



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Rebetiko Meets Digital Humanities: Automatic Analysis of
Rebetiko Lyrics via Data Science Methods“

verfasst von / submitted by

Varvara Arzt

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 869

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Byzantinistik und Neogräzistik

Betreut von / Supervisor:

Mag. Dr. Nina-Maria Wanek

Mitbetreut von / Co-Supervisor:

Dr. Dimitrios Kousouris, BA MA PhD

*In memory of the dearest friend and peace campaigner
Roger Franklin*

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisors Priv. Doz. Mag. Dr. Nina-Maria Wanek and Univ. Ass. Dr. Dimitrios Kousouris, for their support. Moreover, I would not have discovered this fascinating research topic without attending the course on Byzantine church music held by Priv. Doz. Mag. Dr. Nina-Maria Wanek as well as the course on rebetiko held by Univ. Ass. Dr. Dimitrios Kousouris.

I would also like to thank Mag. Dr. Claudia Resch for her great course on corpora and digital text collections, in which I have learned a lot about corpora in general as well as available text analysis tools. I was able to apply this knowledge during the creation of the *Rebetiko Corpus* as well as its analysis.

Moreover, I would also like to thank Univ.-Prof. Dr.-Ing. Benjamin Roth and Dr. Renato Rocha Souza for their course on machine learning and natural language processing, in which I picked up many skills that I was able to then transfer into my master's thesis.

I would like to express my gratitude to the Austrian Centre for Digital Humanities and Cultural Heritage, where I was able to do an internship in summer 2021. During this internship, I could obtain a broader understanding of how large scientific projects in the area of digital humanities are conducted, and I have learned a lot about machine learning.

Rebetiko is a beautiful music genre, for which wonderful online resources exist. I would like to thank the creators of and contributors to *rebetiko.sealabs.net*. Without the data available on this website I would not be able to create the *Rebetiko Corpus* in the form in which it exists today. The love and enthusiasm for rebetiko in the *rebetiko forum* (ρεμπέτικο φόρουμ) is deeply impressive, and I have learned a lot by browsing the discussions. I would also like to thank Nikos Politis, a rebetologist and a very devoted contributor to the *rebetiko forum*, for answering my questions. I would also like to thank the creators and the contributors of the *Virtual Kounadis Archive* for their dedication and their commitment to open-data. All these resources and communities have been a huge inspiration for me during my work on my master's thesis. In times of the COVID-19 pandemic, which prevented me from visiting the rebetiko venues in Greece, simply listening to the rebetiko songs and following the discussions in the *rebetiko forum* transferred me at least to an imaginary rebetiko venue.

Moreover, I would like to thank Prof. Stathis Gauntlett for his valuable recommendations regarding the corpus creation and data analysis.

Also, I would like to thank the scholars from the State University of Saint-Petersburg, who raised and supported my interest in linguistics as well as in the Greek and Albanian languages, especially my then supervisor Prof. Alexander Rusakov as well as Dr. Aleksej Andronov, Dr. Alexander Novik and Prof. Fatima Eloeva.

I would like to thank my dear family and friends. A special person, who has influenced my life immensely, is my grandmother Rita, an Assembler programmer and a passionate literature and theatre lover. Every conversation with her is a great joy, whether we discuss mushrooms or machine learning. I am very grateful to my dear husband Andi, who always supports me and has encouraged me immensely during the work on this thesis. Thank you for the discussions on machine learning and data science, your valuable recommendations, and the readiness to spend long evenings proofreading this thesis.

I would also like to thank the crew of the frigate *Shtandart* for being a wonderful part of my life.

Finally, I would like to dedicate this work to the late Roger Franklin, a dearest friend, a cosmopolitan and a peace campaigner, who made the world a better place to live in. Every time I think back to the time that I spent in Tickmorend, I feel Roger's warmth and kindness.

