultural meanings, all of which revolve around concepts, images, and experiences of power. The loudness and intensity of heavy metal music visibly empower fans, whose shouting and headbanging testify to the circulation of energy at concerts."

Walser's description focuses precisely upon the element of variety, suggesting that heavy metal has surpassed the boundaries of a monolithic music *genre*, becoming so diverse that it is nowadays best defined as a meta-*genre* (Weinstein, 2014). Indeed, throughout the course of its history heavy metal has undergone a process of diversification and originated numerous *subgenres*. Traditionally, the label *heavy metal* is used with two different meanings. On the one hand, *heavy*

metal is an umbrella term, including all the *subgenres* that developed out of the first forms of heavy metal in the 1970s; on the other, *heavy metal* indicates the first wave of heavy metal, before it began its process of fragmentation.

Owing to its richness and diversity, heavy metal lends itself to being the object of linguistic analysis. The present work examines the lyrics of 10 *subgenres* of heavy metal with the tools of Natural Language Processing (NLP) applied to an *ad hoc* compiled corpus.

Corpus compilation and pre-processing

In order to compile a corpus that would represent the heavy metal *genre*, we chose to divide it into its 10 most prominent *subgenres*. Thus the Heavy Metal Corpus (HMC) is made up of the following sections: black metal (BM); death metal (DM); glam metal (GM); grindcore (GC); heavy metal (HM); metalcore (MC); nu metal (NU); power metal (PM); progressive metal (PrM), and thrash metal (TM).

The documents that constitute the HMC were downloaded from two freely accessible and specialised websites: *darklyrics.com* and *metalkingdom.com*. Both

websites are collections of heavy metal music lyrics. The latter offers a categorisation of various albums by musical *subgenre*, which was particularly useful for the purpose of this work. In fact, many heavy metal albums are a mixture of different *subgenres* and would not be good candidates for inclusion in the HMC since the aim of the present work is to capture quintessential differences amongst *subgenres*. For this reason, only the albums that were assigned to one *genre* label were included.

For each *subgenre*, we downloaded a sufficient number of songs to reach 100,000 tokens approximately. The HMC consists of 1,091,054 tokens and 5,409 songs in total (Table 2).

All the lyrics are in English. Only a few songs showcase verses written in other languages. As for pre-processing, the major transcription errors were corrected (e.g. *goins* → *going*), symbols were removed, and the word *chorus,* used as a placeholder to avoid repeating the refrain, was replaced with the actual verses.

	Number of tokens	Number of songs
Black metal	107436	553
Death metal	108981	601

Glam metal	110433	460
Grindcore	109800	1185
Heavy metal	109494	485
Metalcore	110027	339
Nu metal	107052	341
Power metal	108501	493
Progressive	109981	441
Thrash	109349	511
Total	1091054	5409

Table 2. Composition of the HMC.

Methods

The methods employed to extract and process information are based on Natural Language Processing (Manning & Schütze 1999) and Text Mining (Bird *et al*. 2009; Dale *et al*. 2000; Feldman, Sanger 2007; Jurafsky, Martin 2008; Mertz 2003).

The corpora were analysed through a Python 3.6 script written by the author. In particular, the following libraries were used:

- LDA for topic modelling

- LexicalRichness to measure lexical diversity
- Matplotlib for data visualisation
- NLTK for text-processing, tokenisation, lemmatisation, collocation extraction
- NumPy for computing mean and standard deviation
- Spacy for Part-of-Speech tagging
- Vader for Sentiment Analysis

Analysis

Lexical richness

As can be observed in Table 2, the number of songs in the different *subgenres* varies from a minimum of 339 (in the MC corpus) to a maximum of 1185 (in the GC corpus), in spite of being of very similar size. Turning to the mean length of the lyrics, we can observe the effect of each *subgenre* on song length. We computed the mean length for each *subgenre* and created a new global distribution of means. The global mean of such distribution is 235.32 tokens with a standard deviation of 72.79.

