

Σύγκριση/Comparaison/Comparison

Αρ. 34 (2025)

Ψηφιακές Λογοτεχνικές Σπουδές



The Lexicon of Greek Rap

Panagiotis Delis

Copyright © 2026, Panagiotis Delis



Άδεια χρήσης [Creative Commons Attribution-NonCommercial-ShareAlike 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Βιβλιογραφική αναφορά:

Delis, P. (2026). The Lexicon of Greek Rap: a corpus-stylistic study. *Σύγκριση/Comparaison/Comparison*, (34), 115–138. ανακτήθηκε από <https://ejournals.epublishing.ekt.gr/index.php/sygkrisi/article/view/42779>

PANAGIOTIS DELIS

National and Kapodistrian University of Athens

The Lexicon of Greek Rap: A Corpus-Stylistic Study

1. Introduction

Rap, as it has developed internationally, can be defined as a complex oral art form embedded within strict or more elastic rhythmic frameworks often accompanied by musical excerpts through sampling (a technique that borrows a segment of an existing piece and reuses it in a new composition), percussive sound patterns, and electronic timbres. Beyond its evident entertainment and artistic dimensions, rap performs functions that range from political protest and social critique to the articulation of personal experience and the construction of collective identities. Within the broader culture of hip-hop, which also comprises DJing, breakdance, and graffiti, rap constitutes a nexus between orality and wider aesthetic, technological, and ideological developments (Werner, 2019). Within Digital Literary Studies, rap provides a particularly productive test case because its poetic organisation is anchored in performance and interaction. A digitally assisted, corpus-stylistic approach, supported by digital tools, moves beyond close reading and still captures genre features such as orality, direct address and flow. In this article lyrics are treated as a poetic discourse whose stylistic patterning, namely lexical sets, clause patterns, and performance cues, can be measured across time and then interpreted culturally. This dual commitment to quantitative mapping and qualitative, context-aware interpretation underpins the contribution of the study.

For a more comprehensive understanding of the genre, it is necessary to foreground several fundamental dimensions of this musical phenomenon. Rhythm and the beat are constitutive for hip-hop production, as rap is rarely encountered without pronounced rhythmic patterns, which function as the principal mechanism for anchoring speech to music. A second basic structural component is flow and delivery, terms that capture each rapper's subjective "signature", i.e. the manner in which words are aligned to the beat and manipulated through prosody, pauses, and emphases; in short, the way the artist articulates the lyrics. Another characteristic is the relatively limited melodic content, insofar as the lyrics and their semantic load carry greater weight. Although in more recent iterations (e.g., melodic rap, R&B crossovers) sung delivery has gained ground, the declamatory character remains a structural constant. A further defining feature of rap is its interactional "contact layer" (Morgan, 2009), including overt pronominal reference (I/you), vocatives, imperatives, discourse particles, and short interjections that keep the performer and audience in continuous alignment. This is especially salient in Greek, which is a pro-drop language where null subjects are the default; when speakers choose overt subject pronouns, they typically mark emphasis, contrast, or a shift in topicality. Therefore, the frequent surfacing of first- and second-person forms in rap are read as deliberate stance-marking and proximity work rather than mere grammar (Holton, Mackridge, & Philippaki-Warburton, 2012). Vocatives and directives likewise compress social distance, building the impression of live exchange even in recorded tracks.

With respect to rap's structural organisation, a typical track follows canonical templates such as verse-chorus-verse, which accommodate recurring refrains in which the central message or "hook" of the piece is often encoded. Sampling and collaborations enhance the fluidity of discourse and the genre's intertextuality through interaction with earlier tracks (samples) and with multiple artists (features/featurings). Linguistically, the genre is marked by flexibility, as via metaphor and wordplay, as well as rhyme and slang, it exhibits pronounced linguistic inventiveness, an aspect even more salient in improvisations (freestyles), an art regarded as constitutive of rap, particularly in live settings, where the performer's ability to compose rhythmic lines on the spot is foregrounded. Finally, a cornerstone of the genre is the artist-audience interaction (Baym, 2018), realised and intensified through the audience's active participation, whether through vocal exchanges (call-and-response) or through reactions that feed back into the performance.

Rap in the United States effectively first appears in the late 1970s within Bronx communities (Rose, 1994; Krims, 2000). As rap artists achieve unprecedented commercial success during the 1980s and 1990s, the genre's so-called "golden age" takes shape, with currents ranging from highly charged forms such as politicized rap (Public Enemy) and gangsta rap (N.W.A.) to more lyrical explorations (Nas, *A Tribe Called Quest*). In parallel, rap "migrates" rapidly abroad, influencing the musical cultures of many countries. These circulations are not merely musical but infrastructural, as distribution platforms, visual grammars, and scene-internal authenticity discourses travel alongside sounds and lexicons (Terkourafi, 2010; Pennycook, 2007; Cutler, 2007). International flows thus shape local stylistic repertoires both directly, through loanwords, ad-libs, brand/persona references) and indirectly, through expectations about flow, hook design, or persona work.

In Greece, rap emerges on a mass scale in the 1990s through the first Greek-language groups such as Terror X Crew, Active Member, and others. The lyrics of the period often focus on social, political, and class issues, drawing on Athenian slang and orality, while the scene displays multiple faces, on the one hand underground formations, and on the other the first efforts toward commercial recognition. In this initial decade the hip-hop community is also consolidated, acquiring its own characteristics and practices that revolve around a nonconformist and insurgent way of life, with music at its centre. In the 2000s, the field splinters into further currents, as some foreground a "tougher" thematic orientation associated with the unease of those directly or indirectly involved in rap with the social conjuncture, while others pursue more lyrical or overtly emotional form, while at the same time, commodification becomes pronounced, the genre gains public visibility, and is no longer confined to the obscurity of the underground scene. The 2010s find the genre in a transitional state with respect to structure and style, in which the incorporation of trap and the rise of streaming services configure a new condition of globalisation. Consequently, elements such as references to luxury brands, foreign-language refrains, and the online promotion of tracks assume a central role. In what follows rap is approached as a form of poetic discourse. This stance aligns with a substantial francophone tradition that reads rap as poiesis in its own right, attentive to lineation, sonic patterning, intertextual play and performance voice. Studies have traced rap's ludic and metapoetic operations, its reworking of canonical repertoires, and the pragmatic staging of the

emcee's ethos, thereby motivating literary methods alongside sociolinguistic ones (Carinos & Hammou, 2017; Bonin & Krastev-McKinnon, 2019; Labourie, 2017; Pecqueux, 2007; Rolland, 2021). The present study leverages a decade-by-decade Greek rap corpus (1990s, 2000s, 2010s) and situates it against two Greek reference baselines, namely the general language, and the broader song-lyrics, to separate rap-specific stylistic signals from features of Greek lyric discourse at large. In addition, a decade-stratified U.S. rap corpus provides a comparative horizon. This configuration enables us to trace both the stability of Greek rap's interactional infrastructures and the re-weighting of stylistic resources associated with trap aesthetics, platform-mediated promotion, and conspicuous brand discourse during the 2010s.

In sum, what is argued is that Greek rap's stylistic evolution is best understood as a diachronic re-balancing of small but powerful units, namely clitics and pronouns, particles, vocatives, and brief ad-libs, onto which changing aesthetic and commercial ecologies are grafted. The article proceeds as follows: Section 2 locates the study within stylistics and DLS, Section 3 details corpora and methods, Section 4 presents findings for the 1990s, 2000s, and 2010s, Section 5 offers a comparative discussion with U.S. rap, and Section 6 concludes with implications and future directions.