CONTENTS

1	INTRODUCTION	1
2	WHAT IS REBETIKO	3
2.1	The Origins of the Term Rebetiko	6
2.2	Chronological Boundaries of Rebetiko	17
2.3	Periodisation of Rebetiko	26
2.4	The Question of Marginality of Rebetiko	38
2.5	From the Oral Tradition to the Commercialisation of Rebetiko	41
2.6	Examples of Rebetiko Songs	45
3	REBETIKO CORPUS	51
3.1	Existing Rebetiko Song Lyrics and Metadata Sources	51
3.2	Creation of the <i>Rebetiko Corpus</i>	53
3.2.1	Automatic Data Extraction	53
3.2.2	Challenges of the Corpus Creation	55
3.2.3	Key Data of the <i>Rebetiko Corpus</i>	60
3.2.4	Corpus Publication on Zenodo	63
4	DATA ANALYSIS	65
4.1	Data Analysis with Voyant Tools	67
4.1.1	Experiments with Voyant Tools: Word Clouds	68
4.1.2	Experiments with Voyant Tools: Trends	88
4.2	Data Analysis with Custom Tools Written in Python	94
4.2.1	Challenges for NLP Tasks with the <i>Rebetiko Corpus</i>	94
4.2.2	Custom Frequency Graph Over Time	99
4.2.3	Topic Modeling	101
5	CONCLUSION AND FUTURE WORK	109
6	APPENDIX	111
6.1	Composers Included in the <i>Rebetiko Corpus</i>	111
6.2	Use of Pronouns in the <i>Rebetiko Corpus</i>	112
6.3	Example of Lemmatisation Performed by UDPipe with Punctuation	114
6.4	LDA Topic Modeling: Topics 3 and 5	115
7	ABSTRACTS	117
7.1	Abstract	117
7.2	Zusammenfassung	118
	BIBLIOGRAPHY	119

LIST OF FIGURES

Figure 2.1	Gramophone Record of the Song “Tempstress from Smyrna”	15
Figure 2.2	First Known Gramophone Record with the Genre Name <i>Rebetiko</i> Men- tioned on a Label.	20
Figure 3.1	Excerpt of the List of Works by Vamvakaris	54
Figure 3.2	Source Code in Wikitext	54
Figure 3.3	<i>Rebetiko Corpus</i> in Navicat	58
Figure 3.4	Distribution of Songs in the <i>Rebetiko Corpus</i> over Time	62
Figure 4.1	Word Cloud for the <i>Rebetiko Corpus</i> without Normalisation	70
Figure 4.2	Word Cloud for the <i>Rebetiko Corpus</i> after Removal of Stop Words . .	74
Figure 4.3	Word Cloud for the <i>Rebetiko Corpus</i> after Removal of Stop Words and Lemmatisation	75
Figure 4.4	Word Cloud for Lemmatised Nouns and Proper Nouns of the <i>Re- betiko Corpus</i>	77
Figure 4.5	Word Cloud for the Dragatsis Subcorpus after Removal of Stop Words and Lemmatisation	81
Figure 4.6	Word Cloud for the Vamvakaris Subcorpus after Removal of Stop Words and Lemmatisation	83
Figure 4.7	Word Cloud for the Georgakopoulou Subcorpus after Removal of Stop Words and Lemmatisation	85
Figure 4.8	Word Cloud for the Tsitsanis Subcorpus after Removal of Stop Words and Lemmatisation	86
Figure 4.9	Graph Showing the Use of the Word <i>hashish</i> and its Synonyms . . .	89
Figure 4.10	Graph Showing the Use of the Word <i>mangas</i> and its Synonyms . . .	91
Figure 4.11	Graph Showing the Use of Masculine and Feminine Nouns denoting a ‘Careless Person of the Street’	93
Figure 4.12	Example of a Lemmatisation Performed by UDPipe	95
Figure 4.13	Custom Graph: Use of the Word <i>hashish</i> in the <i>Rebetiko Corpus</i> . . .	99
Figure 4.14	Custom Graph: Use of the Word <i>hookah</i> in the <i>Rebetiko Corpus</i> . . .	100
Figure 4.15	Custom Graph: Use of two Verbs Standing for <i>smoke</i> in the <i>Rebetiko Corpus</i>	100
Figure 4.16	LDA Topic Modeling: General View	104
Figure 4.17	LDA Topic Modeling: Topic 1 <i>Love</i>	105
Figure 4.18	LDA Topic Modeling: Topic 2 <i>Hashish Consumption</i>	106
Figure 4.19	LDA Topic Modeling: Topic 4 <i>Family, Homeland and Foreign Land</i> . .	107
Figure 6.1	Example of Lemmatisation Performed by UDPipe with Punctuation	114
Figure 6.2	LDA Topic Modeling: Topic 3	115
Figure 6.3	LDA Topic Modeling: Topic 5	115