Figure 1 shows the mean length of the lyrics in each *subgenre*. Black metal (194.27 tokens), death metal (181.33 tokens), glam metal (240.07 tokens), heavy metal (225.76 tokens), power metal (220.08 tokens), prog metal (249.39 tokens),

and thrash metal (213.99 tokens) do not deviate substantially from the mean. However, three *subgenres* stand out. Metalcore and nu metal are made up of much longer songs, 324.56 and 313.93 tokens respectively, whilst grindcore lyrics are significantly below average with as few as 92.65 tokens. Indeed, grindcore is a *subgenre* characterised by extremely short as well as harsh and non-verbose lyrics.
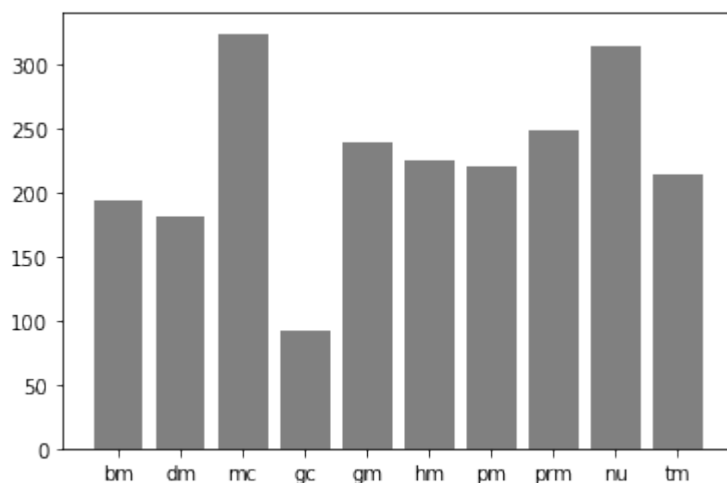

Fig. 1 Mean length of lyrics in each *subgenre*.

Measuring word length can provide insights into the complexity of a text. Generally, simpler texts are made up of shorter words, whilst longer words indicate a more complex text. We calculated the mean length of words for each subgenre as

well as the mean length of content words only. The results are summarised in Table 3.

	Mean length of words	Mean length of content words
Black metal	4.12	5.10
Death metal	4.24	5.23
Glam metal	3.48	4.17
Grindcore	4.14	5.07
Heavy metal	3.65	4.38
Metalcore	3.58	4.40
Nu metal	3.57	4.42
Power metal	3.82	4.72
Progressive	3.79	4.70
Thrash	3.88	4.69

Table 3. Mean length of words and content words.

The *subgenres* that display more complex words are black metal, death metal, and grindcore, whilst glam metal, heavy metal, and metalcore appear to be constituted by simpler words.

In order to assess lexical diversity, we calculated the Type-Token Ratio (TTR) for every song in each *subgenre* and computed the average so as to obtain a comprehensive TTR score for each *subgenre* (Table 4). The *subgenres* showing the highest TTR values are grindcore (0.69), death metal (0.62), and black metal (0.61). Thrash metal, prog metal, and power metal follow with 0.57, 0.53, and 0.53 respectively. At the lower end of the scale are heavy metal (0.48), nu metal (0.47), glam (0.44), and metalcore (0.4).

Since text length is a determining factor in TTR, we opted to utilise a measure that overcomes length issues, i.e. Moving Average Type-Token Ratio (MATTR). The results accord with those yielded by the TTR. The ranking remains mainly unchanged with the exception of grindcore texts which now rank third instead of first. In all likelihood, this is due to the extreme brevity of such texts which could have inflated the TTR value slightly. In general, we can claim that death metal, black metal, and grindcore lyrics tend to be richer from a terminological point of view, whilst glam metal and nu metal appear to be less diverse.

A further parameter for evaluating the richness of a text is lexical density. This was computed by dividing the number of content words by the total number of

tokens. The results are shown in Table 4. The *subgenres* with the highest incidence of content words are death metal (60.07%), grindcore (58.82%), and black metal (58.43%). At the other end of the spectrum are metalcore (51.52%) and nu metal (51.61%).