2. Toward a stylistic approach to rap

Although rap is conceptualised as a highly popular musical genre and has traced a significant history in Greece, academic research focusing on its linguistic and stylistic dimension remains particularly limited for both Greek-language rap (Delis, 2024) and English-language rap (Werner, 2019). The need for a more extensive, systematic mapping of the lexicon and the diachronic changes of Greek rap, using tools from corpus stylistics, is especially evident. Framing rap as poetic discourse allows corpus stylistics to query features classically associated with poetry, including patterned repetition, rhyme, parallelism and intertextual citation, while remaining sensitive to performance. Francophone scholarship has made this case explicitly, examining rap's poetic procedures, its dialogic ethos and its systematic reworking of literary traditions (Carinos & Hammou, 2017; Labourie, 2017; Pecqueux, 2007; Rolland, 2021). The present paper undertakes an initial attempt to address this need by focusing on the following research questions:

1. What is the stylistic profile of Greek rap, as revealed by corpus-based analysis?
2. How are its stylistic parameters differentiated across the 1990s, 2000s, and 2010s?
3. In what way do developments in Greek rap converge with or diverge from American rap, in the light of the globalized music industry?

In adopting a stylistic perspective, *style* is conceptualised as a recurrent linguistic patterning that is functionally motivated within a discourse community (McIntyre & Walker, 2019). Corpus stylistics links measurable patterns to interpretations of voice, stance and genre, which suits rap well because interactional and performance cues recur in short, frequent words and phrases (Biber, 2011; Werner, 2019). Central to the analysis is the notion of the corpus, understood as the systematic collection of large quantities of data on the basis of defined principles and criteria (McEnery & Hardie, 2012; Goutsos & Fragkaki, 2015). In practice, this entails principled sampling,

transparent inclusion/exclusion criteria, and basic normalisation steps (tokenisation, lowercasing, handling of punctuation), so that downstream measures are comparable across subcorpora and over time (McIntyre & Walker, 2019). Corpus stylistics constitutes a rapidly developing subfield of linguistics that mobilises the application of theories, models, and research frameworks from stylistics to the analysis of corpora and is situated at the intersection of stylistics and digital approaches to literature (McIntyre & Walker, 2019). The techniques of corpus stylistics have been applied more frequently to prose (fiction, articles, newspapers), whereas the analysis of poetry with digital tools remains an area that has not yet been adequately explored (Škrabal & Piorecký, 2022). The rap-as-poetry debate is established in francophone research which documents rap's poetic and intertextual operations (Carinos & Hammou, 2017; Bonin & Krastev-McKinnon, 2019; Labourie, 2017; Rolland, 2021), though, what remains under-developed is large-scale corpus-based mapping for Greek rap specifically. This relative under-exploration is noteworthy given that rap lyrics share with poetry lineation, foregrounding, and patterned repetition, yet add a performance layer that is highly amenable to systematic mapping through concordances and collocation networks (McIntyre & Walker, 2019; Werner, 2019). Although the datasets that would enable robust distant reading of contemporary poetry remain relatively limited (Moretti, 2013), rap, by virtue of its rhythmic and poetic nature, can offer a fertile field of inquiry when treated through a carefully curated, corpus-based lens.

The methods of corpus stylistics consist in a first-order quantification through the creation of keyword lists, which foreground words that are statistically more likely to appear in a specialised corpus when it is contrasted with a reference corpus. In the present usage, keyness refers to over- or under-representation of lexical items in a target corpus relative to a reference, typically assessed with standard statistics (e.g., log-likelihood) and then interpreted functionally via concordance contexts (Biber, 2011; McIntyre & Walker, 2019). Collocation is examined with symmetric windows and association measures (e.g., MI), allowing us to recover local semantic prosodies and interactional routines that raw frequencies alone might obscure (Fischer-Starcke, 2009). Subsequently, one may undertake keyword-list analysis, in which semantic fields are identified, i.e. sets of words that express semantically related dominant concepts in the data, so as to decode the meaning functions and linguistic patterns with which they are associated (Fischer-Starcke, 2009; Biber, 2011; Culpeper, 2009). For rap specifically, we pay particular attention to what we term the contact layer, overt person reference, vocatives, imperatives, interjections, and discourse particles, which realises addressivity and turns recorded lyrics into simulated live interaction. This section situates the study and delineates the research gap. Greek rap across the 1990s, 2000s and 2010s is analysed using corpus-stylistic methods. Two Greek reference corpora are employed to distinguish rap-specific features from those shared with Greek song and with general Greek usage. A decade-stratified English-language rap corpus is additionally used for cross-lingual comparison. Within this scope, the most relevant literature is summarised. Prior work on hip-hop discourse has shown how shout-outs, direct address, and participation frameworks function as indexical devices that align performers and audiences (Morgan, 2009; Rose, 1994; Alim & Smitherman, 2012), a set of insights that corpus methods can now scale across decades. In the present study, keyword lists, concordances, and collocates are employed in order

to identify with confidence the principal linguistic and conceptual cores of each decade (Biber, 2011; Scott & Tribble, 2006). Two genre-specific challenges are addressed during preprocessing, on the one hand code-mixing, with Greek–English slang, brand names, ad-libs, and on the other hand orthographic variability, with transliteration and creative spellings. Rather than eliminating these phenomena, they are treated as stylistically meaningful. The stylistic dimension lies in the subsequent interpretation of how the numerical and statistical findings contribute to a literary and cultural reading of the genre. Interpretation proceeds iteratively, as quantitative signals motivate targeted concordance reading, and qualitative insights feed back into category refinement, which is essentially a cyclical workflow standard in corpus stylistics (McIntyre & Walker, 2019). This article contributes three things. First, it provides a decade-by-decade, corpus-based account of Greek rap that keeps the performance layer in view through a focus on address, vocatives, particles and ad-libs. Second, it employs two Greek baselines, the general language and the song-lyrics field, to separate rap-specific signals from features common to Greek song. Third, it compares the Greek results with a decade-stratified English-language corpus to show where they converge and where they diverge, while keeping local histories intact and not forcing a single global model. The approach remains bottom-up, as categories are induced from keyword and collocation signals and are refined through concordance reading, yielding a portable description of layered stylistic resources and a reproducible workflow that can be adapted to other languages and scenes.

3. Data and Methodology

For the present study, decade-stratified corpora focused on Greek rap (GreekRapCorpus) for the 1990s, 2000s, and 2010s, with a total size of 224,726 tokens were compiled. The GreekRapCorpus comprises officially released Greek-language rap tracks distributed across three decades (1990s/2000s/2010s). Selection followed principled inclusion criteria (rap as primary genre; Greek as main lyric language; studio or commercially released live recordings), while duplicates, re-issues without substantive lyrical changes, and non-rap cross-overs were excluded. Lyrics were transcribed from publicly available sources and cross-checked against audio where feasible to reduce paratextual noise (stage banter, metadata, credits). This configuration seeks representativeness at the level of stylistic diversity rather than artist completeness, consistent with corpus design recommendations for stylistic analysis (McIntyre & Walker, 2019). For comparison with American rap, corresponding corpora of English-language rap lyrics (EnglishRapCorpus) were collected, likewise organised by decade (from the 1960s to the 2010s), totalling 655,621 tokens (Delis, 2019). In parallel, two reference corpora for Greek were used, namely the Corpus of Greek Texts (CGT) (Goutsos, 2010), of approximately 30 million words covering a range of text types, and a song-lyrics corpus drawn from stixoi.gr (SongLyricsCorpus), of approximately 2 million words, in order to distinguish the features that set rap apart from the broader lyric tradition.