LIST OF TABLES

Table 6.1	List of Composers with more than 20 Pieces in the <i>Rebetiko Corpus</i>	111
Table 6.2	Use of First-person Pronouns in the <i>Rebetiko Corpus</i>	113
Table 6.3	Use of Second-person Pronouns in the <i>Rebetiko Corpus</i>	113
Table 6.4	Use of Third-person Singular Pronouns in the <i>Rebetiko Corpus</i>	113
Table 6.5	Use of Third-person Plural Pronouns in the <i>Rebetiko Corpus</i>	113

LISTINGS

Listing 3.1	General Parser for Pages Containing Song Lyrics	56
Listing 3.2	Specific Parser for the Entry for Vamvakaris	57
Listing 3.3	A Song by Vamvakaris in JSON Format (Before Data Correction)	57

INTRODUCTION

Rebetiko is a Greek popular music genre, with its roots going back at least to the second half of the 19th century. It had its heyday between the 1920s and the 1950s, and continues to prosper to this day. Although it has been considered the music of the underworld, rebetiko had and still has an immense influence on the Greek culture and language. Numerous contemporary musicians have reinterpreted both well-known as well as rare rebetiko songs, including Glykeria, Haris Alexiou, Giorgos Dalaras, Marina Satti and Anna Vissi.¹ In addition, there exists a great number of bands that are ascribed to the genre of *neo-rebetiko*, including the band *Trio Tekke*. Rebetiko also attracts the attention from musicians outside of Greece and the Greek diasporas: for instance, in 2020 the British musician Brendan Perry released an album “Songs of Disenchantment: Music from the Greek Underground” on which he reinterpreted old well-known rebetika. Due to the similarities in the genres’ formations, rebetiko is commonly seen as *Greek blues*, but while it often is associated with the ‘marginal’ social groups, it is notable that representatives of a wide social strata have contributed to the music genre (Gauntlett, forthcoming).

While rebetiko is a fascinating combination of musical and textual components, as well as dance, this thesis focuses on the song lyrics and their analysis. Traditionally, rebetiko lyrics have mostly been analysed using a qualitative approach (e.g. (Gauntlett 1985a; Tragaki 2009)). Nowadays, due to advances in technology and interdisciplinary research in the field of digital humanities, the quantitative analysis of cultural artefacts is growing in popularity.² To be able to perform any kind of quantitative research on rebetiko, a sufficiently large corpus of lyrics of rebetika is needed. Interestingly, the importance of building such a corpus of rebetika was already emphasised by a group of researchers in the 1990s, who had planned to create the so-called *Corpus Rebeticorum*, but unfortunately this project was never realised (Gauntlett/Payvanas/Chatzinikolaou 1994).

Thus, the main idea of this thesis is to create and publish a corpus of rebetiko, and to use it for quantitative analysis of rebetiko lyrics.³ More precisely, the thesis tackles the question of corpus creation, both regarding its requirements for the subsequent analyses and from a practical stand point, i.e. how to collect and prepare a sufficiently large amount of data in a reasonable amount of time. Moreover, applicable tools and approaches for the analysis of rebetiko from the field of data science are identified and discussed, including their limitations. Finally, and most importantly, these approaches are applied to the corpus, to try to answer the question if it is a viable way to try to confirm or reject well-known hypotheses

1 For example, Anna Vissi’s reinterpretation of the song “Five Years in Prison (The Sound of a Hookah Pipe)”, released in 2008, and originally recorded in 1936 by Vangelis Papazoglou.

2 An example of such research is the analysis of lyrics of Bob Dylan songs (Sipl/Fuchs/Burghardt 2018), in which a computer-assisted analysis of a dataset with Dylan’s lyrics was performed.