	TTR	MATTR	LD (%)	Hapax (%)	Out-of-vocabulary/extremely rare words (%)
Black metal	0.61	0.77	58.43	4.88	1.26
Death metal	0.62	0.78	60.07	4.99	1.21
Glam metal	0.44	0.68	54.94	1.7	0.35
Grindcore	0.69	0.76	58.82	4.75	1.19
Heavy metal	0.48	0.70	54.85	2.27	0.38
Metalcore	0.40	0.70	51.52	1.62	0.22
Nu metal	0.47	0.67	51.61	2.53	0.66
Power metal	0.53	0.75	54.09	2.28	0.48
Progressive	0.53	0.75	53.62	3.14	0.64
Thrash	0.57	0.76	57.94	3.27	0.66

Table 4. Statistics on lexical richness.

In order to explore lexical richness further, we examined the number of hapax legomena in each *subgenre* (Table 4). Death metal, grindcore, and black metal present the highest percentage of uniquely occurring words (4.99%, 4.88%, and 4.75% respectively), whilst glam metal shows an extremely low incidence (1.7%).

We moved on to examine the presence of extremely rare terms. In order to do so, we lemmatised the words in our corpora and compared them with a list of 97,565 English lemmas extracted from Google Books and ordered by frequency. Any word not included in the list is either extremely rare or does not exist in the English language.[6] The results (Table 4) show that BM and DM corpora have the highest concentration of very rare or out-of-vocabulary terms, whilst in the MC and GM corpora this phenomenon is almost irrelevant. In the BM corpus, uncommon terms include, amongst others, Middle English words (e.g. *beholde, thou, shalt, dwelth*), northern/Gothic mythology (e.g. *Claunek, Yggdrasil*), words related to the spiritual/religious/magic dimension (e.g. *latria*, *Tipheret*, *Tarchimache, Necronomicon, Zagreus*), foreign terms (e.g. *sjeler* from Norwegian, *kalde* from

---

[6] Out-of-vocabulary words can be neologisms as well as misspelled words or foreign words.

Danish, *vinter,* the Nordic word for *winter*). In the DM corpus, unusual terms are mainly related to divinities (e.g. *Azathoth, Nigurrath, Alithea)* as well as to the semantic area of science (e.g. *nanomass, multifractal, nanotech*) and medicine *(*e.g. *pneumonectomy, oximeter, integumental*).

Topic extraction[7] and word frequency

Content words provide indications of the most relevant semantic areas in a series of texts. We extracted the most frequent content words for each *subgenre* and compared them. The word *time* appears amongst the 10 most frequent words in each corpus (Table 5) thus it can be said to be a *trait d'union* connecting the different *subgenres*. We turned to trigram collocations in order to gauge whether use of the noun *time* was homogeneous across different corpora. Generally, *time* is used to express urgency and immediacy (e.g. *the time has, time has come, it is time, is time to, time is now*), or to refer to past events (e.g. *long time ago, long time gone, a long time*). In the DM corpus, other usages appear which accord with the gloomy, dark, and ineluctable atmosphere of the *subgenre*. We find, amongst

---

[7] In order to extract topics, LDA (Blei et al. 2003) was used. LDA is a generative statistical model that automatically identifies hidden topics in a corpus.

others, expressions such as *time to die, time heals nothing, worse each time, end of time, time is black. Night time* is common both in the GM and in the HM corpus, although it is used positively in the former, i.e. *a good time.*

The noun *life* occurs in most corpora. Its occurrence is particularly interesting in the DM and TM corpora (it is the most frequently occurring content word in the latter). The concordances, however, show that the term is often employed only to recall its antonym *death* (e.g. "I took your life from you", "bodies void of life", "abolish life"), which is the third and second most frequent word in DM and TM respectively, or to vent one's frustration or anger, as in "my plague is my life", "boundless hate for this life". Death is indeed a very relevant topic in the DM corpus, appearing in three out of five topics identified by the LDA algorithm. *Death* is often associated with *flesh* and *soul*, thus pointing to two distinct usages of the term, one related to the physical dimension, taking the shape of crude narrative realism, the other transcending the physical sphere to embrace the spiritual dimension of death.

In the TM corpus, the themes of death and dying and the juxtaposition between life and death are found in all the extracted topics, often in combination

with *evil, hell*, and *god*. The most relevant collocations concerning death are *time to die, gonna die, wanna die, prepare to die, wish to die, smell of death, king of death, fear of death, dying time*. There are only 13 instances of the third person *dies* in the TM corpus, 10 referring to inanimate objects (e.g. *winter, dust, nature*), and only three to a human being. Thus death is generally wished upon either the subject or an enemy. The themes of suffering, terror, and devastation, typical of thrash metal, emerge from the presence of words such as *hell, evil, kill, pain*, and *blood*.