Table 1. *Corpora and Subcorpora Employed in the Study: Composition, Size, and Analytical Roles*

Corpus/Subcorpus	Language	Period covered	Size (tokens)	Composition	Analytical role
GreekRapCorpus '90s	Greek	1990s	~72,902	Greek rap lyrics (decade-stratified)	Target corpus for decade analysis
GreekRapCorpus '00s	Greek	2000s	~83,154	Greek rap lyrics (decade-stratified)	Target corpus for decade analysis
GreekRapCorpus '10s	Greek	2010s	~68,670	Greek rap lyrics (decade-stratified)	Target corpus for decade analysis
English-Rap-Corpus (ERC)	English	1960s–2010s	655,621 (total)	English-language rap lyrics, decade-	Cross-lingual comparator by decade

				stratified	
Corpus of Greek Texts (CGT)	Greek	Mixed (Modern Greek)	~30,000,000	General reference corpus (multiple genres)	General - language baseline (Goutsos, 2010)
Greek Song-Lyrics Corpus (SLC)	Greek	Mixed	~2,000,000	Greek song lyrics from stixoi.gr	Song-lyrics baseline to separate rap-specific signals

To ensure comparability across subcorpora, basic preprocessing was applied uniformly: tokenisation, lowercasing, and punctuation normalisation. Greek diacritics were preserved, while contractions and clitics were left as surface forms to retain stylistically meaningful variation (e.g., *μου/σου* in clitic vs possessive use), and no lemmatisation was performed so as not to collapse rhyme-relevant morphology (Biber, 2011). Code-mixing (Greek–English items, brand names, ad-libs) was retained as data rather than filtered, given its analytical salience in rap discourse and performance (McIntyre & Walker, 2019). A keyword is defined as a word that is statistically over-represented in a target corpus relative to a reference corpus. Keyness was computed with log-likelihood at $p < .001$ and a minimum frequency threshold of five in order to limit sparse-data artefacts (Rayson & Garside, 2000). A concordance (KWIC) is a list of all instances of a word with a fixed window of co-text on either side and is used to inspect usage patterns and to guard against selective quotation. Collocation was

examined within a symmetric window of four words to the left and four words to the right. Association strength was estimated with mutual information and log-likelihood, and patterns were then interpreted through the concordance lines (Church & Hanks, 1990; Evert, 2005; Fischer-Starcke, 2009; Brezina, McEnery, & Wattam, 2015; McEnery & Hardie, 2012). Two Greek baselines were employed for keyness, namely the Corpus of Greek Texts, which represents general written usage across genres, and the Song-Lyrics Corpus, which represents the wider field of Greek popular lyrics. The first comparison foregrounds departures from general written norms and makes the oral texture salient, while the second highlights what is distinctive of rap within the ecology of Greek songs. For each decade of the GreekRapCorpus two keyword profiles were computed against each baseline. Signals that persist in both comparisons are taken as high-confidence rap features. Signals that surface only against the general baseline are interpreted as broadly lyric. Signals that surface only against the song-lyrics baseline are interpreted as rap-specific within the song field. This two-baseline design avoids attributing to rap what may in fact belong to Greek lyrics in general and provides a clearer map of the style space. Finally, the decade-specific profiles for the GreekRapCorpus were compared with the decade-stratified EnglishRapCorpus to contextualise cross-lingual convergence and divergence.

For the digital analysis (extraction of keywords, frequencies, and collocations) AntConc (Anthony, 2019) was utilised, as it allows the generation of concordance lines and the automatic retrieval of the most frequent items. At a second stage, the results were interpreted from the perspectives of stylistics and sociolinguistics. For this qualitative analysis, all concordance lines of very high-frequency items (e.g., *μου/σου/είμαι*) were surveyed, and the same process was followed for lower-frequency but analytically salient items (e.g., ad-libs, brand/persona lexemes). Rather than imposing a priori categories from the literature, analytic bundles were induced from the data, with decade-specific keyword lists and collocation profiles in the GreekRapCorpus (with CGT/SLC as references) revealing recurrent lexical sets. From these signals we operationalised: (i) interpersonal address (first/second-person pronouns, vocatives, imperatives, discourse particles), (ii) craft/scene lexicon (e.g., *rap*, *MC*, *ρίμα* ‘rhyme’, *μικρόφωνο* ‘microphone’), (iii) ad-libs and short interjections (e.g., *yah*, *hey*, *whoa*), (iv) brands/personae/money talk (luxury labels, celebrity personae, denominations), (v) scene framing (time/space deictics such as *απόψε* ‘tonight’, *νύχτα* ‘night’, *club*, *γειτονιά* ‘neighbourhood’), and (vi) aggression/face-threat resources.

4. Findings and Analysis

Section 4 presents the findings of the corpus-stylistic analysis of Greek rap. Each subsection examines one decade, namely the 1990s, 2000s, and 2010s, using keyword lists, concordances, and collocation profiles. For each decade the results are reported against two reference baselines, the Corpus of Greek Texts and a Greek Song-Lyrics Corpus, so that broadly lyric features can be separated from rap-specific signals. The discussion integrates quantitative evidence with close reading and focuses on address, craft talk, performance vocables, and the commodity lexicon. After the decade-by-decade analysis, a short sub-section isolates what remains distinctive of rap when compared with other Greek songs, and a synthesis traces cross-decade trajectories

and the reweighting of stylistic resources. The section closes with a bridge to the cross-lingual comparison with English-language rap in Section 5.

4.1. GRC '90: Emerging motifs and local anchors

In the Greek rap subcorpus of the 1990s (approximately 72,902 words), central keywords such as *μου*-me/my (rank 1), *σου*-you/your (rank 2), *είμαι*-am (rank 8), *εγώ*-I (rank 9), *μαλάκα*-jerk (rank 19) decisively differentiate the texts from the Corpus of Greek Texts (CGT). In a pro-drop language such as Greek, the overt surfacing of first- and second-person forms is pragmatically marked: their over-representation in the 1990s subcorpus signals stance, proximity, and an insistence on dyadic alignment with the addressee, rather than mere grammatical necessity (cf. Biber, 2011; Werner, 2019). The frequent use of *μου* underscores the artist's connection to personal experience, as depicted in *η καρδιά μου* (my heart), *οι ρίμες μου* (my rhymes), *η ζωή μου* (my life), thereby configuring a confessional tone or an autobiographical narrative. *Μου* functions as an index of self-awareness, allowing the artist to articulate internal processes while simultaneously asserting an individual identity vis-à-vis the audience and prevailing social conditions. In this way, the listener apprehends the rapper's personal trajectory and experiences as an essential component of artistic creation. With *σου*, the artist addresses the listener directly, seeking a dialogue or a virtual conversation. This rhetorical technique intensifies immediacy and interpersonal connection, rendering the listener an active co-participant in the narrative and cultivating a sense of proximity in which the listener can identify with, or feel personally implicated in, what is being said. Concordance reading shows that *μου/σου* frequently occur within directive and evaluative frames, e.g., *πες μου* 'tell me' / *δώσε μου* 'give me', which simulate a live, back-and-forth exchange. Vocatives and short interjections (e.g., *ρε*, *ε*, pause-fillers) often flank these frames, tightening addressivity and projecting audience uptake.

The word *είμαι* appears frequently for self-definition and a claim to authenticity on the part of the artist. The rapper explicitly states "who they are" through the lyrics, assuming the role of cultural bearer or authentic spokesperson of the street; this foregrounds the creator's significance as a central figure in rap who defends their identity and constructs the self as MC or representative artist. Thus, rap is projected as a cultural force, and essentially a pillar of expression and meaning-making for the group or local community. Collocationally, *είμαι* tends to appear in self-positioning schemas (*είμαι* + N/Adj), combining role claims and value stances (e.g., *είμαι MC* '[I] am an MC' / *είμαι αληθινός* '[I] am real' / *είμαι από [γειτονιά]* '[I] am from the hood'). Such formulae fuse craft identity with local anchoring and authenticity work.

The use of *μαλάκα* is associated with direct, oral language drawn from urban everyday life and slang. The lemma under scrutiny is retained un glossed, as its pragmatic value spans camaraderie, surprise, and insult, resolved by co-text and delivery. Although in formal contexts it is considered profanity, in Greek rap it often functions as an expression of familiarity, surprise, or spontaneity and becomes a marker of connection to local identity, reflecting oral culture and the construction of an "authentic" street voice, while at the same time establishing a code of familiarity with an audience that shares the same linguistic usage. In performance terms, *μαλάκα* functions as a rhythmic vocative that both marks in-group intimacy and cues audience response.