3 The idea of creating a corpus of rebetiko song lyrics was sparked by my lifelong interest in languages and mathematics, as well as by a course on rebetiko that I attended during the summer term 2019, which was titled “Rebetika: Zwischen urbanem Untergrund und kulturellem Mainstream” (Lecture VO 090111) and held by Univ. Ass. Dr. Dimitrios Kousouris at the Department of Byzantine and Modern Greek Studies of the University of Vienna.

about rebetiko via quantitative analyses, and if maybe even new aspects of rebetiko can be uncovered.

This thesis is structured as follows. Chapter 2 is devoted to the music genre of rebetiko. Knowledge of the history and the main characteristics of rebetiko are indispensable for the creation of a suitable corpus and its analysis, thus this chapter discusses the object of research in detail. It includes an outline of the history of the term rebetiko itself and the chronological boundaries and periodisation of rebetiko. Moreover, this chapter discusses the issue of the supposed ‘marginality’ of rebetiko and examines the influence of the oral tradition of early rebetika on later periods of rebetiko, and relates it to the question of authorship. Finally, this chapter also includes a section with examples of rebetiko songs, including a brief analysis of their lyrics, with the goal to prepare the reader for the remainder of the thesis. Chapter 3 discusses existing rebetiko song lyrics and metadata sources, describes how the *Rebetiko Corpus* was created, and gives an overview of the data included in the corpus. Chapter 4 is devoted to the analysis of the *Rebetiko Corpus*. I try to explore the world of rebetiko lyrics with the well-known text analysis tool *Voyant Tools*, as well as with custom scripts implemented in the programming language *Python*, and demonstrate the viability of this approach as a research tool. Finally, Chapter 5 contains the conclusions of this thesis, summarises its main results, and gives an outlook on possible future research in the intersection of rebetiko and data science.

This thesis contains numerous translations from Greek into English. All translations, unless otherwise stated, are mine. Throughout the thesis, I am using the term *Rebetiko Corpus* to refer to the corpus of rebetiko songs, including lyrics and metadata, that was created in the scope of this master’s thesis.

WHAT IS REBETIKO

Before diving into the fascinating world of rebetiko¹, it is important to clarify the term itself. One can define rebetiko as a Greek popular music genre that originated in the second half of the nineteenth century and whose end of development is either dated to the mid-1950s, or it is seen as a tradition that continues its existence also nowadays (Zelepos 2001, 5). Rebetiko is often defined as the music of the urban underworld. Especially at the beginning of its existence, this music genre had a bad reputation due to the fact that it was associated with marginal social groups. It is important to note that rebetiko is a combination of music, verse and dance. Due to the similarities in their formation, rebetiko is often defined as *Greek blues*, but at the same time one could find numerous similarities also with other music genres that have their origins in marginal social groups: for instance, flamenco or fado (Tragaki 2009, 116). Many representatives of rebetiko were refugees from the Asia Minor who arrived in Greece in 1923, but at the same time it would be wrong to connect the origins of rebetiko only with the Asia Minor, as the rebetiko tradition in mainland Greece dates back to the nineteenth century (Zelepos 2001, 23-24).

The music that is called rebetiko today had originally no established name, so over time a range of synonymous terms was used for this genre, such as *mourmourika*, *tsachpinika* or *vlamika*. The record companies mostly avoided to label their records as rebetiko, using such generic English terms like a ‘folk song’ or a ‘popular song’ instead. Even more often, recordings of rebetika were simply labeled with the rhythm in which a particular song was performed, e.g. *zeibekiko* or *tsifteteli*. In contrast to tango or flamenco, for example, the word *rebetiko* is not connected directly to a specific dance or rhythm, but to the figure of rebetis², who can be seen, on the one hand, as a performer of rebetiko and, on the other hand, as a representative of a particular social group.

