

Valentin Werner\*

# Assessing hip-hop discourse: Linguistic realness and styling

<https://doi.org/10.1515/text-2019-2044>

**Abstract:** This study provides a corpus-linguistic take on hip-hop discourse (as represented in rap), relating to one of the most influential cultural mass movements to date. To this end, a custom-built corpus of lyrics by US-American rap artists (LYRAP) was compiled, containing performed hip-hop discourse over a 25-year period. This material is used to test the alignment of hip-hop discourse with African American English in terms of morphosyntax, and to determine the amount of styling present in the lyrics. In addition, a comparative perspective with pop lyrics (as represented in the LYPOP corpus) is established, and highly characteristic lexical and discourse features of hip-hop discourse are identified. The analyses suggest that “linguistic realness” (in terms of conveying a street-conscious identity) is created on multiple structural levels, but that different artists style their lyrics to various extents to achieve this realness, and that a complete congruence of African American English with hip-hop discourse cannot be traced.

**Keywords:** hip-hop, rap, lyrics, pop culture, style, African American English (AAE)

## 1 Introduction

### 1.1 Context

Even though initially regarded as yet another (ephemeral) form of youth culture, hip-hop has grown into a mass movement, and is currently viewed in a line with other influential musical genres (such as blues, jazz, soul, and funk; Scharenberg 2004: 16). Rooted in Afro-American and Caribbean oral traditions (Smitherman 2000: 269), this alternative urban form of expression emerging from marginalized urban African American (and Latino) communities of the 1960s/1970s, mainly representing socially peripheral youth of the time, became more and more mainstream starting from the late 1980s, and can now be seen as

---

\*Corresponding author: Valentin Werner, English and Historical Linguistics, University of Bamberg, Bamberg 96045, Germany, E-mail: [valentin.werner@uni-bamberg.de](mailto:valentin.werner@uni-bamberg.de)

a multi-faceted global phenomenon (Mitchell 2015: 227). Commonly, a distinction is drawn between “hip-hop” (henceforth, HH) as a term for the overall cultural phenomenon, including the practices of breakdancing, graffiti writing, DJ-ing,<sup>1</sup> and “rap” (alternatively labeled “MC-ing”) as arguably the central element of the overall HH complex (Olivo 2001: 67).

As the “aesthetic placement of verbal rhymes over musical beats” (Alim 2006: 4), rap represents the field where language – in the form of lyrics – features most prominently, and thus where hip-hop discourse (HHD) is shaped to a large extent. Hence, it deserves attention to understand the communicative practices in what now has arguably become a dominant part of pop culture world-wide (see also Armstrong 2004: 335). From a broader perspective, the overall sociocultural impact of HHD, alternatively referred to as “Hip Hop Nation Language” (Alim 2015) or “Blinglish” (Jansen 2012),<sup>2</sup> can further be inferred from reports that label rap an influential modern form of poetry (Dalzell 2014: 18) and studies discussing its pedagogical and political potential (e.g. Newman 2005). Moreover, linguistic features associated with HHD (i) can be found in other pop and consumer culture manifestations (such as advertisements, video games, in movies and pop literature as well as in the worlds of art and fashion; see Richardson 2006: 97–104; Jansen 2012: 343; Dalzell 2014: 16–17) and (ii) have even entered wider colloquial usage in the US and elsewhere (Kimminich 2004: viii).

The present study takes account of the extensive sociocultural impact of HHD and provides additional perspectives gained through a corpus-based analysis. To this end, after a description of the steps in preparing the data and the methods used (Section 2), the main part of the study provides a linguistic analysis of HHD as represented in rap lyrics, and draws the link to features of African American English (Section 3.1). In addition, it compares rap lyrics with a “close relative”, namely pop lyrics, to identify salient morphosyntactic differences as well as differences regarding the lexicon and further discourse features (Section 3.2). In passing, a number of claims on HHD voiced both in sociolinguistics and other disciplines are re-assessed against the corpus evidence. The concluding part (Section 4) contextualizes the overall results, and highlights areas where further research is required.

---

<sup>1</sup> See Porter (2001: 150–155) for a musicological history of HH.

<sup>2</sup> The former takes into account that HH is seen as global movement, and has a broad scope representing a “synergistic combination of speech, music, and literature” (Alim 2015: 851); the latter (a blend of *bling* ‘jewelry’ and *English*) alludes to overt displays of material wealth in HH culture (see Section 3.2).

## 1.2 Literature review

In spite of the central status of rap lyrics (also in contrast to the music; Danesi 2015: 165), and the resulting largely text-based nature of HH (Dalzell 2014: 18), analyses from a linguistic perspective are comparatively scarce, as a main focus of previous works has been on other aspects. HH-related scholarly studies have, among others, been undertaken from the perspectives of musicology (e.g. Komaniecki 2017; Connor 2018), sociology (e.g. Lena 2016), cultural studies (e.g. Richardson and Scott 2002), political science (e.g. McNair and Powels 2005), and psychology (e.g. Richards 2017).

Thus, despite a few recent publications (see especially the collection Ross and Rivers 2018), it can be claimed that actual language use represents an under-researched topic both in empirical linguistics and in HH studies in general (see also Cutler 2007: 519). There exist a number of works with cursory statements on the language of rap (e.g. Edwards 1998; Widawski 2015), and those that consider diverse topics such as the linguistic representation of identity as well as the world-wide dissemination and appropriation of HHD from an anthropological perspective (Morgan 2001; Alim 2006, Alim 2009, and Alim 2015; Pennycook 2008; see also the contributions in Alim 2009; Terkourafi 2012a). There are also studies with highly specific foci such as the usage of linguistic features associated with HHD in White teenagers and rappers (Cutler 2007, Cutler 2009, and Cutler 2014; Bucholtz 2011), spelling conventions (Olivo 2001), language use by individual artists (Edwards 2002; Armstrong 2004; Edwards and Ash 2004), in particular cultural contexts (see, e.g. O'Hanlon 2006 on Australian HH), or in forms of electronic communication (Beers Fägersten 2008).

The summary provided in Alim (2015: 851–852) serves as a helpful starting point for the present linguistic description of HHD as represented in rap, and identifies a number of properties that can be subject to corpus-based scrutiny.<sup>3</sup> First, in line with a number of other researchers (see, e.g. Morgan 2001: 188; Kautny 2015: 102; Widawski 2015: 6–7), Alim mentions the strong link between HHD and African American English (AAE), due to the historical roots of rap in the respective oral traditions. This leads us to expect the presence of characteristic features of AAE in the rap lyrics. At the same time, Alim also draws attention to

---

<sup>3</sup> It is apparent that the corpus-based approach does not lend itself to making generalizations on the phonology of rap, nor on the interaction of rhythm and lyrics, for instance. Relevant information can be found in Smitherman (2000: 269–275), Ch'ien (2011: 69) and Connor (2018). Note, however, that phonological deviations from standard usage may be reflected in the spelling, as in *thang* 'thing' (T.I.: Pledge allegiance to the swag) or *dey* 'they' (Nicki Minaj: Set it off). See further Olivo (2001), Beers Fägersten (2008: 230–233), and Kreyer (2016: 96–100).

the existence of internal (including regional) variation within HHD (see also Alim 2006: 10), and others have argued that, to a certain extent, these features are deliberately used (Green 2002: 160). Therefore, it will be worthwhile to determine how customary their use among rap artists actually is.

Following common sociolinguistic definitions (see, e.g. Coupland 2001; Eckert 2004) such “styling” can be understood as a practice of conscious usage of particular linguistic features either as “visible manifestation of social meaning” (Eckert 2004: 44) or to “construct a social image”, that is, “to shape the way [rappers] wish to be seen by others” (Cutler 2014: 5) as a (street-conscious and “real”) member of the AAE community. Indications of this have been found in smaller-scale sociolinguistic work (e.g. Alim 2002; Álvarez-Mosquera 2015). On a related note, Alim (2015: 852) argues for a strong tie between HHD and actual “sociopolitical circumstances” of its performers, and for the salience of the issues of identity and language ideologies. It is hypothesized that this relation will surface in topic choice, mirrored in the vocabulary used in the lyrics.

Although, apparently, “sociolinguistics has identified [...] a ‘language style’ associated with hip-hop” (Cutler 2007: 532) based on small samples of data, a systematic analysis of the lexicon and the lexicogrammar of rap from a more holistic perspective is still lacking. Even those studies that use a corpus of rap lyrics to arrive at generalizations (such as Lüdtke 2007; Jansen 2012; Aleshinskaya 2013; Kautny 2015), restrict their analyses to a qualitative view, thus leaving the full potential of a corpus-based take underexploited. A recent exception is Kreyer (2016), who takes a quantitative approach and tests whether HHD can be established as a distinctive register.