The keywords that emerged from comparing the GRC with the song-lyrics corpus (SLC), and that are linked to rap's linguistic identity vis-à-vis the lyric production of other musical genres, are: *rap* (rank 3), *πρόκληση*-challenge (rank 7), *ρίμα*-rhyme (rank 14), and *μικρόφωνο*-microphone (rank 16). These underscore, on the one hand, the close connection between Greek hip-hop and the international current, as for instance the shared terminology *rap* and the notion of *challenge* as a competitive element characterizing the global scene and, on the other hand, a pronounced orientation toward technical mastery. *Rhyme* highlights the lyrical dexterity regarded as a central criterion for the MC, while *microphone* relates to live performance, stage presence, and performing before an audience, thereby capturing the emphasis placed on the art of the word, competitive skill, and the substantive linkage of Greek rap to global hip-hop culture. Examined together, *rap/ρίμα/ μικρόφωνο/πρόκληση* delineate a “craft & performance” bundle, which is characteristic of the 1990s Greek scene, employing *rap* as the named practice and scene affiliation, *ρίμα* as the locus of verbal ingenuity, *μικρόφωνο* as a metonym for live presence and stage authority, and *πρόκληση* indexing agonistic display (battles, cyphers, lyrical one-upmanship) that mirrors the competitive ethos of international hip-hop. These interactional and craft-centred infrastructures, realised through overt address, directive frames, vocatives, and rhyme/microphone discourse, form a stable base that subsequent decades will retain while re-weighting specific resources (e.g., particles in the 2000s; ad-libs and brand/persona lexicon in the 2010s).

4.2. GRC '00: Broadening of themes and lyrical prospects

The comparison of the 2000s subcorpus (~83,154 words) highlighted as keywords *μου*-me/my (rank 1), *σου*-you/your (rank 2), and *μάγκα*-tough guy (rank 11). The frequent use of the pronouns *μου* and *σου* in Greek rap lyrics indicates a strengthening of the interpersonal function and a sharper turn toward personal stories and emotions, as the artist narrates experiences, desires, and thoughts belonging to their own sphere (*η καρδιά μου, η ζωή μου* ‘my heart’, ‘my life’), while with *σου* they address directly, and often intimately, a real or imagined interlocutor (*σου λέω άκου* ‘[I] tell you listen’). These linguistic choices intensify the sense of dialogue, as the creator does not remain within a monologue but introduces the listener, or a specific addressee, into the narrative, thereby enhancing immediacy and emotional involvement. Beyond pronouns, the 2000s material exhibits denser clause-linking and stance management, as particles and complementisers such as *μα, να/μη(ν), αν, πως* proliferate, producing rapid parataxis and turn-initial stance pivots. Crucially, vocatives and the particle *ρε* frequently frame imperatives, forming allocutive envelopes that simulate live exchange. Typical configurations range from clause-initial summons + command (*Ξεκόλλα, ρε χαμένε!* ‘snap out of it, yo lame’), medial vocative insertion (*Μα, άντε βγάλε... μάγκα...* ‘Go on then, show it-what you got, tough guy’), and clause-final *ρε* tagging a directive (*Βγες έξω, ρε* ‘Step outside, huh’). These patterns show that address terms co-engineer directive force rather than merely decorate it.

By contrast, the word *μάγκα* reflects the social expectations and stereotypes of the period concerning endurance, honour, resistance, and one's stance toward adversity. The lemma *μαγκιά* operates as a cultural code that links the artist to a shared sense of urban identity, projecting a model of the tough or independent individual

who defends the values of solidarity, *μπέσα* ('keeping one's word'), or the street, as well as the social stereotypes of dynamism and the code of honor that govern a particular conception of the masculine role. Concordances place *μάγκα* predominantly in vocative position and near second-person imperatives (*έλα, άκου, δεσ, 'come', 'listen', 'see'*), softening or intensifying the directive depending on prosody and co-text, while pairings with *μπέσα* foreground an ethics of solidarity and reciprocity, because in performance terms, *μάγκα* works as an addressive tag that indexes urban belonging and cues audience uptake.

The items that emerged from comparing the 2000s corpus with the general song-lyrics corpus are MC (rank 2), *ρίμες* 'rhymes' (rank 6), and *rap* (rank 7). These indicate that rap in this period focuses particularly on the refinement of the lyrics and on verbal play, namely rhyme, rhythm, and originality. The rapper self-defines not merely as a singer or lyricist but as a craftsperson of language who occupies a distinctive place in hip-hop culture. The notion of *rhyme*, beyond simple end-rhyme, signals the importance of linguistic inventiveness and the creator's poetic skill, through which they seek to impress or even challenge via elaborate lyrical structures. Finally, *rap* acquires the dimension of a technical field for self-affirmation and the display of lyrical abilities; in combination with MC and *rhymes*, it underscores the emphasis on personal skill and on the artist's stage presence. Examined as a whole, *MC/ρίμες/rap* form a metapoetic craft bundle, with *MC* naming role and scene position, with *ρίμες* focalising technique and verbal ingenuity, and with *rap* labelling the practice and its evaluative field. Typical frames (*κάνω/σπάω ρίμες* 'I pop rhymes'; *οι ρίμες μου* 'my rhymes'; *είμαι MC/ο MC της...* 'I am the MC') fold self-presentation into craft evaluation, crystallising a poetics of technical mastery in which stage authority is earned through explicit demonstrations of rhymework. A further recurrent layer involves interjections and resets, e.g., *άντε, γο, ωπ*, that regulate breath, mark section boundaries, or trigger crowd response. They often co-occur with hook entries or verse transitions, functioning as micro-timing devices that keep flow and audience synchronised.

In sum, the 2000s retain the 1990s' interactional substrate (overt address, directive frames) but reweight it toward particle-mediated stance work and explicit craft talk, laying the groundwork for the 2010s' performance vocables and brand/persona lexicon.

4.3. GRC '10: Commodification, trap, and globalisation

The 2010s subcorpus (~68,670 words) is marked by the clear influence of the trap current and the internationalised culture of rap. Words such as *yah* (rank 3), *dab* (rank 5), *gangsta* (rank 14), *gucci* (rank 16), *ferrari* (rank 23) attest to the entry of a new aesthetic in which the projection of a luxury lifestyle and foreign-language expressions assume significant prominence (Nieborg & Poell, 2018; Prey, 2020; Baym, 2018; Morris & Powers, 2015). Moreover, performance vocables and energy cues are frequently manifesting themselves in the corpus in different forms. Firstly, as ad-libs such as *yah/whoa/uhh* that punctuate hooks and verse entries, stretching bars and signalling breath and uptake (e.g., *yah, yah ... , whoa, whoa, whoa*). Secondly, as sound effects indexing and sonifying motion as for instance *vroom, vroom, vroom* around the lemma *Ferrari*, or *skrrrt* for braking, in order for speed and display to arrive on the beat. *Dab* functions less as a lexical verb and more as a danced deictic, as in *κάνω dab*,