Until the beginning of 1930s, rebetiko was mainly performed in hashish dens (*tekedes*), gambling dives, prisons and cafés amans, or cafés santours.³ It then gradually moved into

- 1 For representing Greek text with Latin script, a combination of the ALA-LC and ELOT 743 romanisation tables was used, except for cases for which a way of representation has already been established (like in the case of the words rebetiko or mangas). More details can be found at <https://www.loc.gov/catdir/cpsd/romanization/greek.pdf> or <https://web.archive.org/web/20150922010901/http://sete.gr/files/Media/Egkykliai/040707Latin-Greek.pdf> (accessed November 1, 2021). In general, it is quite difficult to ensure consistent transliterations due to the fact that there exists more than one transliteration style. As none of them is a perfect match for the Modern Greek spelling, numerous authors present their own way of transliteration. More information about the topic of transliteration of Modern Greek text can be found in (Bucuvalas 2018).
- 2 As rebetis, pl. rebetes (Greek sg. *ρεμπέτης*, pl. *ρεμπέτες*), a representative of the rebetiko music genre is defined. The meaning of this term is two-fold. On the one hand, it defines a performer and creator of such music. On the other hand, it can also define a person who follows a lifestyle similar to one of rebetis, a kind of vagabond who does not accept the existing moral standards. When speaking about this second definition, it is important to take into account that the figure of rebetis was stigmatised to a certain extent.
- 3 Cafés amans define coffee shops in which a certain kind of music concerts was given. The name for such cafés was given in analogy to French cafés chantants of the period of the Belle Époque that offered a westernized music repertoire and appeared in the Ottoman Empire and Greece in the mid-19th century (Anagnostou 2018, 288). Around the same time, cafés amans, or cafés santours, also appeared in the cities of the Ottoman Empire and originally offered an oriental music repertoire. Later on their repertoire was enriched with Greek music

taverns and urban entertainment centres (*kosmika kentra*). The change in the performance context also shows the general evolution of the music genre: from being associated with the marginal social groups to a music genre understandable to and popular with a much broader social strata.

Many rebetiko songs refer to themes such as unrequited love, hashish consumption, illnesses, sorrow, prison life, gambling, woman's beauty and mother's love. The protagonist of a rebetiko song sees the only way to escape the harsh realities of life in alcohol, hashish or death. Emotions like *πόνος* (English 'pain'), *καημός* (English 'sorrow'), *ντέρτι* (English 'yearning'), *κέφι* (English 'spirit of joy, overwhelming emotion filled with passion'), *μεράκι* (English 'passion, strong desire'), *αγάπη* (English 'love'), *σεβντά* (English 'love') and *βάσανο* (English 'suffering') are very frequent in rebetiko song lyrics.

Regarding the musical aspects of rebetiko, until 1936 many rebetiko songs followed a traditional system of modes called *dromoi* (Greek *δρόμοι* 'roads, paths'), which are connected to the Byzantine ecclesiastic chant as well as to the Turkish makam mode system.⁴ After 1936, when the censorship of Metaxas was imposed and the process of the westernisation of the music began, the traditional system of *dromoi* modes was gradually replaced by the Western well-tempered system (Tragaki 2009, 62).

Most rebetologists distinguish between two types of rebetiko: the Smyrna and the Piraeus style, named after geographical places which were considered to be the main centres of these styles. One of the main differences is the instrumentation which was used in these styles. Rebetika of the Smyrna school, which prevailed in the rebetiko recordings until the beginning of the 1930s, were often sung by female singers and accompanied by violin, santour, oud, qanun and cello. Songs of the Piraeus style, which prevailed from the 1930s onwards, were in contrast usually sung by male singers and accompanied by bouzouki and baglamas.

It seems also important to mention that whereas there were many female singers of rebetiko, the roles of the instrument player and the composer were mainly reserved for male exponents of rebetiko. The appearance of female composers and lyricists, or at least an acknowledged appearance on the labels of the records, was very rare, especially until the second half of the 1940s. As some of the few exceptions one can mention the compositions of Roza Eskenazi⁵, Soula Kalfopoulou, Ioanna Georgakopoulou and Eleni Moraitou. Interestingly, due to the fact that starting in the 1930s record companies were interested in producing larger numbers of records, some songs in which the lyrics were meant to be sung by a male singer, were then also sung by female singers: for example, the song «Με ζουρνάδες με νταούλια» (English "With Zurnas and Davuls"), composed by Panagiotis Toundas, was sung by Stellakis Perpiniadis (disc number: Columbia DG-498, recorded in 1934), but also by Rita Abatzi (disc number: HMV AO-2136, recorded in 1934) and Roza Eskenazi (disc number: Odeon GA-1806, recorded in 1935). The protagonist of this song wins money in the baccarat

genres such as rebetiko, demotic song, light popular song and amanedes (Tragaki 2009, 53). The word *aman* reflects the exclamations that were frequent in songs performed in these cafés. The similar term *café santour* is often used as a synonym of *café aman* and refers to a music instrument that was present in such cafés, namely the santour (Anagnostou 2018, 6, 18). The first permanent café aman in Athens was established at least as early as 1874 (see Morris, Roderick Conway. Greek Café Music. In (Bucuvalas 2018)