### 1.3 Research questions

As stated above (Section 1.1), the present study aims to highlight the additional perspectives on American HHD that can be gained through a corpus-based analysis, and seeks to expand the description of HHD beyond analyses of the lyrics of individual artists or beyond isolated linguistic features that are studied as “identity markers” within the US HH community. More specifically, it tackles the following issues:

- i. HHD as variety/(sub-)register: Is it justified to speak of a unified “language” (cf. the term “Hip Hop Nation Language”; Alim 2015) or rather a variety of HHD? Is it justifiable to speak of HHD as a separate (and stable) register (Kreyer 2016), or is it simply a particular (performed and styled) variant of African American English?

- ii. Internal variability: To establish some kind of “linguistic realness” (Edwards and Ash 2004: 175) are there common linguistic properties used by all artists or can internal variability pertaining to linguistic usage patterns be witnessed?
- iii. Rap vs. pop: What differentiates (commercially successful) rap from (commercially successful) pop?

## 2 Data and method

Two corpora are used as a base for empirical study. The first one is a specialized corpus of HHD as represented in rap lyrics (LYRAP). It contains all lyrics available from 1991 to 2016 for a set of 15 American HH artists. Rappers included are 50 Cent, Drake, Eminem, Jay Z, Kanye West, Lil Wayne, Ludacris, Nas, Nicki Minaj, Notorious BIG, Snoop Dogg, T.I., Tupac, Tyga, and Wiz Khalifa (see Appendix A for a quantitative overview). The choice of material by these performers was motivated by the fact that they can arguably be seen as having made the transition from “underground” to “mainstream” in their careers; that is, they represent a selection of commercially highly successful artists (see also <http://rapalytics.com/articles/underground-mainstream-what-sells-rap>). Note that LYRAP epitomizes what could be labeled “classic” American HHD, and does not necessarily allow generalizations for rap as a global *multilingual* phenomenon (as described in Pennycook 2008 or Alim 2009, for instance).

The second corpus is a (smaller) purpose-built corpus of pop lyrics (LYPOP), which serves as a point of reference for LYRAP. This corpus has already been used in other studies (e.g. Werner and Lehl 2015; Werner forthcoming) and is available upon request on the platform CQPweb for research purposes.<sup>4</sup> It contains all lyrics from the top ten selling albums of the years 2000 to 2015 as determined by the *Official Charts Company* (<http://www.officialcharts.com/>), and thus covers a wide span of genuinely “pop” songs (for a discussion of the label, see Werner 2012; Werner 2018) that possess international appeal.<sup>5</sup>

All lyrics were retrieved from the online lyrics database *AZlyrics*, chosen as a source due to its inherent quality control (see also Kreyer 2016: 89). First, when submitting transcriptions of lyrics (see <http://www.azlyrics.com/add.php>), users

---

<sup>4</sup> Thanks are due to Stefan Evert (University of Erlangen-Nuremberg) for providing the infrastructure and Fabian Vetter (University of Bamberg) for support in preparing the data.

<sup>5</sup> Due to the sampling criteria, three rap albums form part of LYPOP, representing 6.7% of the tokens. This slight overlap does not invalidate the overall findings.

have to specify their source (CD booklet, another website, lyrics video, own transcription, etc.). Second, multiple corrections by other users are possible (via a “submit corrections” button on the pages for individual songs), so that the database contains a dynamic feature to facilitate improvements on stored transcriptions. The latter feature appears to be used regularly, as the number of acknowledgements for corrections for the individual songs reveals. Spot checks (matching the performed lyrics against the transcription) suggested that the input of the lay transcribers is generally accurate (see also Queen 2013: 224 for comparable approaches with other types of material).

Although the actual retrieval could be done semi-automatically with the help of the *DownThemAll!*-browser plugin, further preparatory steps were necessary to ensure comparability and compatibility with the corpus tool used. These consisted of:

- removal of file doublets caused by the inclusion of the same song in more than one release or by the presence of remix versions;
- extraction of the lyrics proper from the HTML files with regular expressions in *Notepad ++*;
- removal of unwanted annotation potentially interfering with further analyses (metainformation such as artists’ names or structural indications, e.g. *chorus*, *verse 1*, etc., regularly appearing between square brackets) and correction of individual symbols (e.g. <’> for <´> in items such as *don’t* or *ain’t*);
- automatic part-of-speech-tagging with the CLAWS tagger (Garside and Smith 1997) to facilitate searches for morphosyntactic patterns (cf. Section 3.1 and Appendix B).

The study relies on the *AntConc* suite (Anthony 2014) for keyword and concordance analyses. A summary overview of the data is shown in Table 1.

The data of the present study improves on material used previously both in terms of size<sup>6</sup> and in terms of the scope of the time period covered.<sup>7</sup> With the aim of an increased representativeness, the current analysis can thus also be seen as a step toward establishing a broader and up-to-date linguistic picture of US rap lyrics.

---

6 Cf. Edwards (1998): 41 songs/no word count given; Edwards and Ash (2004): 6,300 words; Bloomquist and Hancock (2013): 9 albums/no word count given; Kreyer (2016): 200,000 words.

7 Cf. Morgan (2001): 1990–1998; Bloomquist and Hancock (2013): 1991–2003; Kreyer (2016): 2003–2011.

**Table 1:** Overview of the data used in the study.

	LYRAP	LYPOP (reference corpus)
Period covered	1991–2016	2000–2015
Songs	3,910	1,804
Tokens <sup>1</sup>	2,095,927	534,696
Types	43,414	13,285
Average tokens per song/file	536	296
Type-token ratio	0.021	0.025
Standardized type-token ratio (per 300 words) <sup>2</sup>	53.44	38.74

<sup>1</sup>Counts based on raw text data; note that items with clitics such as *ain't*, *don't*, etc. are counted as one word. If counted separately (i.e. <'> removed from the token definition), the counts increase/decrease insubstantially (2,194,544 tokens, 39,157 types for LYRAP), not affecting overall results. Spelling variants are preserved both in LYRAP and LYPOP.

<sup>2</sup>The base of  $n = 300$  words was chosen for the standardized type-token measure as it fits the average size of the shorter lyrics in LYPOP. If a larger base was chosen, many of the songs included would receive the standardized type-token measure of 0, which would inhibit comparability. Note that the type-token results for LYRAP vs. LYPOP quantitatively confirm earlier views of a higher verbal density of HHD in contrast to pop lyrics (Kautny 2015: 101).

It is evident that the choice of the reference corpus is of central importance in a keyness analysis.<sup>8</sup> For the present study, at least two options would have been available in principle. First, an established monitor corpus (say, the *Corpus of Contemporary American English*) could serve as a means of comparison. This would be the method of choice if we wanted to contrast rap lyrics to general usage in English, and would be comparable to the approach taken in Beers Fägersten (2008), for instance, where a corpus of HH-related online postings was compared to corpora from the BROWN family (see, e.g. <http://clu.uni.no/icame/manuals/BROWN/index.htm>). Lyrics, however, due to their circumstances of production and situatedness (Werner 2012: 43), constitute a very specific register, so it is almost certain that marked differences will emerge in an analysis of the type described. Therefore, an alternative approach, namely the comparison to a “close relative” as a reference corpus (Culpeper 2009: 35) – represented by the lyrics material in LYPOP – was considered more fitting. This approach has fruitfully been implemented before (e.g. in Motschenbacher 2016; Kreyer 2016), and will most likely lead to the identification of highly characteristic, and thus more telling, properties of LYRAP, the target corpus.

<sup>8</sup> A minor issue pertains to the choice of the keyness statistic used. I opted for log likelihood (default in *AntConc*) as this approach seemed adequate for the register/genre-oriented research presented (Pojanapunya and Watson Todd 2018).

### 3 Properties of hip-hop discourse

#### 3.1 Establishing “linguistic realness” through morphosyntactic means

As mentioned above, HHD is closely associated with a number of features of AAE. At the same time, HHD features have been labeled either “vernacular” or “nonstandard” (see, e.g. Lüdtke 2007: 109; Jansen 2012: 345). It is evident that such views only apply if Standard (American) English is taken as a yardstick rather than considering AAE as a baseline of comparison. Yet, to characterize HHD, and to determine whether some kind of unified and distinguishable rap style exists among the set of artists represented in LYRAP, it is helpful to test the actual overlap between AAE and HHD. To this end, a morphosyntactic feature catalog based on elements mentioned in the relevant literature and on the set of items with a high (“A”) pervasiveness rating in Urban African American Vernacular English in the *electronic World Atlas of Varieties of English* (eWAVE; Wolfram 2013) was created. This feature catalog contains 27 different categories, and is described in detail in Appendix B. It is evident that such a list is far from comprehensive; less pervasive, but still potentially characteristic structures (for instance, those rated “B” in eWAVE)<sup>9</sup> are partially ignored. However, it complements previous studies that have relied on the analysis of a single morphosyntactic feature (notably, copula absence, as in (1); see, e.g. Alim 2006: 117; Kreyer 2016: 102–106) as a characteristic marker of HHD.