dab 'I dab', recruiting the audience into a visual pose that doubles as chorus. The presence of items such as *yah* and *dab* indicates that rappers of the 2010s draw authentic vocabulary and influences from the American and global hip-hop scene, thereby confirming the phenomenon of globalisation. Additional statistically significant lemmas emerging from the keyword lists are those related to brands and persona work. The lemma *Gucci* oscillates between store, taste marker, and alias, as in *το Gucci Store (yah), αν δεν είναι double G...* 'At the Gucci Store (yah), if it ain't double G...', while *Gucci Mane*, a renowned hip-hop artist, is invoked to anchor status and sound *Όλα είναι Gucci Mane* 'Everything is Gucci Mane'. This fusion of brand and persona magnifies credibility through spending, access, and affiliation. At the same time, the terms *gangsta*, *gucci*, and *ferrari* highlight a shift toward aesthetic values that accord greater weight to lifestyle and image, with numerous references to luxury goods and symbols of wealth. Material goods and consumerism become central elements of the narrative, along with money talk functioning as as timing, stance, and power. The lemma *λεφτά* frames biography and gatekeeping, as in *Ποτέ δεν είχαμε λεφτά... τώρα είν' όλοι fans* 'We never had cash... now everybody's a fan', where scarcity authenticates abundance and polices who may speak. *Φράγκα* pushes immediacy and control, as in *Φέρε τα φράγκα αμέσως* 'Bring me the cash right now' treating value as an on-demand transfer and choreographing the scene. Proceeding to the next category, namely scene framing and participation, hooks repeatedly install local chronotopes, such as *club*, *νύχτα* 'night', *γειτονιά* 'neighbourhood', as participation scripts, aligning setting, crowd, and beat. In Greek rap of the 2010s, these terms simultaneously strengthen the connection with the international hip-hop scene and project a new social trend toward ostentatious consumption, which relocates rap within a framework of global, commodified culture. Moreover, lemmas such as *mothafucka* (rank 13) and *gangsta* (rank 14) also underscore rap's tendency to incorporate lexemes and symbols that are directly associated with global hip-hop culture as well as with its commodified facet. The *gangsta* persona functions as a reference to the harder or more extreme manifestations of American rap, while the use of items such as *mothafucka* signals the radical verbal tradition of international rap, introducing elements of immediacy and provocation. All this bears witness to a dynamic linkage between Greek rap and the international scene, in which a consumerist mentality and raw expressions constitute key components of style, steering local production toward a more globalised framework.

In short, the 2010s retain the addressive substrate of earlier decades but re-weight it toward performance vocables, kinetic sound icons, and brand/persona lexicon, aligning Greek rap with the platformed aesthetics of global trap while maintaining local participation frames.

4.4. Rap-specific salience against the Greek SongLyricsCorpus (SLC)

Shifting the baseline from the general written language to the Greek SongLyricsCorpus allows us to isolate what is specifically rap within the broader ecosystem of popular Greek lyrics. The comparison unveils that rap keeps a stable core of craft and performance talk, strengthens a metapoetic and scene-oriented vocabulary, and maintains an interactional envelope that is denser than what is usual in non-rap songs. In other words, when rap is measured against other song texts rather than

against general Greek, its distinctiveness remains visible in the way it names its practice, foregrounds technique, and stages address.

In the 1990s the material brings to the surface a compact craft cluster that includes *ραπ*, *ρίμα*, *μικρόφωνο* and *πρόκληση*. These items present rap as a named practice, rhyme as the locus of verbal technique, the microphone as a metonym of live authority, and challenge as an agonistic frame. The same comparison elevates colloquial vocatives such as *ρε* and the context dependent *μαλάκα*, which together signal a face-threatening register that is rarer in mainstream song and aligns with street-styled orality. Group and role identifiers, for example *Ημισκούμπρια* and *MC*, anchor performers within a recognisable scene. Even plural clitics such as *σας* and *τους* remain more visible than in the general song field, which suggests that rap continues to carry thicker alignment and disalignment work with addressees and third parties.

The 2000s intensify metapoetic self-reference and role labelling. *MC*, *ρίμες* and *rap* remain key against the song-lyrics baseline, which confirms that craft talk distinguishes rap from adjacent lyric genres rather than from the written language alone. The period also brings scene-internal tagging and infrastructure into clearer focus, with forms that point to crews, affiliations and distribution formats. On the interactional side, *μάγκα* and high-frequency interjections such as *άντε* appear as keywords in the SLC comparison. In concordance contexts they flank imperatives and vocatives and help simulate live exchange, which is a usage profile less entrenched in mainstream pop or λαϊκά lyrics. The distinctiveness of the 2000s therefore lies not only in what rap revolves around but also in how participation is recruited and managed within the line.

The 2010s preserve the craft spine but the SLC comparison shows a different layer becoming decisive. Trap-coded performance vocables such as *yah* and *dab*, Anglicised ad-libs, and a conspicuous brand and persona lexicon that includes *gucci*, *ferrari* and *fendi*, as well as celebrity monikers, now appear at scale. Money talk and high denomination currency items become more visible. These resources configure a platform-facing and globalised style that remains relatively rare outside rap, which is why they retain their distinctiveness even when the comparator is other songs rather than general Greek.

Read across decades, the SLC baseline clarifies a pattern of continuity followed by reweighting. The craft and performance bundle that centres on rap, rhyme, the microphone and competitive display is already in place in the 1990s and continues to define the genre's self-presentation within the wider song field. The 2000s expand metapoetic and affiliation talk while thickening the address and directive envelope through vocatives and interjections, which gives the discourse a stronger flavour of simulated liveness than the average song lyric. The 2010s add a performance and commodity references composed of ad-libs, kinetic sound icons, brands and personae. This palette is largely orthogonal to mainstream song conventions, so it remains highly distinctive even when rap is benchmarked against the song-lyrics corpus.

Methodologically, the SLC contrast complements the CGT contrast. The general baseline highlights the distance from written norms and renders the oral texture salient. The song-lyrics baseline highlights markings of craft, role, scene and globalisation that make rap stand out among songs. Using both baselines facilitates the separation of features that rap shares with Greek song from the inventory that is robustly

rap specific, and this separation is the basis for the cross-decade synthesis that follows.

4.5 Synthesis of cross-decade trajectories

Viewed diachronically, Greek rap maintains a consistent pattern of address, namely an I/you layer with vocatives and directives, onto which different stylistic resources are progressively layered. This pattern is visible against the general-language baseline and remains visible when the reference is the song-lyrics baseline. What fluctuates across decades is not the presence of this layer, but its weighting and the specific lexical sets that co-select with it. In other words, the genre continues speaking *to* someone even as the linguistic means of doing so diversify.

At the level of addressivity, the 1990s over-realise first- and second-person forms and tolerate bald imperatives, a striking choice for a pro-drop language where overt subjects and unmitigated directives are pragmatically marked. This produces a lyric texture of direct alignment and challenge. In the 2000s, the same dyadic scaffolding is densified rather than displaced, as clause-linkers and particles (e.g., *μα, να/μη(ν), αν, πως*) proliferate and are recruited turn-initially to pivot stance, while vocatives (e.g., *ρε, μάγκα*) are threaded through imperative frames to fine-tune pressure, solidarity, or face-threat. The 2010s conserve this contact layer but overlay it with performance cues and kinetic tags that manage energy and uptake (*yah, whoa, skrrrt, vroom*), binding Greek material to a recognisably trap-coded delivery. Relative to the song-lyrics baseline, direct address is not exclusive to rap, yet rap sustains a denser and more functionally stacked use of pronominals, vocatives and directives.

The craft/metapoetic stratum exhibits continuity with shifting emphases. In the 1990s, a compact cluster (*rap, ρίμα, μικρόφωνο, πρόκληση*) foregrounds technique, stage authority, and agonistic display as core values. In the 2000s, craft talk becomes more self-indexical, with *MC, ρίμες, rap*, folding role claims and technical evaluation into first-person self-presentation. By the 2010s these craft markers persist but increasingly co-occur with platform-coded performance lexicon (ad-libs, hook signals) and with commodity/persona terms, so that technical prowess is narrated alongside access, affiliation, and conspicuous style. As presented by the song-lyrics baseline, these craft tokens are rap-specific rather than generic features of Greek lyricism.

The display economy, through brands, personae, and money talk, intensifies over time. While the 1990s and 2000s ground credibility in locality, stance and technique, the 2010s elevate brand/persona interlock (*Gucci* as store/label alongside *Gucci Mane* as persona), and monetised imagery (*λεφτά, φράγκα, εκατοστάευρα*), integrating these with performance vocables in hooks and verse entries. The result is not a replacement of Greek frames but an accretion, as global icons of speed, luxury and hype are grafted onto familiar participation chronotopes (e.g., *γειτονιά, νύχτα, club*). These items retain salience when benchmarked against the song-lyrics baseline, indicating that the commodity layer is a rap-specific development in the 2010s rather than a general shift in Greek song.