Blood is the most frequent word in the BM corpus and the second most frequent in the DM corpus. Its presence points to the ruthless nature of black metal and death metal lyrics. In the BM corpus, the most relevant adjectival collocates of *blood* are *warm* and *innocent,* whilst in the DM corpus they are *stagnant* and *human*. As for the verbal collocates, whilst *run* is common in both corpora, *flow* is more characteristic of BM, and *drain* of DM. In the BM the prevailing sense is sight as the high occurrence of the verb *see* and the relevance of the noun *eye* in topic modelling reveal. But topic modelling also suggests that the esoteric or spiritual dimensions are relevant, as attested by the presence of words such as *soul, Satan, and Hell*.

The GC corpus stands out for the presence of foul language. *Fucking* is the most frequent word, *fuck* and *shit* rank sixth and eighth respectively. This reveals the extremely crude nature of grindcore lyrics, which aim at the ruthless and merciless denigration and criticism of social conventions and religion. Such denigration and criticism are expressed through bursts of violent anger, as can be observed in the expressions "fuck life", "fuck reality", and "fuck your gods". In more extreme cases, the lyrics become so exaggerated as to embrace cannibalism, necrophagia, and necrophilia, as is the case of "corpses to dig up and fuck" or "I live to fuck the dead". It should be pointed out, however, that, contrary to other *genres* in which the lyrics express a truthful point of view on the world and on society, in extreme grindcore recourse to such violent and disturbing images is purely fictional, stemming from the same fascination for violence and gore on which splatter cinema is based. Indeed, many grindcore musicians have openly declared their passion for splatter films.

Black	Blood (335), death (330), black (291), life (260), night (255), soul (242), world (234), time (229), see (223), god (218)
Death	Life (444), blood (345), death (340), dead (340), flesh (278), god (254), time (253), see (250), die (246), world (244)

Glam	Love (818), got (759), know (483), night (435), like (410), never (364), time (363), take (337), baby (336), get (334)
Grindcore	Fucking (522), life (373), like (317), see (257), time (251), fuck (244), death (240), shit (206), world (199), nothing (195)
Heavy	Know (451), time (445), see (402), got (369), like (358), never (344), come (328), night (326), life (316), away (316)
Metalcore	Never (557), know (467), let (384), away (381), see (367), time (350), like (341), way (303), take (299), feel (296)
Nu	See (536), like (501), get (423), time (396), away (383), know (383), feel (357), got (355), never (352), take (336)
Power	Time (480), life (428), world (383), see (361), never (348), way (309), away (289), take (266), light (262), come (257)
Progressi ve	See (422), time (416), know (347), never (346), life (346), away (331), world (325), way (280), eyes (274), like (274)
Thrash	Life (501), death (457), die (352), time (328), see (308), take (283), hell (273), know (257), never (252), blood (233)

Table 5. The 10 most frequent content words.

Although love is a universal source of inspiration in music, the only *subgenre* in which the word *love* is relevant is glam metal where it is the most frequent term. Love is conceived of as both emotional experience and sexual desire, as suggested by the results of topic modelling in which *love* occurs with *want, dream, know* as well as with *fire, burn,* and *night.* Collocates include *your, our, fighting, my, give,*

real, hot, and *making*. That glam metal revolves mainly around love is also shown by the extremely high frequency of the noun *baby,* used to address the lover. The subject of love is indeed treated in a stereotypical and glossy fashion, as can be gauged from expressions such as "my love is real", "I want your love forever", "our love burning like a flame", and "I love you so much it hurts". Whilst the noun *night* is relevant in many *subgenres* as it evokes the typical gloomy setting in which demons appear and evil deeds are committed, in glam metal it becomes the background for romantic affairs in which the band are involved, as exemplified in phrases such as "this is the night I've waited for", and "I want this night to last forever". Glam metal lacks the depth and sophistication typical of other *subgenres* of heavy metal and seems to focus predominantly upon reaching the consensus of a wider audience, often at the expense of compositional creativity. Hence it is apparent why glam metal is frowned upon by groups of diehard heavy metal *aficionados*.