- (1) He Ø running like a bitch with his tail between his legs (Nas: Eye 4 an eye freestyle)

The following analysis rests on the assumption that an extended number of AAE features indexes “linguistic realness” (Edwards and Ash 2004: 175; see also McLeod 1999; Ochmann 2013: 427), or the situational appropriateness and authenticity (see Coupland 2014 for discussion) of the lyrics, and is thus seen as a desired property for which rappers strive (see also Alim 2002, Alim 2006). In different terms, it has been stated that rappers use “linguistic symbols, which signify membership, role and status” (Cramer and Hallett 2012: 257–258). In this regard, it has further been argued that rappers consciously use AAE features to express their defiance of mainstream language, perceived as White and standard

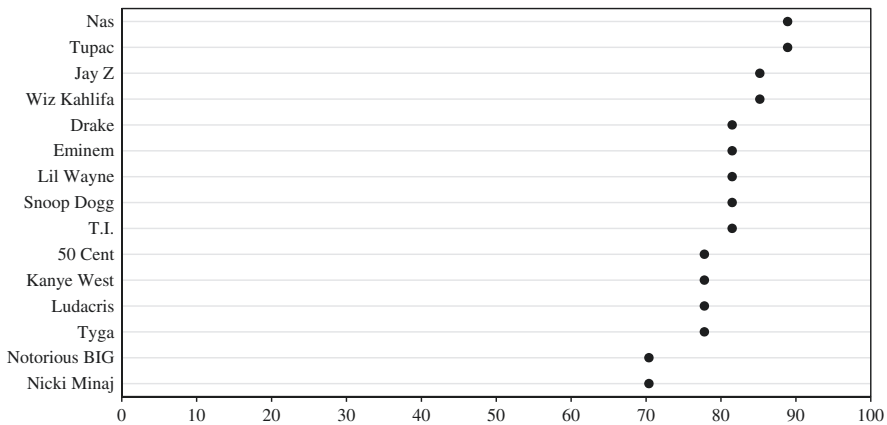
---

<sup>9</sup> Some of the features rated “B” in eWAVE (e.g. completive/perfect *done*) were in fact mentioned as highly characteristic in other sources, and were thus included in the feature catalog.



(Morgan 2001: 188; Edwards and Ash 2004: 167; Terkourafi 2012b: 8; Álvarez-Mosquera 2015). All this eventually pertains to the issue of styling (see Section 1.2) and how social meaning in terms of conveying a street-conscious HH identity is created.

Figure 1 shows an overview of the percentage of the 27 features from the catalog present (i.e. they occurred at least in three different songs) in the lyrics of the 15 artists included in LYRAP, and thus creates a kind of “linguistic realness” index for the individual rappers. A value of 100% would imply that an artist uses all of the 27 features tested.<sup>10</sup>



**Figure 1:** Percentage of morphosyntactic features associated with AAE/rap (individual LYRAP artists).

On first sight, the perspective that emerges from Figure 1 is a fairly homogeneous one, which suggests that all artists use the vast majority (>70%) of features. Therefore, the general principle to “highlight AAE [...] features” (Morgan 2001: 194), potentially to create texts that are perceived as authentic and linguistically appropriate by the projected AAE-literate audience, seems to hold for the linguistic area explored.

In terms of audience design/referee design (Bell 1984), earlier research has suggested that only very few lexicogrammatical features of AAE are used in HHD to ensure appeal to White audiences (Edwards 1998: 143–144), which represent the majority of HH consumers (Armstrong 2004: 339; Mitchell 2015:

<sup>10</sup> An exact breakdown of the presence of features for individual artists is available at <https://osf.io/vf98d/>.

226). In fact, this finding is not borne out by the data as represented in Figure 1, suggesting that the styling indeed seems to be geared toward an AAE-speaking (or at least AAE-understanding) audience. However, none of the artists uses all of the features, and a certain amount of internal variability in the scores can be observed (range: 70–89%; SD = 5.3%). This indicates that some artists apparently rely on a broader or more restricted set of morphosyntactic features to achieve linguistic realness in their lyrics.

Viewed differently, this finding could also be interpreted in terms of more pronounced styling, that is, conscious (over-)use of the AAE features to increase audience appeal (see Section 1.2), in particular by the artists with higher scores (Alim 2006: 78; Cutler 2007: 522). This perspective emphasizes the performative nature of lyrics, as argued in Aleshinskaya (2013: 423), for instance, and should be seen in the context of rappers “forging a linguistic-cultural connection with the streets” (Alim 2006: 111).

A closely related issue pertains to the ostensibly counterintuitive finding that the lyrics of Eminem, a White artist, also contain 81% of the AAE features tested. Previous research has suggested that White American rappers consciously use AAE features to increase audience appeal, even though the issue of authenticity is apparently at stake here (Armstrong 2004: 336; Cutler 2009: 81). In the particular case of Eminem, his personal background of having grown up in a largely black community, and his successful integration into the associated (HH) culture do not stand in contrast to his authenticity as a rapper, but rather led to an “honorary Black status” (Cutler 2009: 87) within the community. For these reasons, the presence of the AAE features in Eminem’s lyrics is predictable to a large extent.

A second type of assessing linguistic realness is presented in Figure 2, where a flipped perspective is taken. In this view, the individual features tested serve as categories and the percentages of use for each feature among the set of 15 rap artists in LYRAP is shown. A value of 100% occurs if the feature is used by all artists in LYRAP.

Figure 2 shows that 15 out of the 27 features tested appear in the lyrics of all 15 LYRAP artists (dots), and thus seem to be most strongly associated with linguistic realness. Consequently, this set of features comprises the core of morphosyntactic AAE features used in HHD. There is a second group of seven features appearing in the lyrics of at least 50% of the LYRAP artists (triangles).

The remaining five features (squares) are used by the minority of rappers only. The latter finding is particularly striking in view of the fact that these features either have received a high pervasiveness (“A”) rating in eWAVE (such as double modals as in (2); 33% usage rate; Wolfram 2013) or have been

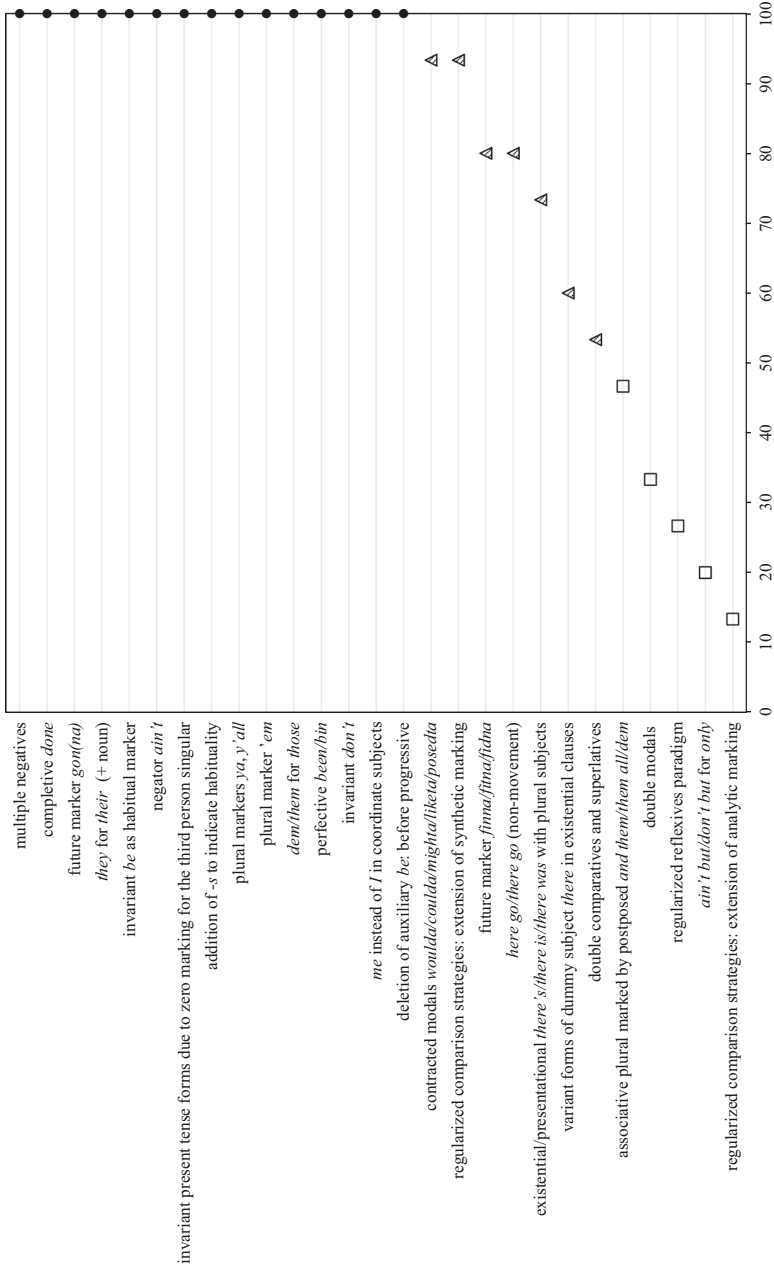


Figure 2: Percentage of morphosyntactic features associated with AAE/rap (feature-based view).

identified as highly characteristic of HHD in earlier studies (such as *ain't but for only* as in (3); 20% usage rate; Lüdtke 2007).