Finally, registers of aggression and antagonism persist but shift their surface: the 1990s lean on Greek insults and confrontational vocatives; the 2000s retain these while tightening control through particles and vocative placement; the 2010s amplify code-mixed face-threat (e.g., *mothafucka, bitch*) alongside Greek items, aligning the

acoustic grain of provocation with a global hip-hop timbre. This register co-exists with affiliative cues and crew/index terms, sustaining rivalry and belonging in the same textual space. The song-lyrics comparison likewise suggests that code-mixed face-threat resources are denser and more routinised in rap than in mainstream song texts. Thus, the evidence supports a layered account of stylistic change, specifically a contact layer that remains constant; a craft/metapoetic layer that peaks across the 1990s–2000s, and a performance-vocables/commodity layer (ad-libs, sound effects that index motion, brands, money) that accelerates in the 2010s. Using both baselines clarifies what rap shares with Greek song and what remains distinctively rap. This layered re-weighting explains how Greek rap remains diachronically coherent while staying synchronically adaptive, an interpretation that will frame the cross-lingual comparison with U.S. rap that follows.

5. Comparison of Greek and English-language rap

The English-language rap corpus (ERC) is examined across a broader chronological continuum (1960–2010). The two corpora differ in historical depth but share the same construction and analysis pipeline, an issue of significance for interpretation. The English-language corpus captures the emergence and consolidation of rap in the United States and therefore carries a longer social history of racialised and gendered language, while the Greek corpus captures a shorter arc and a later phase of globalisation. Comparisons therefore rest on functions rather than on one-to-one lexical histories, a point considered appropriate for stylistic analysis because address, stance and performance can be compared as discourse work even when emblematic signs differ. To ensure comparability with the Greek analysis, the ERC was assembled decade by decade and processed with the same pipeline of tokenisation, keywording, concordances and collocation profiling. This longer temporal horizon captures the move from proto-forms in the 1960s and 1970s to consolidation and diversification from the 1980s onward. In the earlier decades (1960s–1970s), only a few primary forms of rap are identified, whereas from the 1980s onward canonical keywords emerge, *nigga* (rank 1), *bitch* (rank 2), and *hoe* (rank 3), which carry strong cultural, racial, and gendered charge (Delis, 2019; Alim & Smitherman, 2012). These items do not remain stable in meaning across time. Their distribution and polarity shift by decade, which allows a diachronic reading of stance, address and group boundaries. More specifically, *nigga* is resemanticised as an in-group expression but continually provokes controversy outside the African American community, while terms such as *bitch* and *hoe* predominantly cast the social actors they reference in a negative light, with a limited number of artists reappropriating *bitch* (Delis, 2019). Within this trajectory, *bitch* develops an ambivalent profile that combines persistent misogynistic use with strategic reappropriation by some women rappers for self-assertion and clique building, whereas *hoe* remains more consistently dysphemic. These patterns corroborate corpus findings in Delis (2019) and align with the impoliteness-based account of taboo and naming strategies presented by Delis (2022). An impoliteness perspective facilitates aligning the two scenes at the level of function. Prior work illustrates that taboo items and naming strategies serve both aggression and affiliation depending on who speaks and who is addressed. English-language rap routinises

taboo as a resource for stance and for crowd work, while Greek rap routinises vocatives, particles and imperatives and later adds ad-libs and brands as performance cues, indicating that the pragmatic work is similar, yet the inventories differ. Thus, reading the two corpora through this lens clarifies why convergence appears at the level of address and confrontation while divergence persists in the choice of emblematic signs.

Accordingly, the principal points of convergence between Greek and American rap relate to pronounced self-referentiality centered on the artist's ego, the intensified interpersonal function that creates a dialogue with the listener, and social critique that frequently encompasses political and societal issues. In both traditions direct address is central to performance. The findings, therefore, foreground the classic topos of authenticity in hip-hop (Cutler, 2007). Lyrics are staged as dialogue through overt I/you reference, vocatives and directives, while metapoetic talk evaluates craft and flow. Greek rap realises this contact layer through dense pronominal and vocative work already visible in the 1990s, whereas English-language rap combines pronominal address with routinised taboo as forms of naming and stance. Simultaneously, both scenes continually develop their stylistic mechanisms, such as flow and rhyme techniques. By the 2010s both scenes also expand a performance-management and display register. In the ERC this includes ad-lib driven hype and money talk alongside the continued presence of charged terms. In the GRC it includes ad-libs, kinetic sound icons and brand and persona lexicon layered onto Greek participation frames. Divergences concern, inter alia, linguistic influences stemming from local slang and domestic cultural references; gender representation, as for instance the use of charged terms for women, which has a different weight and historical context in American rap, and, finally, differences in the trajectories and evolution of the genre, which exhibit specific local particularities when compared with the way rap has been historically structured in the United States. In Greece imported English invective appears later and remains more selective. Gender is often staged through local vocatives and insults rather than through wholesale adoption of *bitch* or *hoe*. In the Greek 2010s, brand references and persona building are more prominent. By contrast, the English-language corpus (ERC) reflects the long social history of racialised and gendered terms that shape address and provocation in the earlier decades. These differences reflect distinct media ecologies and social histories as well as each scene's position within global hip-hop. Therefore, the two corpora point to a shared pragmatics of confrontation and solidarity realised through different emblematic signs, a result that the following section mobilises for a cross-lingual synthesis.

6. Conclusions and future directions

On the basis of the quantitative and qualitative findings, Greek rap in each temporal phase (1990s, 2000s, 2010s) exhibits a high degree of orality through direct address to the listener, drawing on slang, pronominal reference, and dialogic formats (e.g., *Φίλε, πες μου τι γίνεται* 'Friend, tell me what's going on'), and cultivates a sustained sense of proximity. This "new orality" (Hoover, 2002) emerges as a key resource shaping both stage presence and lyric architecture, as the artist does not remain within a monologue but invites the listener to participate in a virtual exchange.

Consequently, freestyle and on-stage improvisational practices help consolidate the performer's image as a popular "spokesperson" and intensify the sense of the "here and now," elevating discourse into a social act of live communication. Decisive in this respect is the systematic recurrence of personal pronouns and first-person verbs (*μου, σου, είμαι*), attested across all decades, which underscores an autobiographical tendency, a confessional dimension, and the continual need for artists to position themselves as central subjects. Through such repetition, lyrics acquire rhythmic and aesthetic cohesion, functioning as a stylistic device that provides both musicality and thematic focus to the artist's ongoing, personal narrative. The "I" remains a stable point of reference that generates an internal rhythm, heightens the sense of a continuing personal monologue, and turns the listener into a confidant of the individual narrative and of the creator's evolving persona over time. These patterns support a layered account of style in which a pattern of address remains stable while small but powerful inventories are reweighted over time.

As for the implications for Digital Literary Studies, this study illustrates how corpus stylistics can model a performance-centred poetics without erasing local texture. Treating lyrics as poetic discourse allows quantification of small and frequent items that carry interactional weight. The approach scales from micro-units to cultural interpretation by linking keyness and collocation to contact, craft and display layers. Employing two baselines within the same language demonstrates how to separate genre-internal signals from general properties of song. Cross-lingual comparison then proceeds at the level of function and register rather than by forcing lexical equivalences. Such workflow can be redeployed for other genres with strong performance layers such as battle poetry or slam and for other media where address and timing are central.

A diachronic view makes clear that Greek rap is far from static. It begins with a more activist, socially oriented form, where "authenticity" and a critical stance are primary, and moves toward a more globalised version in which notions of status, consumption, and international lifestyle enter the lexicon. Even so, social and political critique remains salient for many artists, producing a field of tension between a commercial/surface image and deeper thematic concerns. The song-lyrics baseline indicates that the commodity and brand or persona terms remain diagnostic for rap in the 2010s rather than a general shift in Greek song.