The most frequent term in the HM corpus is *know*. When the verb is governed by the personal pronoun *I,* it often shows awareness of one's destiny (e.g. "I know I will take the blame", "I know I can not turn back time", "I know why, each dawn I

die"). Often it is preceded by the pronoun *you* (e.g. "you know", "don't you know", "you will know", "you've got to know"), a vague entity which only at times coincides with the listener. *You* can also personify a fellow human being in distress, or an enemy.

Although ranking only 28th in the most frequent content words list, *run* is a particularly relevant term in the HM corpus. It appears in one of the topics extracted through LDA, and is often associated with the adverb *away,* or with the prepositions *from* and *for,* to express the need to escape from a threatening being in order to survive (e.g. "better run", "run for your life"), albeit encountering difficulties (e.g. "try to survive, run but you cannot hide").

The NU and MC corpora share the majority of the most frequent words. Indeed, the two *subgenres* are similar in their interest in the self and its torments. The verbs *feel* and *know* are particularly relevant as they are a means to express a subject whose internal moods in a perpetual state of conflict and agony. Wh-adverbs are central to both corpora, i.e. *what* (558 occurrences in MC, 730 in NU), *when* (416 occurrences in MC, 366 in NU), *why* (143 occurrences in MC, 314 in NU), *how* (213 occurrences in MC, 226 in NU). High frequency of the word *when*

suggests that metalcore is heavily determined by and focused upon time (e.g. "when will we ever get what we deserve?", "scars do not heal when you keep cutting", "I'll meet you at the gates when the light of day fades my soul"). In nu metal, rather, *what*, *why*, and *how* express the subject's restless quest for answers throughout an endless inner journey (e.g. "what about me?", "what do you want me to do?", "why am I so angry inside my head?", "why do you have to put this on me?", "how could you do this?", "how can you say that you feel sorry?").

The intimate themes that characterise nu metal and metalcore are dismissed in power metal, where dynamism is key. In the PM corpus, verbs of movement or change of state are very frequent, e.g. *go* (167), *die* (165), *fall* (120), *fly* (117), *rise* (114), *fight* (103), as well as verbs related to the senses, i.e. *see* (334) and *hear* (148). *Feel* (208) is also very common but generally bears no relation to the emotional sphere. Rather, it is often employed to describe an internal state of empowerment and call to action, e.g. "can you feel the power flames will take us higher", "we do not feel ashamed, it's time to raise the flag", "I feel oh so strong and I know my own way". Power metal is indeed concerned with the theme of war, fighting malignant forces, and with the warrior's self-assertion, which often results

in a valiant battle against fate (e.g. "I'm facing my fate without fear", "we're facing our fate, to kill or to die is our destiny").

Progressive metal is a heterogeneous *subgenre*. It shares many of the nouns common to other *subgenres*, such as *time*, which is the first most frequent noun with 412 occurrences, *life* (337), *world* (320), and *eyes* (263). Yet contrary to what is observed in other corpora, in the PrM corpus the word *mind* is numerically significant (219). Amongst its occurrences are "you can find all you need in your mind", "I see the eye of time, facets of a crystal mind", and "open up your mind and seize the day". The most relevant collocation is *in my mind,* which indicates that cognition is a space in which events take place, and that the mind can connect the individual to higher dimensions. Amongst the most commonly employed verbs are *see* (403), *know* (331), *feel* (218), and *find* (184). The latter, in particular, denotes an attempt to arrive at an intangible destination, as expressed by the phrases "find my way", "find the truth", "find the answers", "find my thoughts", and "find the meaning in my life". Arguably, the umbrella theme of progressive metal is the exploration of various aspects of existence but the erratic nature of the

music is also reflected in a diversity of topics which does not lend itself to synthesis.