- (2) I **might could** fit you in on a Wednesday (Drake: Digital dash)
- (3) Damn, you **ain't but** sixteen nigga (Tupac: Shorty wanna be a thug)

This suggests that (i) a one-to-one association between AAE and HHD falls short in the area of morphosyntax (cf. Alim 2015: 852); and (ii) a number of the features tested seem to be less closely associated with HHD than previously assumed (cf. Alim 2006: 114); or, in the context of audience design/styling, they are not consistently employed as consciously used markers of linguistic realness.

Overall, the finding that HH artists seek to achieve linguistic realness through various means can be interpreted in the following manner: There seems to exist a certain canon of features (15/27 features tested are used by all artists, a further seven by at least 50% of artists), but also some internal variability. Thus, the evidence gained from the analysis of LYRAP confirms earlier studies on styling that have claimed conscious adaptations within HHD toward informal AAE with the aim of maintaining authenticity (Edwards 1998: 143; Alim 2006: 124), even though HHD and AAE do not align completely. The presence of internal (linguistic) variability further reflects previous (cultural) analyses that have argued against a reductionist view of authenticity and that see HH (and HHD) as an inherently complex phenomenon (see, e.g. Ochmann 2013: 429).

### 3.2 Rap vs. pop: Keyword analysis

In addition to morphosyntactic features, the lexicon of HHD has received considerable attention. Some observers have even claimed that HHD is merely “slang”, that is, mainly characterized by lexical features deliberately deviant from standard use (see, e.g. Dalzell 2014) and conveying “social signals” (Eble 2004: 375). While this view apparently falls short of linguistic realities (see Section 3.1), the lexicon still forms an essential part of HHD. It is the linguistic area where – besides linguistic realness – “cultural realness” (Edwards and Ash 2004: 175) is established in terms of content choice and use of specialized vocabulary, again with the aim of conveying a street-conscious identity.

In the following, I will establish the central topics in LYRAP, using the keyword function of *AntConc*. Table 2 shows the top 30 entries emerging from the keyword analysis.

Table 2: Top 30 keyword list (LYRAP vs. LYPOP).

Rank	Absolute frequency	Keyness value	Keyword	Domain
1	11,912	5,072,199	<i>nigga</i>	ghetto/profanity
2	6788	2,928,104	<i>niggas</i>	ghetto/profanity
3	8478	2,522,299	<i>shit</i>	profanity
4	6869	2,474,632	<i>bitch</i>	sex/profanity
5	6871	1,904,137	<i>fuck</i>	sex/profanity
6	9472	1,371,590	<i>ain't</i>	grammar
7	12,468	1,333,136	<i>they</i>	grammar
8	2838	1,263,012	<i>niggaz</i>	ghetto/profanity
9	5541	1,208,669	<i>money</i>	ghetto
10	6079	1,075,665	<i>ya</i>	spelling
11	2695	1,013,955	<i>bitches</i>	sex/profanity
12	3290	821,130	<i>ass</i>	sex/profanity
13	2173	793,605	<i>y'all</i>	grammar
14	3079	696,736	<i>'em</i>	grammar
15	1653	653,878	<i>hoes</i>	sex/profanity
16	5383	641,032	<i>them</i>	grammar
17	1863	595,105	<i>wit</i>	spelling
18	13,585	538,457	<i>got</i>	
19	2380	530,702	<i>yo</i>	spelling
20	1451	522,487	<i>pussy</i>	sex/profanity
21	2021	467,463	<i>em</i>	grammar
22	18,531	452,134	<i>like</i>	
23	14,046	445,317	<i>get</i>	
24	1131	431,173	<i>hood</i>	ghetto
25	46,582	420,866	<i>a</i>	
26	15,871	419,521	<i>up</i>	
27	5674	419,045	<i>man</i>	
28	1065	415,140	<i>weed</i>	ghetto
29	1334	387,871	<i>dick</i>	sex/profanity
30	1283	375,605	<i>l'ma</i>	grammar/spelling

### 3.2.1 General aspects

Due to the fact that there are no “official” transcription guidelines for the lyrics databases, spelling variants may appear. In Table 2, both *'em* (rank 14) and *em* (rank 21) as well as *niggas* (rank 2) and *niggaz* (rank 8) occur. While the two variants in the former pair in all likelihood do not carry any specific connotations (and thus may be treated together), it has been argued (see, e.g. Olivo

2001: 72; Beers Fägersten 2008: 233–234) that the *-z* spelling in *niggaz*, in addition to orthographically reflecting a voiced sibilant, may serve as an iconic indexical marker of HH authenticity.<sup>11</sup>

Note further that – in addition to function words such as *a* and *up* (not discussed) – the search retrieves a number of features already mentioned under the label of morphosyntax (see Section 3.1), which is arguably testimony to an integrated lexicogrammatical profile of HHD. In particular, the list represented in Table 2 yields the negator *ain't* (rank 6), as in (4), plural marker <sup>(s)</sup>*em* (ranks 14 and 21), as in (5), and the nonstandard second person plural form *y'all* (rank 13), as in (6).

- (4) It **ain't** the Truman show it's the human show (Nas: Where's the love)
- (5) You showed **'em** where the Mag at (50 Cent: Call me)
- (6) Fire it up, **y'all** shift the lane (Kanye West: Perfect bitch)

The comparatively high frequency of function words such as *them* and *they* is caused by the use of demonstrative *them* for *those* (rank 16), as in (7), and possessive *they* for *their* (rank 7), as in (8).

- (7) Yeah, we ridin in **them** heavy Chevys (T.I.: Heavy Chevys)
- (8) Niggaz playin **they** hands wrong (Notorious BIG: Tonight)

The list further yields two forms not discussed in Section 3.1 for the reason that they are mainly spelling-related and consequently not part of the set of pervasive morphosyntactic features of AAE in eWAVE. These are the reduced futurate form *I'ma* (rank 30), as in (9), and the reduced second person (possessive) pronoun/determiner *ya* (rank 10) and *yo* (rank 19), as in (10) and (11).<sup>12</sup>

- (9) Look, **I'ma** tell you this (Lil Wayne: Let's go)
- (10) That's some crazy ass shit for **ya** (Tyga: Wrong in the right way)

---

<sup>11</sup> Similar patterns have also been found in electronically mediated discourse and youth culture more generally. Note that these spellings may represent the transcribers' rather than the artists' (written) production.

<sup>12</sup> *Ya* and *yo* may alternatively be used as phonesthetic devices (Werner 2012: 41) or discourse particles (Beers Fägersten 2008: 220).

- (11) I hit **yo** ass up (Eminem: The re-up)

While (9) nicely illustrates continuing grammaticalization of the *going to* future in an informal domain along the lines *I'm going to* > *I'm gonna* > *I'ma* (occasionally also in the apostrophe-free and thus arguably further reduced form *Im(m)a*; cf. Beers Fägersten 2008: 232), (10) and (11) can be seen as orthographic reflexes of the non-rhotic pronunciation associated with AAE (cf. *nigga(s/z)* or *sto'* (“store”) in (12); see, e.g. Olivo 2001: 72–73; Lüdtkke 2007: 109–111; Jansen 2012: 374 for discussion and further examples). A similar finding applies to *wit* (“with”; rank 17), as in (13), where a stop regularly replaces the labiodental fricative in AAE.

- (12) They got a **nigga** that own the **sto'** (Lil Wayne: Tha block is hot)

- (13) Squeeze in ya tight skirt, match **wit** ya best purse (T.I.: King and queen)

Another item figuring prominently is *like* (rank 22). Mainly two usages are responsible for its high keyness in LYRAP. As exemplified in (14), it may serve as part of quotative BE *like*. Alternatively, *like* is used in similes, as in (15) or (16), highlighting either the prowess in rapping (known as “skills”), success in the music business, monetary affluence or (less often) the moral integrity or cultural realness of the artist (cf. Green 2002: 156), as in (17), where a cultural reference to a highly popular sitcom starring African-American actors is made, for instance.