4 Dromoi used in rebetika were often connected not only to the echoi (Greek sg. *ήχος*, pl. *ήχοι*) of the Byzantine church music, but also to the Ancient Greek modes (Liavas 2009, 24). As Zelepos points out, this argument was also used to emphasise the idea of continuity of Greece in the context of rebetiko (Zelepos 2001, 133).

5 Roza Eskenazi, who is primarily known as a rebetiko singer, was not only a songwriter of some of the songs, but also an instrument player accompanying rebetika with zills and tambourine.

card game and wants to make *glendi* (English ‘party’) in a tavern with beautiful girls from different parts of Greece, enjoying the *zurna* and *davul* music, and good wine. There are many more examples of songs that were supposed to be sung by male exponents⁶, but were sung by female ones. One also finds some reverse examples. For instance, the song «Τράβα ρε αλάνη» (English “Piss off, You Rover (Alanis)”) composed by Kostas Skarvelis was sung by Rita Abatzi (disc number: His Master’s Voice AO-2291, recorded in 1936) and Roza Eskenazi (disc numbers: HMV AO-2147 and Sahibinin Sesi AO-2147, recorded in 1934), but also by Stellakis Perpiniadis (disc number: Columbia DG-2035, recorded in 1934) and Kostas Roukounas (disc number: Odeon GA-1793, recorded in 1934). The female protagonist of the song “Piss off, You Rover (Alanis)” finds out that her boyfriend has another girl and advises him not to show up anymore in her district, and that from now on she will have fun with a butcher. This song seems to have been a big success among the audience, as four different gramophone records were produced within two years in Athens.

After laying the groundwork in this brief introduction, the following sections will present a deeper look into the history and characteristics of this music genre. Section 2.1 is concerned with the history and the etymology of the term *rebetiko*. It also presents a deeper analysis of such figures of *rebetiko* as *rebetis* and *mangas*. Numerous examples of particular *rebetiko* recordings will help the reader to get a better understanding of the history of *rebetiko*. The Section 2.2 examines the chronological boundaries of *rebetiko*. It observes different existing hypotheses concerning the lower boundary of the music genre *rebetiko*, namely based on the formation of the term *rebetiko* or the appearance of songs similar to those that are considered today as ‘typical’ *rebetika*. Different views regarding the upper boundary are also presented, ranging from defining an endpoint of *rebetiko* in the mid-1950s, to the opinion that *rebetiko* is a music tradition that is still alive in our days. Section 2.3 presents several possible *rebetiko* periodisation patterns and observes their main characteristics. Moreover, it demonstrates main styles of *rebetiko*, such as the Smyrna and the Piraeus style. It also gives an alternative view on *rebetiko*, putting *rebetika* performed in Thessaloniki and America into additional separate categories. The Section 2.4 investigates, whether *rebetiko*, which has long been seen exclusively as the music of the Greek underworld, is really that marginal. It raises the question of drug consumption in Greece and Europe in general at the end of the nineteenth and the beginning twentieth century. It also discusses the roles of ‘jobbing versifiers’ as well as professional music ensembles in the context of *rebetiko*, who only had a very loose connection to the marginal social groups. Section 2.5 examines the period of transition from the oral tradition to the commercialisation of *rebetiko*. It studies characteristic elements of *rebetiko* such as spoken interjections as well as the numerous re-recordings or reinterpretations of the same song, so that the question of authorship often remains unclear. The Section 2.6 presents three *rebetiko* songs from different periods of the music genre and provides a short analysis of their lyrics. It is supposed to prepare the reader for Chapters 3 and 4, in which the corpus of *rebetiko* song lyrics will be presented and then used to perform a computer-assisted analysis of *rebetiko* lyrics on a large scale.