Greek rap is, on the one hand, linked to American influences (flow, brands, the global-trap aesthetic) and, on the other, retains local elements such as slang, references to Greek social realities (e.g., political corruption, unemployment), and the emergence of *μαγκιά*-toughness as a conceptual frame. This duality, between the local context and global rap culture, underpins the basic stylistic choices observed in the corpora under study. The presence of "soul" (*psychi*) and *magkiá* in the lyrics signals artists' desire to function as vehicles of social protest and political discourse, mobilising distinctly local expressive means that reflect popular/urban culture and its linguistic particularities. At the same time, the use of foreign-language terms evidences efforts to adapt to global trends: artists seeking international interaction adopt words and symbols from the global scene, thereby fashioning a hybrid style. The incorporation of external elements onto a local substrate does not entail the erasure of specificity, rather, it yields new stylistic, aesthetic, and cultural hybrids. The English data

show that resemanticised racialised and gendered terms structure address and provocation in ways not mirrored in Greek rap (Delis, 2019). Delis (2022) demonstrates how taboo and naming strategies operate as resources for both aggression and affiliation in Anglophone rap, which offers an interpretive lens for Greek patterns where antagonism is realised through local vocatives, particles and performance cues rather than through a wholesale adoption of English invective (Delis, 2024).

Drawing on methods from the digital humanities, and, more specifically, the methodology of corpus linguistics (McEnery & Hardie, 2012; Schreiber, Siemens, & Unsworth, 2016; Dimitroulia et al., 2024), rap can be examined through a comprehensive, “distant reading” perspective that enables more objective inferences about the genre’s language, tendencies, and macro-level shifts. The analysis used two reference baselines, the general Corpus of Greek Texts and a song-lyrics corpus, in order to separate broadly lyric features from the rap-specific inventory that emerged from bottom-up keywording and collocation. In this way, research is not confined to selected samples or personal impressions, rather, it can precisely map rap’s diachronic developments, stylistic invariants, and thematic trajectories in both local and global frames. Future research prospects should include multimodal approaches that correlate lyrics with audiovisual material. On this view, treating rap as poetic discourse is not merely a terminological choice but a methodological commitment that integrates literary description with corpus-based measurement, in line with work that theorises rap’s poetic procedures and performance voice (Carinos & Hammou, 2017; Pecqueux, 2007; Rolland, 2021). The present cartography can be regarded as a first step toward a broader and more in-depth scholarly treatment of the topic.

Future work should align lyrics with audio to model flow and delivery with prosodic and timing features. It should extend collocation analysis to rhyme-scheme and metre, annotate code-mixing more systematically, and sample subgenres and regional scenes to improve representativeness. Cross-lingual replication with comparable pipelines would test how the layered account travels beyond Greek and English, while multimodal analysis that integrates video performance and platform metadata would clarify how stylistic layers are staged, circulated and received.

References

- Alim, H. S., & Smitherman, G. (2012). *Articulate while Black: Barack Obama, language, and race in the U.S.* Oxford University Press.
- Anthony, L. (2019). *AntConc* (Version 3.5.8) [Computer software]. Waseda University. <https://www.laurenceanthony.net/software/antconc/>.
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- Baym, N. K. (2018). *Playing to the crowd: Musicians, audiences, and the intimate work of connection*. New York University Press.
- Biber, D. (2011). Corpus linguistics and the study of literature: Back to the future? *Scientific Study of Literature*, 1(1), 15-23.
- Bonin, M., & Krastev-McKinnon, N. (2019). La poésie ludique du rap. *Esprit*, 460, 84-88. <https://www.jstor.org/stable/26854237>.
- Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139-173.
- Carinos, E., & Hammou, K. (2017). Approches du rap en français comme forme poétique. In S. Hirschi, A. Belhadjin, & M. Murat (Eds.), *La poésie délivrée*. Presses universitaires de Paris Nanterre. <https://doi.org/10.4000/books.pupo.10358>
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22-29.
- Culpeper, J. (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, C. (2007). Keepin' it real: White hip-hop artists' discourses of language, race, and authenticity. *Journal of Linguistic Anthropology*, 17(1), 52-73.
- Delis, P. (2019). *Power, powerlessness and empowerment in African-American and White performers' hip-hop rhetoric: A contrastive corpus-driven critical discourse analysis* [Unpublished master's thesis]. National and Kapodistrian University of Athens.
- Delis, P. (2022). Impoliteness in hip-hop music: African American and White artists' racist and sexist rhetoric. *Journal of Language Aggression and Conflict*, 10(1), 197-218. <https://doi.org/10.1075/jlac.00062.del>
- Delis, P. (2024). *The analysis of textual and stylistic features of Greek hip hop songs through corpus linguistics* [Conference presentation]. 2nd Postgraduate Workshop in Pragmatics, National and Kapodistrian University of Athens.
- Δημητρούλια, Τ., Γούτσος, Δ., & Φραγκάκη, Γ. (2024). *Εισαγωγή στις Ψηφιακές Ανθρωπιστικές Επιστήμες: Ένας πρακτικός οδηγός*. Καρδαμίτσα [Dimitroulia, T. Goutsos, D. & Frangkaki G.. *Introduction to Digital Humanities. A Practical Guide*].
- Evert, S. (2005). *The statistics of word co-occurrences: Word association measures*. [Doctoral dissertation, University of Stuttgart].
- Fischer-Starcke, B. (2009). Keywords and frequent phrases of *Pride and Prejudice*: A corpus-stylistic analysis. *International Journal of Corpus Linguistics*, 14(4), 492-523.
- Goutsos, D. (2010). The Corpus of Greek Texts: A reference corpus for Modern Greek. *Corpora*, 5(1), 29-44.

- Γούτσος, Δ., & Φραγκάκη, Γ. (2015). *Εισαγωγή στη γλωσσολογία σωμάτων κειμένων*. ΣΕΑΒ [Goutsos, D & Fragkaki, G., *Introduction to Corpus Linguistics*]. <http://hdl.handle.net/11419/1932>.
- Holton, D., Mackridge, P., & Philippaki-Warbuton, I. (2012). *Greek: A comprehensive grammar* (2nd ed.). Routledge.
- Hoover, D. L. (2002). Frequent word sequences and statistical stylistics. *Literary and Linguistic Computing*, 17(2), 157-180.
- Krims, A. (2000). *Rap music and the poetics of identity*. Cambridge University Press.
- Labourie, M. (2017). *Le rap comme poétique du langage ordinaire* [Master's thesis]. DUMAS. <https://dumas.ccsd.cnrs.fr/dumas-01706805>.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- McIntyre, D., & Walker, B. (2019). *Corpus stylistics: Theory and practice*. Edinburgh University Press.
- Moretti, F. (2013). *Distant reading*. Verso.
- Morgan, M. (2009). *The real hip hop: Battling for knowledge, power, and respect in the LA underground*. Duke University Press.
- Morris, J. W., & Powers, D. (2015). Control, curation and musical experience in streaming music services. *Creative Industries Journal*, 8(2), 106-122.
- Nieborg, D. B., & Poell, T. (2018). The platformization of cultural production. *New Media & Society*, 20(11), 4275-4292.
- Pecqueux, A. (2007). *Voix du rap: Essai de sociologie de l'action musicale*. Presses de la Maison des sciences de l'homme.
- Pennycook, A. (2007). *Global Englishes and transcultural flows*. Routledge.
- Podhorná-Polická, A. (2020). RapCor, Francophone rap songs text corpus. In *Proceedings of RASLAN 2020* (pp. 45-52). <https://nlp.fi.muni.cz/raslan/2020/paper6.pdf>.
- Prey, R. (2020). Locating power in platformization: Music streaming and playlist curation. *Social Media + Society*, 6(4), 1-11.
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the Workshop on Comparing Corpora* (pp. 1-6). Association for Computational Linguistics.
- Rolland, N. (2021). « Dans la rue et le bitume ne poussent que les fleurs du mal »: Jouer et déjouer Charles Baudelaire et sa poésie dans le rap français. *Itinéraires*, 2020(3). <https://doi.org/10.4000/itineraires.9377>
- Rose, T. (1994). *Black noise: Rap music and Black culture in contemporary America*. Wesleyan University Press.
- Schreibman, S., Siemens, R., & Unsworth, J. (Eds.). (2016). *A new companion to digital humanities*. Wiley-Blackwell.
- Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. John Benjamins.
- Škrabal, M., & Piorecký, K. (2022). The Corpus of Contemporary Czech Poetry: A database for research on contemporary poetic language across media. *Digital Scholarship in the Humanities*, 37(2), Article qhab072.
- Terkourafi, M. (Ed.). (2010). *Languages of global hip hop*. Continuum.