- (14) They **be like**, Oh God girl you bad (Nicki Minaj: Here and there girl you bad)

- (15) They have people waiting at the airport like five days **like** I'm a Beatle or somethin (Jay Z: I did it my way)

- (16) I gotta hold up my estate **like** Donald Trump (Lil Wayne: Get down)

- (17) So I stay dirty **like** “Sanford and Son” (50 Cent: Tony Yayo explosion)

Turning to content words (all nouns, apart from ambiguous *fuck* and *shit*), the current keyword analysis converges with previous studies that have identified a number of dominant topics/semantic fields around which HHD revolves (cf. Edwards 1998: 132; Scharenberg 2004: 25; McNair and Powles 2005: 350; Lüdtkke 2007: 163; Danesi 2015: 165; Kreyer 2016: 94). The most central ones are (i) sex (*bitch(es)*, *fuck*, *ass*, *hoes*, *pussy*, *dick*), and a broader notion of (ii) “ghettocentricity” (Scharenberg 2004: 25) that includes various

(partly related) subfields such as self-conscious Blackness (*nigga(s/z)*, ranks 1, 2, 8; *hood*, rank 24), crime and drugs (*weed*, rank 28), monetary success (rank 9) as well as profanity (*shit*, *fuck*, *nigga(s/z)*, and all sex-related terms). These vulgarisms are often viewed as a trademark feature of HHD, carrying covert prestige (Beers Fägersten 2008: 223–234), and have been claimed to represent “means of communicating anger, outrage and ‘realness’ to audiences of young, urban Blacks” (Edwards 2002: 65; see further below).

Overall, the notion of ghettocentricity is linguistically reflected in a drastic demarcation of HHD from standard usage associated with (White) mainstream that could be related to the broader issue of “anti-language” (see Section 4 and Scharenberg 2004: 26; Dalzell 2014: 18). More specifically, LYRAP exemplifies semantic shifts in a number of items, such as *nigga* (18) or *shit* (19), where a transformational process of “counter-labeling” (Kirkland 2015: 836) occurs in which the (standard) negative or taboo connotations of terms are subverted (see also Smitherman 2000: 279–280; Kreyer 2016: 101). In this regard it has also been argued that the subversion of standard orthography may serve as a supporting formal means to underscore the deviation from mainstream (White) conventions and therefore to assert the linguistic identity of the rappers (see Olivo 2001: 74–80; Dalzell 2014: 21; Danesi 2015: 255 for discussion and further examples).

(18) A young rich **nigga**, I’m buying for my whole crew (Wiz Kahlifa: On me)

(19) My flow be the **sick shit**, gravely **ill** (Jay Z: Murder gram)

Further items (not included in the top 30 shown in Table 2) where such semantic shifts occur without a shift in orthography are *ill* and *sick*, as in (19), *motherfucker*, as in (20), *hustle*, as in (21), or *bad(dest)* and *pimp*, as in (22). It is evident that this list of examples is highly reductive, but it serves well to illustrate the range of semantic inversion (see also Morgan 2001: 198; Beers Fägersten 2008: 223; Dalzell 2014: 18; Kirkland 2015: 839) and linguistic creativity in HHD.

(20) Move smooth as a **motherfucker**, me and my nine, I’m as cool as a **motherfucker**, I’m a get mine (Tupac: Picture me rollin’)

(21) I had so much **hustle** plus I was down to ill (Jay Z: A week ago)

(22) The **baddest pimp** nigga that you ever gon see (Snoop Dogg: Stacey Adams)



The remaining content items appearing in the top 30 keyword list presented in Table 2 are *man* (rank 27) and *get/got* (ranks 23 and 28). The high keyness of the former is caused by its use as a ritualized discourse marker at the beginning or end of lines, as in (23) or (24), arguably to establish a conversational tone, while the salient discourse topic of “manliness” (as identified by McLeod 1999: 142; Edwards and Ash 2004: 175; Scharenberg 2004: 25; Kirkland 2015: 844), illustrated in (25) and (26), may also contribute to its pervasiveness.

- (23) **Man** what’s up this is Marv (Ludacris: Bots radio)
- (24) Yeah, 9th Wonder! Don’t judge me **man** (Drake: Think good thoughts)
- (25) But keep your nuts cause this is a **man’s** game (Jay Z: In my lifetime)
- (26) But I’m a **man** so I’m able to handle my situations correctly (Lil Wayne: Grown man)

The high keyness of *get/got* mainly results from an informal usage of *got* as a possessive, as shown in (27), and *get*-passives, as shown in (28).

- (27) I **got** a big ego, bitch don’t touch me (Tyga: Bitch betta have my money)
- (28) Okay till the anger dissipates you **get punched** in the face (Eminem: Ridaz)

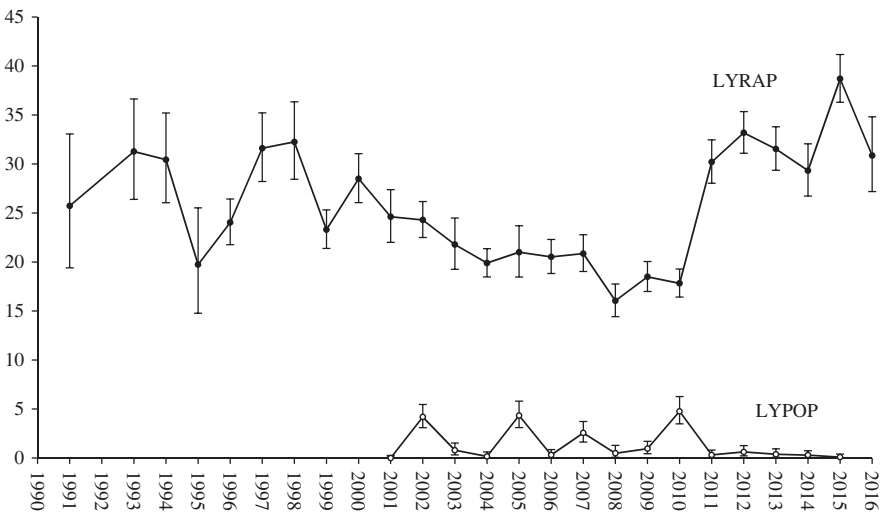
Thus, these two features can be added to a catalog of morphosyntactic features characterizing HHD (see Section 3.1).

Kreyer (2016: 94) suggests “warfare, defense, weapons, army” as another salient semantic field. In the present keyword analysis, relevant items such as *shots* (rank 108), *gun(s)* (ranks 127 and 138), *clip* (rank 149), *glock* (a make of guns; rank 248), *nine* (‘gun with nine millimeter caliber’; rank 201, see example (22) above) only appear further down, but still prominently, in the keyword list. A similar finding applies to the self-referential field of “music/rapping” (Lüdtke 2007: 163), where *drop* (+ *rhymes/beats/etc.*; rank 45), *rap* (rank 46), *flow* (rank 90) or *rappers* (rank 134) occur on lower positions.

Danesi (2015: 165) further mentions “contemporary political issues or aspects of black history” as a central concern in HHD (see also Ch’ien 2011: 61; Terkourafi 2012b: 2). However, this semantic area cannot be traced among the top keyness items of LYRAP beyond the items mentioned above (*nigga(s/z)*, ranks 1, 2, 8; *hood*, rank 24).

### 3.2.2 Profanity

On a somewhat related note, researchers have claimed that rap has developed from “political or conscious rap” to the “more controversial and nihilist genre of gangsta rap” (McNair and Powles 2005: 341) characterized by an increasing usage of profanity, allegedly to satisfy market demands (Armstrong 2004: 343). In order to test the accuracy of such claims, I present a short-term diachronic perspective. To this end, the incidence of terms from the top 30 keyness list (Table 2) categorized as profanity (*nigga(s/z)*, *shit*, *bitch(es)*, *fuck*, *ass*, *hoe(s)*, *pussy*, *dick*) is tested over the 25-year period studied. Results are shown in Figure 3.<sup>13</sup>



**Figure 3:** Profanity rates (normalized frequency per 1,000 words) in LYRAP (1991–2016) and LYPOP (2000–2015) with 95% confidence intervals.

Above all, Figure 3 suggests that, while profanity rates in the period studied may vary by a factor of 100% (average 25.84; range 16.04–38.69; SD = 5.78) in LYRAP,

<sup>13</sup> It is evident that the content of individual songs may determine frequencies of profanity and nonstandard features (as argued in Edwards and Ash 2004: 172–173, for instance). However, as our aim here is to deduce a general tendency over time, and as the corpus size is deemed adequate for such a deduction, the year-by-year perspective established is fitting. Note that the lyrics of individual songs for which no release data could be determined is ignored in this part of the analysis. In Figure 3, no data point appears for 1992 as no album contained in LYRAP was released that year.

a long-term consistent trend in HHD toward markedly more profanity cannot be traced. Rather, a differentiated picture emerges, indicating a three-partite division along the following lines: Sub-period 1 (1991–1999) yields an undulating development with marked drops in profanity rates in the years 1995/1996 and 1999, all other rates exceeding the average value. The results for sub-period 1 have to be taken with a grain of salt due to data scarcity for individual years, however, as indicated by the larger (and partly overlapping) confidence intervals. Actually, the years 1991 and 1995 are represented by a single album each (both from Tupac), so these dips may alternatively be viewed as an idiosyncrasy of the reduced profanity rates of this artist (cf. Edwards and Ash 2004: 172–173), who apparently relies on different linguistic means (see Figure 1 in Section 3.1) to assert his realness.