Whilst content words can reveal the themes and semantic areas on which a text is built, function words can, at times, be equally informative. If we focus upon the personal pronouns *I* and *you* in all the corpora (Table 5), we see that only in the GM and the HM corpora are the numbers balanced. In particular, in the GC corpus and the TM corpus there is strong prevalence of *you*, whilst in the MC, NU, PM, and PrM corpora the preferred pronoun is *I*. This confirms that metalcore and nu metal lyrics in particular revolve around the individual and their personal experiences, whilst grindcore and thrash metal lyrics focus upon and are directed towards a fictional enemy, who is the recipient of threats and offences. Personal pronouns are particularly scarce in the BM and DM corpora, signifying that the focus of the lyrics in these *subgenres* transcends personal experiences; rather, they explore universal themes by means of a narrative style that rises above the personal level and embraces mankind as a whole.

Another relevant function word is *your*. *Your* forms significant collocations with the noun *god* in the DM and BM corpora. Indeed, spirituality and religion

permeate the lyrics of certain *subgenres* of heavy metal. Yet whilst the former is a means to elevate one's self and takes on a positive connotation, the latter is to be fiercely opposed as we see in phrases such as "your god is dead", "I will erase god for all mankind", "your god is not there", "god's word is useless", and "there is no god and for all eternity you will cry".

Black	I (1864), you (1079)
Death	I (1990), you (1427)
Glam	I (4196), you (4027)
Grindcore	You (2934), I (1516)
Heavy	You (3181), I (3107)
Metalcore	I (4773), you (3291)
Nu	I (5303), you (4553)
Power	I (2497), you (1999)
Progressive	I (3364), you (2119)
Thrash	You (2893), I (1957)

Table 5. The number of occurrences of *I* and *you*.

Sentiment Analysis

We ran a Sentiment Analysis (SA) algorithm[8] in order to evaluate the atmosphere conveyed by the lyrics in each *subgenre*. We calculated the SA score of each song and subsequently averaged all the scores for each *subgenre*. The results are reported in Table 6.

	SA score
Black	-0.4258
Death	-0.6680
Glam	0.2933
Grindcore	-0.4756
Heavy	-0.1585
Metalcore	-0.3554
Nu	-0.3268
Power	-0.1563
Progressive	0.0112
Thrash	-0.6685

Table 6. Sentiment Analysis scores.

---

[8] Sentiment Analysis algorithms process the lexical content of a text and generate a real number ranging from -1 to +1, where -1 is extremely negative, whilst +1 identifies exceptionally positive texts. Owing to its ability to capture polarisation of opinions, Sentiment Analysis has been widely applied to different types of text.

In general, the predominant sentiment for the *genre* is negative since eight out of 10 *subgenres* have produced negative numbers. This agrees with the results of the lexical analysis conducted above and with Weinstein's claim that: "the confidence in a brighter tomorrow have no place in heavy metal" (2000: 35).

Death metal and thrash metal are the most negatively connotated *subgenres*, followed by grindcore and black metal. The use of negative vocabulary as well as swear words (especially in grindcore lyrics) surely accounts for the negative scores.

One might wonder why when applied to the BM corpus SA does not return a lower score, given that the *subgenre* is generally associated with a dark and gloomy atmosphere. Closer inspection of the data shows that not all the lyrics in the BM corpus produced negative sentiment. In fact, some generated a positive score. This can be explained by the fact that black metal, despite its preference for grim and eerie tones, often depicts aspects of the natural world, drawing inspiration from the wilderness, forests, mountains, winter, storms, and blizzards (Marone 2014). Mythology and folklore figure in many songs too, stemming from a fascination for the distant past. Such themes do not necessarily produce a negative undertone and can account for the results yielded by SA.

Only two *subgenres* are characterised by a neutral/slightly positive overtone; these are glam metal and progressive metal. This is unsurprising given that glam metal mainly deals with love and sex in a rather superficial and stereotypical fashion, so much so that even the impact of terms related to heartache is attenuated. Conversely, the score obtained when running the algorithm on progressive metal lyrics can be explained in light of the variety of themes explored in the *subgenre*, some of which might create a negative atmosphere, others a more optimistic one.

Conclusion

The present work aims to investigate a sample of heavy metal lyrics. Being a particularly heterogeneous *genre*, not only in terms of music but also in the language utilised and the themes explored, heavy metal lends itself to analysis using the tools of linguistics and NLP.