6 See also «Ελενάκι αν δεν σε πάρω» (English “Sweet Eleni, if I don’t Take You”) composed by Kostas Karipis and sung by Giorgos Papasideris (disc number: Columbia DG-489, recorded in 1933), Rita Abatzi (disc number: HMV AO-2112, recorded in 1933) and Roza Eskenazi (disc number: Parlophone B-21732, recorded in 1933) as well as «Κατινάκι μου για σένα» (English “My Katinaki, for You”) composed by Kostas Karipis and sung by Andonis Dalgas (disc number: HMV AO-2078, recorded in 1933), Rita Abatzi (disc number: Odeon GA-1691, recorded in 1933) and Roza Eskenazi (disc number: Columbia DG-447, recorded in 1933).

2.1 THE ORIGINS OF THE TERM REBETIKO

The term rebetiko for a specific form of music became widely accepted not before the second half of the twentieth century (Zelepos 2001, 18). Until then, multiple terms were used interchangeably. Gauntlett suggests that the term rebetiko replaced others due to the publication of an anthology of rebetiko songs by Elias Petropoulos in 1968 (Gauntlett 1983, 82). Additionally, the publication of several articles in the late 1940s and in the beginning of the 1950s has consolidated the genre name rebetiko. In January 1947, Fivos Anoyanakis published an article on rebetiko songs in the newspaper *Rizospastis*.⁷ In this article⁸ titled “The Rebetiko Song”, Anoyanakis was actually the first to discuss the artistic merits of rebetiko and to associate it with the Byzantine chant and the demotic song (Anoyanakis 1961). Another important event was a lecture by Manos Hatzidakis on the value of rebetiko, which he delivered at the Art Theatre of Athens (Theatro Technis) in 1949.⁹ The article “The Worlds of the Folk Art and Tsitsanis” by Sophia Spanoudi was published in 1952 in the Greek daily newspaper *Ta Nea* (Greek *Τα Νέα*). It also had its focus on rebetiko and praised the artistic contributions to Greek music by the rebetiko composer Vasilis Tsitsanis and by singer Marika Ninou.¹⁰ These works contributed to the preference of the genre name rebetiko over other existing names. Besides the above-mentioned texts, it is also important to keep in mind that the process of the commercialisation of the genre, which started in the 1930s with an increase in record production in Greece, also influenced the process of the unification of the genre name.

To get a better understanding of the origins of rebetiko, a discussion of other terms, which used to describe the genre, is useful. According to Gauntlett, terms like mangika¹¹, mortika, seretika, tsachpinika and karipika are almost synonyms with rebetika, while the terms

7 *Rizospastis*, Greek *Ριζοσπάστης*, is the newspaper of the Greek Communist party KKE.

8 The article was also republished in 1961 in the art revue magazine *Epitheorisi Technis*, Greek *Επιθεώρηση Τέχνης* (also known under the French name *Revue d'Art*).

9 The text of Hatzidakis' lecture on rebetiko can be found in (Liavas 2009, 254-260).

10 Spanoudi's article on rebetiko can be found in (Liavas 2009, 261-263).

11 Mangika comes from mangas, pl. manges (Greek sg. *μάγκας*, pl. *μάγκες*). It is a character that appears in many rebetiko songs. According to the Triantafyllidis dictionary, mangas is a “man of the people that is characterised with excessive confidence and conceit as well as with a distinctive appearance and behaviour (dress code, movements, vocabulary, voice etc.)” (cited as in (Manolis Triantafyllidis Foundation 1998). One can define mangas as a ‘wide boy and/or a brave boy’. Manges as a social group existed in Greek cities especially in the first decades of the 20th century, but in rebetiko they continued to exist afterwards, being characterised as fearless outlaws who refuse well established life values, such as family (Tragaki 2009, 26-37). Thus, the figure of mangas had a positive or a negative connotation, depending on the situation in which they appeared.

vlamika¹², koutsavakika¹³ and mourmourika¹⁴ refer to a certain time span or geographical area. For instance, vlamika and koutsavakika were commonly used in Athens and Piraeus of the late nineteenth and early twentieth century. Mourmourika are associated with Smyrna in times before 1922.¹⁵ Other names that were linked to rebetika are chasiklidika¹⁶, or songs of hashish, and songs of prison¹⁷ that define the thematic aspect of songs (Gauntlett 1983, 83).