Sub-period 2 (2000–2010) is characterized by an overall decline in profanity rates by 60%. This trend is robust and statistically significant, as indicated by the confidence intervals, which clearly do not overlap when the beginning and the end of the period are compared. Further, it is particularly striking as the quantitative linguistic evidence stands in stark contrast to the “cultural pessimist” views voiced by sociologists in the middle of the very period (Armstrong 2004: 343; McNair and Powles 2005: 341).

However, sub-period 3 (2011–2016) sees a surge in profanity rates, with all values for individual years exceeding the average, so that it is apt to speak of a recent trend of more explicit lyrics. Still, a long-term substantial increase cannot be traced, as the range of the values of sub-period 3 is only slightly (and non-significantly, apart for the year 2015) higher compared to the ones of sub-period 1.

A possible interpretation is a resurgence of gangsta rap (its heyday commonly being situated in sub-period 1) characterized by high profanity rates, or at least a mirroring of its concerns, after a period of relative reduced explicitness in the 2000s. Note again that this finding stands in contrast to recent introspective claims that “many of today’s biggest [HH] artists have taken a gentler approach towards romance even amidst the genre’s misogynistic reputation” (McNulty-Finn 2014).

The comparison to the rates of LYPOP (white dots in Figure 3), which are available for the years 2001–2015, nicely illustrates that profanity is nearly absent from this corpus (average 1.35; range 0–4.77; SD = 1.65), so that in turn it can indeed be seen as a trademark feature of HHD. Interestingly, peaks in profanity rates in LYPOP occur in the sub-period that was found to be comparatively “clean” in LYRAP (sub-period 2, see above), to be exact in the years 2002, 2005, 2010 (and, less markedly, in 2008). This finding is due to the fact that HH albums formed part of LYPOP in these years (see Section 2).

Another important difference between profanity usage in LYRAP and LYPOP is that in the latter relevant terms are used without semantic inversion as discussed above (see (29)).

(29) Feels like I'm doing the same old **shit** (Paolo Nutini: Cherry blossom)

Overall, this suggests that profanity use is largely undesired in mainstream lyrics as represented in LYPOP, while the contrary is the case for HHD as represented in LYRAP, particularly in recent years. A potential rationale for the latter finding would be that, due to the presence of some profanity in commercially successful lyrics (as could be traced in LYPOP during sub-period 2), HH artists felt the need to more strongly emphasize their linguistic “otherness” by using even more profanity in their lyrics as a kind of reaction (sub-period 3) to re-assert their linguistic realness. This interpretation (cf. also McLeod 1999: 138; Álvarez-Mosquera 2015) remains highly tentative, of course, and further sociolinguistic motivations of the resurgence of profanity rates in HHD remain to be investigated. Yet, the contrastive perspective confirmed that profanity can be viewed as one of the trademark features of HHD.

## 4 Conclusion

The present paper showed how a corpus-based approach can shed light on the linguistic side of HH as a complex cultural phenomenon. More specifically, it can be seen as a contribution to extend the sociolinguistic perspective on rap lyrics, and thus as a step toward alleviating the under-researched status of HHD. It complemented previous studies that have taken a purely qualitative approach (such as Aleshinskaya 2013) in that it provided an empirical foundation for accepting (or rejecting) a number of claims in these works. It further went beyond earlier quantitative corpus-based studies (such as Kreyer 2016), particularly in the area of morphosyntax, but also relating to lexicosemantic aspects, relying on the most representative corpus of American HHD to date.

In the corpus-based assessment of HHD (as represented in rap lyrics), a diversified picture emerged. The quantitative analysis first suggested that there is a broad canonical set of morphosyntactic features (largely associated with AAE) used by rappers (Section 3.1), which is testimony to the systematicity of HHD, thus confirming studies such as Alim (2006: 115). However, the results of this analysis further indicated that styling seems to play a part, as a number of artists apparently use an extended number of features (up to 90% of the ones

tested) in their performances, potentially to increase their linguistic realness and thus audience appeal.

Another general finding that emerged is that the notion of HHD being an “anti-language” (Halliday 1976: 570), in the sense of yielding salient patterns deviating from standard usage, can be extended beyond the domains of orthography (Olivo 2001: 68) and lexicon (Dalzell 2014: 18) to morphosyntax. This can be viewed as an additional sphere where the “antihegemonic subtext” (Danesi 2015: 256) of HHD finds its linguistic expression, also relying on grammatical rather than merely on the low-level surface structures mentioned. This is in line with earlier functional views that saw the main purpose of HHD in “demonstrat[ing] the artists’ authenticity and that they are ‘staying street’” (Kreyer 2016: 106; cf. McLeod 1999: 142–143; Alim 2006: 124), and, contrary to earlier claims (cf. Edwards 1998: 143–144), establishes that morphosyntax is another area where HH-related linguistic realness is created.

On a related note, the widespread presence of AAE features across all artists included in LYRAP supports the notion of a kind of linguistic “normative Blackness” (Cutler 2009: 80) in HHD, even though some internal variability in terms of the range of the morphosyntactic features used by individual artists exists. Overall, the evidence indicated largely uniform (>70%) usage of morphosyntactic patterns, which could motivate terms such as “Hip Hop Nation Language” (Alim 2015) as a unified concept. At the same time, it is apparent that HHD can neither be viewed as a monolithic entity nor as a one-to-one mirror image of AAE. The latter finding further supports the “HHD as a specific register” view proposed in Kreyer (2016).

A keyword analysis (Section 3.2), intended to identify the main topics in the first place, yielded a number of additional results. For example, informal spelling variants and a number of highly salient morphosyntactic features (see also Section 3.1) emerged among the high-ranking items, and can thus be seen as characteristic properties of HHD and its integrated lexicogrammatical profile. The present study converges with previous work that has established “sex” and “ghettocentricity” as two broad topic areas, while others, such as “politics”, did not feature prominently in the LYRAP corpus. Furthermore, semantic inversion was exemplified, and the layout of the corpus was exploited to explore diachronic tendencies in profanity usage. In this regard, the evidence substantiated evaluations of high profanity use as a trademark feature of HHD (also markedly in contrast with pop lyrics), while claims of ever-increasing profanity rates (mirroring a “culturally pessimist” view) could only partly be confirmed.

It is evident, though, that the present study is far from exhaustive, and that a multitude of additional perspectives should be taken in future work to arrive at a comprehensive picture of the HH(D) complex.

A case in point pertains to another dimension of internal variability, namely regional differences. A number of observers have emphasized that HH is situated in an area of conflict between the global and the local (Scharenberg 2004: 16; Alim 2009: 5). Accordingly, its adaptability to local contexts may nicely be illustrated through the usage of specific local vocabulary, establishing a local or regional identity (Alim 2015: 853). As already attempted in a number of case studies such as Cramer and Hallett (2012) or Bloomquist and Hancock (2013), the internal heterogeneity of HHD could further be assessed by an analysis contrasting linguistic features of rappers of various regional provenance (see also Cutler 2007: 532), and, with an approach comparable to the one used in the present paper, could be supplemented by a corpus-based perspective.

Other linguistic areas not considered in the analyses above are metaphors (see Schröder 2012) as well as multi-word units/n-grams. While a study of the former could be used to outline linguistic creativity in HHD, the latter may provide further insights into specific discourse markers, collocational patterns of highly frequent items<sup>14</sup> (see Section 3.2) and formulaic patterns in HHD, for example as to ritualized opening and closing sequences (Lüdtke 2007: 265–266; Beers Fägersten 2008: 216–221), as illustrated in (24) above and (30).

(30) **Aiyyo, wassup wassup** let's keep it real son (Nas: Life's a bitch)

Further, in the realm of sweeping (and largely “cultural pessimist”) claims on HHD, a central issue that needs to be tackled is the claim that HHD has become more and more “mainstream” over time (see, e.g. Porter 2001: 151; Bloomquist and Hancock 2013), potentially related to the contention that “the global entertainment industry [...] promotes stylised vernacular performances” (Mair 2013). Here, again, a comparison against pop lyrics may provide adequate cues whether this is the case from a linguistic point of view. The restricted diachronic perspective taken in the present paper does not yield any indication in this direction. On a related note, it may prove interesting to contrast the lyrics as contained in LYRAP with those of “underground” rappers to determine the amount of mainstreaming, that is, marked differences on a linguistic level, of the former. However, the problem of how to identify “underground” lyrics can, in all likelihood, only be solved on a subjective basis.