Lexical richness was investigated by means of different measures, i.e. average word length, TTR, MATTR, LD, frequency of hapax legomena, percentage of rare or out-of-vocabulary words. All measures indicate that two *subgenres* are particularly rich and complex from a lexical point of view. These are death metal

and black metal. Lyrics in these two *subgenres* are made up of longer words and are riddled with Middle English terms, proper nouns referring to mythology and spirituality, and some foreign words. At the other end of the spectrum are glam metal, metalcore, and nu metal, which do not seem to take pride in lexical creativity and tend to rely upon more common expressions.

Glam metal focuses mainly upon romance and makes use of stereotypical language. Nu metal and metalcore, by contrast, deal with pain and suffering from an intimate perspective, as the vast predominance of the pronoun *I* confirms. Anger and violence are retrievable in death metal, grindcore, and thrash metal. In the latter two they take the shape of extreme blasphemy and obscenity directed towards society and religion. Such a negative attitude is reflected in the SA scores, which are the lowest in the whole *genre*. Black metal too is suffused with negative imagery, so much so that the most frequent term in the BM corpus is *blood*. Yet the sentiment is more varied owing to a richer range of topics explored.

The main topic in heavy metal lyrics is saving one's own life from the dire threats of a powerful enemy. Themes of war and battle are also central to power metal, however, whilst in heavy metal lyrics escape is often encouraged, as

numerous occurrences of the verb *run* followed by *away*, *from*, and *for* indicate, power metal revolves around proving one's worth by fighting the battle at the expense of one's life.

Topic analysis on progressive metal has proven particularly difficult . This can be explained by the extreme diversity of the topics covered, which deal with many different aspects of human existence.

Diversity is indeed the most striking feature of heavy metal as a *genre*, whose language can range from complex and sophisticated words to very crude, bordering on grisly imagery.

REFERENCES

BIRD Steven, LOPER Ewan, KLEIN Edward (2009), *Natural Language Processing with Python*, O'Reilly Media Inc, Sebastopol.

BLEI David, NG Andrew, JORDAN Michael I. (2003), "Latent Dirichlet allocation", «Journal of Machine Learning Research», 3, pp. 993-1022.

BOLT Carol (1973), *Rosie learns French*, Alexander Street Press.

BURROUGHS William S. (1964), *Nova Express,* Grove Press, New York.

COVINGTON Michael A., MCFALL Joe D. (2010), *Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR),* «Journal of Quantitative Linguistics», 17 (2), pp. 94-100.

DALE Robert, MOISL Herman, SOMERS Harold (eds) (2000), *Handbook of Natural Language Processing*, New York, Marcel Dekker.

FELDMAN Ronen, SANGER James (2007), *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, Cambridge.

JURAFSKY Dan, MARTIN James H. (2008), *Speech and Language Processing*, Prentice Hall, New Jersey.

MANNING Chris, SCHÜTZE Hinrich (1999), *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge (MA).

MARONE Vittorio (2014), "A Winterhorde in a Ravenrealm: Immortal's lyrics as an expression of Northeroic Gothic" (PDF), «Aeternum: The Journal of Contemporary Gothic Studies», 1 (2), pp. 40-60.

MERTZ David (2003), *Text Processing in Python*, Addison-Wesley, Boston.

WALSER Robert (1993), *Running With the Devil: Power, Gender, and Madness in Heavy Metal Music*, University Press of New England, Hanover, N. H..

WEINSTEIN Deena (2000), *Heavy Metal: The Music And Its Culture, Revised Edition,* Da Capo Press.

WEINSTEIN Deena (2014), *Just So Stories: How Heavy Metal Got Its Name—A Cautionary Tale*, «Rock Music Studies», 1:1, pp. 36-51, DOI: 10.1080/19401159.2013.846655

Dictionaries

Cambridge Online Dictionary (2008), Cambridge University Press, available at https://dictionary.cambridge.org, last accessed 20th September 2020.

Collins Online English Dictionary, available at https://www.collinsdictionary.com/us/dictionary

"heavy metal, n." Oxford English Dictionary Online, Oxford University Press, available at https://www-oed-com, last accessed 20th September 2020.

Merriam-Webster.com (2011), Merriam-Webster, available at https:// www.merriam-webster.com, last accessed 20th September 2020.