Werner, V. (2019). Assessing hip-hop discourse: Linguistic realness and styling. *Text & Talk*, 39(6), 749-773. <https://doi.org/10.1515/text-2019-2044>.

Περίληψη

Παναγιώτης Δελής

Το λεξικό της ελληνικής ραπ: Μια υφολογική μελέτη σώματος κειμένων

Η ραπ, ως οργανικό τμήμα της κουλτούρας του χιπ-χοπ, αποτελεί ένα είδος προφορικής τέχνης, στο οποίο ο ρυθμικός λόγος συναντά τη μουσική δομή και την κοινωνική απεύθυνση. Το άρθρο εφαρμόζει μεθόδους υφολογίας σωμάτων κειμένων εξετάζοντας το λεξιλόγιο της ελληνικής ραπ ανά δεκαετίες, (1990, 2000, 2010) και αντιπαραβάλλει τα αποτελέσματα με δύο ελληνικά σώματα αναφοράς, το γενικό Σώμα Ελληνικών Κειμένων και ένα Σώμα Ελληνικών Στίχων, καθώς και με ένα σώμα αγγλόφωνης ραπ διαστρωματωμένο ανά δεκαετία. Αναλυτικά, οι στίχοι της ραπ αντιμετωπίζονται εδώ ως ποιητικός λόγος, με την ανάλυση να συνδέει μετρήσιμα μοτίβα με ερμηνεία. Δίνεται έμφαση στις συχνότερα εμφανιζόμενες λέξεις-κλειδιά και κατόπιν ερμηνεύονται με έμφαση στη διαπροσωπική στάση και στα επιφωνήματα. Τα ευρήματα δείχνουν ένα συνεπές μοτίβο απεύθυνσης, που χαρακτηρίζεται από ρητή αναφορά στο πρώτο και δεύτερο πρόσωπο, από κλητικές προσφωνήσεις και προτροπικά σχήματα, πάνω στο οποίο επισωρεύονται μεταβαλλόμενοι υφολογικοί πόροι. Τα ευρήματα αναδεικνύουν στη δεκαετία του 1990 την τεχνική και τη σκηνική αυθεντία, στη δεκαετία του 2000 στάσεις που διαμεσολαβούνται από μόρια και μεταποιητικό λόγο, ενώ στη δεκαετία του 2010 την πληθώρα από ad-libs, και το λεξιλόγιο εμπορικών σημάτων και λημμάτων που σχετίζονται με το σχηματισμό της περσόνας του καλλιτέχνη. Η σύγκριση με την αγγλόφωνη ραπ εντοπίζει συγκλίσεις στην απεύθυνση και στην τεχνική καθώς και αποκλίσεις σε φυλετικά και έμφυλα λεξιλογικά εργαλεία. Αναδεικνύεται δε, ο ρόλος της ραπ ως μέσου αυτοέκφρασης και κοινωνικής κριτικής, αλλά και ως προϊόντος εντεινόμενης παγκοσμιοποίησης και εμπορευματοποίησης. Η μελέτη δείχνει πώς η υφολογία σωμάτων κειμένων στηρίζει τις Ψηφιακές Λογοτεχνικές Σπουδές, καθώς κλιμακώνεται από μικρά λεξιλογικά σύνολα σε πολιτισμική ερμηνεία και διαχωρίζει τα ιδιάζοντα χαρακτηριστικά της ραπ από χαρακτηριστικά που είναι κοινά με το ελληνικό τραγούδι. Υπάρχει εμφανές κενό στην έρευνα για την ελληνόφωνη ραπ με μεθόδους σωμάτων κειμένων. Η μελέτη καλύπτει αυτό το κενό προσφέροντας διαστρωμάτωση ανά δεκαετία και διπλό σημείο αναφοράς, ώστε να διακριθούν τα χαρακτηριστικά που είναι ειδικά της ραπ από όσα μοιράζεται με το ελληνικό τραγούδι.

Λέξεις Κλειδιά

Ψηφιακές Λογοτεχνικές Σπουδές, Υφολογία Σωμάτων Κειμένων, Ελληνική Ραπ, Χιπ-Χοπ, Σημαντικότητα, Προφορικότητα, Παγκοσμιοποίηση, Υφολογία.

Abstract

Panagiotis Delis

The Lexicon of Greek Rap: A Corpus-Stylistic Study

Rap, as an integral component of hip-hop culture, is an oral performance genre in which rhythmic speech aligns with musical patterning and social address. This article applies corpus-stylistic methods to the lexicon of Greek rap across three decades, namely the 1990s, 2000s and 2010s, and sets the results against two Greek baselines, the general Corpus of Greek Texts and a Greek Song-Lyrics Corpus, as well as a decade-stratified English-language rap corpus. What is computed is keyness, concordances and collocations, then the interpretation of the signals with attention to interpersonal stance and performance vocables is what follows. Findings show a consistent pattern of address marked by overt I/you reference, vocatives and directives, onto which changing stylistic resources are layered. The 1990s highlight craft and stage authority, the 2000s foreground particle-mediated stance and metapoetic talk, and the 2010s accelerate ad-libs, and brand or persona lexicon. Comparison with English rap locates convergences in address and craft and divergences in racialised and gendered inventories. Findings underscore rap's role as a vehicle of self-expression and social critique, as well as a product of intensifying globalization and commodification. The study demonstrates how corpus stylistics supports Digital Literary Studies by scaling from recurrent lexical sets to cultural interpretation and by separating rap-specific signals from features shared with Greek song. There is a clear gap in corpus-based research on Greek rap. Existing work on Greek hip-hop is scattered and largely qualitative, with no decade-stratified corpora and no dual-baseline comparisons that would separate rap-specific features from those shared with Greek song. As a result, claims about address patterns, slang, brand lexicon and performance cues often remain impressionistic. A corpus-stylistic programme directly addresses this by quantifying keyness, concordance behaviour and collocation structure and by linking these distributions to stylistic interpretation (Biber, 2011; Culpeper, 2009; McIntyre & Walker, 2019). The diachronic lens is necessary because external drivers of lexical change intensified across the 2000s and 2010s, including the localisation of global hip-hop styles and the platformisation of music production and circulation that re-configured artist-audience address and branding practices (Krimms, 2000; Pennycook, 2007; Terkourafi, 2010; Baym, 2018; Nieborg & Poell, 2018; Morris & Powers, 2015; Prey, 2020). Within hip-hop discourse specifically, operational work on authenticity and styling shows that core resources are linguistically measurable, which further motivates corpus methods for rap (Werner, 2019). Feasibility is ensured by the availability of robust Greek baselines, notably the Corpus of Greek Texts and a compiled Greek song-lyrics corpus, which allow rap-specific signals to be distinguished from properties of general Greek or of song lyrics more broadly (Goutsos, 2010). Prior Greek steps remain preliminary and do not provide a decade-by-decade map of lexical and interactional resources (Delis, 2024). In analytical terms rap lyrics are treated as poetic discourse, following francophone work that has explicitly theorised rap as a contemporary poetic form and explored its intertextual, prosodic and rhetorical

specificities (Carinos & Hammou, 2017; Bonin & Krastev-McKinnon, 2019; Labourie, 2017; Pecqueux, 2007; Rolland, 2021).

Keywords

Digital Literary Studies; corpus stylistics; Greek rap; hip-hop; keyness; orality; globalisation; stylistics.