Ultimately, it has been argued in the present paper that rappers style their lyrics to *achieve* linguistic realness. At the same time, it has to be acknowledged that rappers commonly are active parts of the larger AAE discourse community and thus also *establish* linguistic realness, potentially serving as benchmarks for

---

14 On the lexicogrammatical complexities of profanity usage in AAE see further Spears (2007).

other AAE speakers. However, previous analyses have indicated that the incidence of AAE features is much higher in artists' lyrics than in their casual speech (see, e.g. Edwards and Ash 2004: 170–172 for a case study on Tupac; cf. Green 2002: 160; Alim 2002), so that deliberate styling (reflected in a high incidence of AAE features used) to achieve linguistic realness is a highly likely scenario. Further investigations contrasting different linguistic modes (such as recorded lyrics vs. freestyle lyrics vs. public interviews vs. ethnographic one-on-one interviews; see also Pichler and Williams 2016) on the basis of a wider population of rappers may serve to determine whether this constitutes a universal tendency.

**Acknowledgements:** I would like to thank Kristy Beers Fägersten, Fabian Vetter, three anonymous reviewers, and Srikant Sarangi for helpful advice on the content and structure of this paper. First ideas were discussed at the “Pop culture and linguistics” workshop featuring as part of the fourth meeting of the International Society for the Linguistics of English (*English in a Multilingual World: Challenges and Perspectives*) held in Poznań, Poland, in September 2016. Travel to this meeting was supported by a mobility grant provided by the Bavarian Academic Center for Central, Eastern and Southeastern Europe (BAYHOST).

## Appendix A

Overview of song and token numbers for LYRAP artists.

Artist	n songs	n tokens
50 Cent	288	147,684
Drake	209	103,886
Eminem	279	190,719
Jay Z	263	151,283
Kanye West	161	80,938
Lil' Wayne	643	350,123
Ludacris	158	86,270
Nas	263	146,099
Nicki Minaj	181	75,914
Notorious B.I.G.	88	51,398
Snoop Dogg	311	161,611
T.I.	219	134,059
Tupac	246	164,063
Tyga	253	111,666
Wiz Khalifa	349	143,930

## Appendix B

List of morphosyntactic features used for the characterization of HHD.

Feature	Source(s)	Area
<i>Me</i> instead of <i>I</i> in coordinate subjects	eWAVE 7	pronouns
Regularized reflexives paradigm	eWAVE 11	pronouns
Forms or phrases for the second person plural pronoun other than <i>you</i>	eWAVE 34	pronouns
Associative plural marked by postposed <i>and them/ them all/dem</i>	eWAVE 52	noun phrase
Plural marker <i>'em</i>	Lüdtke (2007)	noun phrase
<i>Dem/them</i> for demonstrative <i>those</i>	eWAVE 68	noun phrase
<i>They</i> for <i>their</i>	Smitherman (2000)	noun phrase
Double comparatives and superlatives	eWAVE 78	noun phrase
Regularized comparison strategies: extension of synthetic marking	eWAVE 79	noun phrase
Regularized comparison strategies: extension of analytic marking	eWAVE 80	noun phrase
Addition of <i>s</i> to indicate habituality	Jansen (2012)	tense & aspect
Invariant <i>be</i> as habitual marker	eWAVE 90	tense & aspect
Future marker <i>gon(na)</i>	Smitherman (2000); Edwards and Ash (2004)	tense & aspect
Completive/perfect <i>done</i>	Edwards (1998); eWAVE 104	tense & aspect
Future marker <i>finna/fitna/fitna</i>	Jansen (2012)	tense & aspect
Anterior/perfective <i>been/bin</i>	Lüdtke (2007); eWAVE 111	tense & aspect
<i>Here go/there go</i> in a non-movement sense	Jansen (2012)	tense & aspect
Double modals	Lüdtke (2007); eWAVE 121	modality
Contracted modals (modal + <i>to/have</i> ) <i>woulda/coulda/mighta/liketa/posedta</i>	Lüdtke (2007)	modality
Multiple negation/negative concord	Aleshinskaya (2013); eWAVE 154	negation
<i>Ain't but/don't but for only</i>	Lüdtke (2007)	negation
<i>Ain't</i> as negator	eWAVE 155–157	negation
Invariant <i>don't</i>	eWAVE 158	negation

(continued)



(continued)

Feature	Source(s)	Area
Invariant present tense forms due to zero marking for the third person singular	eWAVE 170	agreement
Existential/presentational <i>there's/there is/there was</i> with plural subjects	eWAVE 172	agreement
Variant forms of dummy subject <i>there</i> in existential clauses	eWAVE 173	agreement
Copula absence before progressive	eWAVE 174	agreement

## References

- Aleshinskaya, Evgeniya. 2013. Key components of musical discourse analysis. *Research in Language* 11(4). 423–444.
- Alim, H. Samy. 2002. Street-conscious copula variation in the hip hop nation. *American Speech* 77(3). 288–304.
- Alim, H. Samy. 2006. *Roc the mic right: The language of hip hop culture*. London: Routledge.
- Alim, H. Samy. 2009. Intro: Straight outta Compton, straight aus München – Global linguistic flows, identities, and the politics of language in a global hip hop nation. In H. Samy Alim, Awad Ibrahim & Alastair Pennycook (eds.), *Global linguistic flows: Hip hop cultures, youth identities, and the politics of language*, 1–22. London: Routledge.
- Alim, H. Samy. 2015. Hip hop nation language: Localization and globalization. In Jennifer Bloomquist, Lisa J. Green & Sonja L. Lanehart (eds.), *The Oxford handbook of African American language*, 850–862. Oxford: Oxford University Press.
- Álvarez-Mosquera, Pedro. 2015. Underlining authenticity through the recreolization process in rap music: A case of an in-group answer to an identity threat. *Sociolinguistic Studies* 9(1). 51–70.
- Anthony, Laurence. 2014. *AntConc* (Version 3.4.3). Tokyo: Waseda University.
- Armstrong, Edward G. 2004. Eminem's construction of authenticity. *Popular Music and Society* 27(3). 335–355.
- Beers Fägersten, Kristy. 2008. A corpus approach to discursive constructions of a hip-hop identity. In Anneli Ädel & Randi Reppen (eds.), *Corpora and discourse: The challenges of different settings*, 211–240. Amsterdam: Benjamins.
- Bell, Alan. 1984. Language style as audience design. *Language in Society* 13. 145–204.
- Bloomquist, Jennifer & Isaac Hancock. 2013. The dirty third: Contributions of Southern hip hop to the study of regional variation within African American English. *Southern Journal of Linguistics* 37(1). 1–27.
- Bucholtz, Mary. 2011. *White kids: Language, race, and styles of youth identity*. Cambridge: Cambridge University Press.
- Ch'ien, Evelyn. 2011. Creative technology and rap. *World Englishes* 30(1). 60–75.
- Connor, Martin. 2018. *The musical artistry of rap*. Jefferson: McFarland.
- Coupland, Nikolas. 2001. Dialect stylization in radio talk. *Language in Society* 30(3). 345–375.

- Coupland, Nikolas. 2014. Language, society and authenticity: Themes and perspectives. In Véronique Lacoste, Jakob Leimgruber & Thiemo Breyer (eds.), *Indexing authenticity: Sociolinguistic perspectives*, 14–39. Berlin: Mouton de Gruyter.
- Cramer, Jennifer & Jill Hallett. 2012. From Chi-Town to the Dirty-Dirty: Regional identity markers in US hip hop. In Marina Terkourafi (ed.), *The languages of global hip-hop*, 256–276. London: Continuum.
- Culpeper, Jonathan. 2009. Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics* 14(1). 29–59.
- Cutler, Cecilia. 2007. Hip-hop language in sociolinguistics and beyond. *Language and Linguistics Compass* 1(5). 519–538.
- Cutler, Cecilia. 2009. 'You shouldn't be rappin', you should be skateboardin' the X-Games': The coconstruction of Whiteness in an MC battle. In H. Samy Alim, Awad Ibrahim & Alastair Pennycook (eds.), *Global linguistic flows: Hip hop cultures, youth identities, and the politics of language*, 79–94. London: Routledge.
- Cutler, Cecilia. 2014. *White hip-hoppers, language, and identity in post-modern America*. London: Routledge.
- Dalzell, Tom. 2014. Hip-hop slang. In Julie Coleman (ed.), *Global English slang: Methodologies and perspectives*, 15–24. London: Routledge.
- Danesi, Marcel. 2015. *Popular culture: Introductory perspectives*. Lanham: Rowman & Littlefield.
- Eble, Connie. 2004. Slang. In Edward Finegan & John R. Rickford (eds.), *Language in the USA: Themes for the twenty-first century*, 375–386. Cambridge: Cambridge University Press.
- Eckert, Penelope. 2004. The meaning of style. *Texas Linguistics Forum* 47. 41–53.
- Edwards, Walter F. 1998. Sociolinguistic features of rap lyrics: Comparisons with reggae. In Pauline G. Christie, Barbara Lalla, Velma Pollard & Lawrence Carrington (eds.), *Studies in Caribbean Language, Volume II: Papers from the ninth biennial conference of the Society of Caribbean Linguistics, 1992*, 128–146. St Augustine: Society for Caribbean Linguistics.
- Edwards, Walter F. 2002. From poetry to rap: The lyrics of Tupac Shakur. *The Western Journal of Black Studies* 26(2). 61–70.
- Edwards, Walter F. & Leslie Ash. 2004. AAVE features in the lyrics of Tupac Shakur: The notion of 'realness'. *Word* 55(2). 165–178.
- Garside, Roger & Nicolas Smith. 1997. A hybrid grammatical tagger: CLAWS4. In Roger Garside, Geoffrey Leech & Tony McEnery (eds.), *Corpus annotation: Linguistic information from computer text corpora*, 102–121. London: Longman.
- Green, Lisa J. 2002. *African American English: A linguistic introduction*. Cambridge: Cambridge University Press.
- Halliday, M. A. K. 1976. Anti-languages. *American Anthropologist* 78(3). 570–584.
- Jansen, Robert. 2012. Blinglish: Ein sprachliches Phänomen. *Lebende Sprachen* 57(2). 329–399.
- Kautny, Oliver. 2015. Lyrics and flow in rap music. In Justin A. Williams (ed.), *The Cambridge companion to hip-hop*, 101–117. Cambridge: Cambridge University Press.
- Kimminich, Eva. 2004. Eine Zwischenbilanz: Drei Jahrzehnte Hip Hop – Drei Jahrzehnte Hip Hop-Forschung. In Eva Kimminich (ed.), *Rap: More than words*, vii–xxvi. Frankfurt: Lang.
- Kirkland, David E. 2015. Black masculine language. In Jennifer Bloomquist, Lisa J. Green & Sonja L. Lanehart (eds.), *The Oxford handbook of African American language*, 834–849. Oxford: Oxford University Press.
- Komaniecki, Robert. 2017. Analyzing collaborative flow in rap music. *MTO: Journal of the Society for Music Theory* 23(4). 1–19.

- Kreyer, Rolf. 2016. 'Now niggas talk a lotta Bad Boy shit': The register hip-hop from a corpus-linguistic perspective. In Christoph Schubert & Cristina Sanchez-Stockhammer (eds.), *Variational text linguistics: Revisiting register in English*, 87–109. Berlin: Mouton de Gruyter.
- Lena, Jennifer C. 2016. Authenticity and independence in rap music and other genre communities. In Sara Towe Horsfall, Jan-Martijn Meij & Meghan D. Probstfield (eds.), *Music sociology: Examining the role of music in social life*, 232–240. London: Routledge.
- Lüdtke, Solveig. 2007. *Globalisierung und Lokalisierung von Rapmusik am Beispiel amerikanischer und deutscher Raptexte*. Münster: LIT.
- Mair, Christian. 2013. World Englishes and corpora. In Markku Filppula, Juhani Klemola & Devyani Sharma (eds.), *The Oxford handbook of World Englishes*. Oxford: Oxford University Press. <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199777716.001.0001/oxfordhb-9780199777716-e-008> (accessed 12 January 2017).
- McLeod, Kembrew. 1999. Authenticity within hip-hop and other cultures threatened with assimilations. *Journal of Communication* 49(4). 134–150.
- McNair, Janis & John Powles. 2005. Hippies vs. hip-hop heads: An exploration of music's ability to communicate an alternative political agenda from the perspective of two divergent musical genres. In Dorothy Miell, Raymond A. R. MacDonald & David J. Hargreaves (eds.), *Musical communication*, 339–366. Oxford: Oxford University Press.
- McNulty-Finn, Clara. 2014. The evolution of rap. *Harvard Political Review Special Issue: The Beat of the Drum*, April. <http://harvardpolitics.com/covers/evolution-rap/> (accessed 20 June 2016).
- Mitchell, Tony. 2015. African, African American, Middle Eastern and French hip hop. In Andy Bennett & Steve Waksman (eds.), *The SAGE handbook of popular music*, 227–242. London: SAGE.
- Morgan, Marcyliena H. 2001. 'Nuthin' but a G thang': Grammar and language ideology in hip hop identity. In Sonja L. Lanehart (ed.), *Sociocultural and historical contexts of African American English*, 187–209. Amsterdam: Benjamins.
- Motschenbacher, Heiko. 2016. A corpus linguistic study of the situatedness of English pop song lyrics. *Corpora* 11(1). 1–28.
- Newman, Michael. 2005. Rap as literacy: A genre analysis of hip-hop ciphers. *Text* 25(3). 399–436.
- O'Hanlon, Renae. 2006. Australian hip hop: A sociolinguistic investigation. *Australian Journal of Linguistics* 26(2). 193–209.
- Ochmann, Matthäus. 2013. The notion of authenticity in international hip-hop culture. In Sina A. Nitzsche & Walter Grünzweig (eds.), *Hip-hop in Europe: Cultural identities and transnational flows*, 423–446. Münster: LIT.
- Olivo, Warren. 2001. Phat lines: Spelling conventions in rap music. *Written Language & Literacy* 4(1). 67–85.
- Pennycook, Alastair. 2008. *Global Englishes and transcultural flows*. London: Routledge.
- Pichler, Pia & Nathanael Williams. 2016. Hipsters in the hood: Authenticating indexicalities in young men's hip-hop talk. *Language in Society* 45(4). 557–581.
- Pojanapunya, Punjaporn & Richard Watson Todd. 2018. Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory* 14(1). 133–168.
- Porter, Russel A. 2001. Soul into hip-hop. In Simon Frith, Will Straw & John Street (eds.), *The Cambridge companion to pop and rock*, 143–157. Cambridge: Cambridge University Press.
- Queen, Robin. 2013. Working with performed language: Movies, television, and music. In Christine Mallinson, Becky Childs & Gerard van Herk (eds.), *Data collection in sociolinguistics: Methods and applications*, 217–227. London: Routledge.

- Richards, Abigail V. 2017. A qualitative study of group therapy incorporating rap music with inmates. Master's thesis, Missouri State University.
- Richardson, Elaine. 2006. *Hiphop literacies*. London: Routledge.
- Richardson, Jeanita W. & Kim A. Scott. 2002. Rap music and its violent progeny: America's culture of violence in context. *Journal of Negro Education* 71(3). 175–192.
- Ross, Andrew S. & Damian J. Rivers (eds). 2018. *The sociolinguistics of hip-hop as critical conscience: Dissatisfaction and dissent*. Cham: Palgrave Macmillan.
- Scharenberg, Albert. 2004. Globalität und Nationalismus im afro-amerikanischen Hip Hop. In Eva Kimminich (ed.), *Rap: More than words*, 13–44. Frankfurt: Lang.
- Schröder, Ulrike. 2012. Applying conceptual metaphor and Blending Theory to culture-specific speech functions in rap lyrics. *Text & Talk* 32(2). 211–233.
- Smitherman, Geneva. 2000. *Talkin that talk: Language, culture, and education in African America*. London: Routledge.
- Spears, Arthur K. 2007. African American communicative practices: Performativity, semantic license, and augmentation. In John Baugh & H. Samy Alim (eds.), *Talkin black talk: Language, education, and social change*, 100–111. New York: Columbia Teachers College Press.
- Terkourafi, Marina (ed.). 2012a. *The languages of global hip-hop*. London: Continuum.
- Terkourafi, Marina. 2012b. Introduction. In Marina Terkourafi (ed.), *The languages of global hip-hop*, 1–18. London: Continuum.
- Werner, Valentin. 2012. Love is all around: A corpus-based study of pop music lyrics. *Corpora* 7(1). 19–50.
- Werner, Valentin. 2018. Linguistics and pop culture: Setting the scene(s). In Valentin Werner (ed.), *The language of pop culture*, 3–26. New York: Routledge.
- Werner, Valentin. Forthcoming. Lyrics and language awareness. *Nordic Journal of Modern Language Methodology* 7(1).
- Werner, Valentin & Maria Lehl. 2015. Pop lyrics and language pedagogy: A corpus-linguistic approach. In Federica Formato & Andrew Hardie (eds.), *Corpus linguistics 2015*, 341–343. Lancaster: UCREL.
- Widawski, Maicej. 2015. *African American slang: A linguistic description*. Cambridge: Cambridge University Press.
- Wolfram, Walt. 2013. Urban African American Vernacular English. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://ewave-atlas.org/languages/15> (accessed 23 May 2016).

## Bionote

### Valentin Werner

Valentin Werner is assistant professor of English and Historical Linguistics at the University of Bamberg, Germany. He mainly works in the areas of applied and variational linguistics, focusing on the language of pop culture, the corpus-based analysis of learner language, and the corpus- and questionnaire-based study of varieties of English. He has published two books on the present perfect (UBP, 2014; Mouton de Gruyter, 2016) and a volume on the language of pop culture (Routledge, 2018).