The etymology of the term rebetiko remains unclear, although a wide range of different hypotheses exists. It is widely accepted that the adjective rebetiko comes from the noun rebetis that, according to the Triantafyllidis dictionary, designates “someone that leads an untroubled, carefree life, and probably marginal, refusing the habits of the society and the official and commonly accepted values” (as cited in (Anagnostou 2018, 3)).

According to Panos Savvopoulos, the term *rebetiko* was an artificial construct, created for commercial purposes by the record industry. He also claims that it follows the Greek state policy of that time of cleaning the language from borrowings, thus it was constructed in order to replace terms with foreign origins such as *μποέμικα* (English ,boemika, bohemian songs‘) or *βλάμικα* (English ,vlamika, songs of vlamis‘) (Savvopoulos 2006, 16-19). This hypothesis is doubtful, because there existed other synonymous definitions for this kind of songs that were of Greek origin: for instance, *κουτσαβάκικα* (English ,koutsavakika, songs of koutsavakides‘). Savvopoulos also claims that the word *rebetis* is a derivative of *rebetiko*, the term that according to the present day knowledge first appeared on a record label produced in 1912 in Constantinople (Savvopoulos 2006, 14-15). The fact that the term *rebetiko* was a commercial construct as well the direction of the word derivation *rebetis* < *rebetiko* does not seem plausible since the word *ρεμπέτα* denoting ,gang of burglars‘¹⁸ was discovered to

12 Vlamika comes from vlamis, pl. vlamides (Greek sg. *βλάμης*, pl. *βλάμηδες*). The word *vlamis* comes from Albanian *vëllam* ,blood brother‘ < *vëlla* ,brother‘ and is similar in meaning to koutsavakis and mangas (Manolis Triantafyllidis Foundation 1998).

13 Koutsavakika comes from koutsavakis, pl. koutsavakides (Greek sg. *κουτσαβάκης*, pl. *κουτσαβάκηδες*). The word *κουτσαβάκης* is derived from *κουτσός*, which stands for ,lame‘. According to the Triantafyllidis dictionary, koutsavakis is similar to the image of mangas and was common for the Athens of the second half of the 19th century. As Zelepos points out, a koutsavakis was a person that had a certain dress code, namely trousers, a shirt, a suit coat and a hat. They used to limp and wear a suit coat only on one shoulder (Zelepos 2001, 33-35). It was a kind of satire on the Athenian fashion at that time to wear European, or to be more precise French clothes. It is remarkable that the Italian word *moda* enters the Greek language in the mid-18th century (and is equivalent to the Greek *συρμός*, eng. ,fashion‘). As Droulia points out, it actually meant the abandonment of the Ottoman and traditional Greek costume and the adoption of the European one (Droulia 1999, 129). The change of the way of dressing as a social and cultural code delivers a particular message. In this case it can be interpreted as the willingness to free oneself from the Ottoman yoke. This strive is understandable, even though it often was overdone, therefore frequently being an object of satire, not only by koutsavakides. Greek poet Alexandros Soutsos, who spent a lot of time in Paris and wore French clothes himself, wrote satires about how faithfully Greeks followed the French fashion (Droulia 1999, 132). More details about the costume as a mean of expressing one’s identity at the times of nation building of Modern Greeks can be found in the article by Loukia Droulia (Droulia 1999).

14 Mourmourika are songs of mourmourides. The onomatopoeic word *mourmouris*, pl. *mourmourides* (Greek sg. *μουρμούρης*, pl. *μουρμούρηδες*) defines a person, who usually mumbles, murmurs.

15 One also finds gramophone records made in Asia Minor after 1922 that are labeled as mourmourika: for example, the instrumental “Vlach Song” (Greek «Βλάχικο») of unknown authorship recorded in Istanbul in 1928 (disc number: Homocord G. 28009). More details at <https://www.vmrebetiko.gr/item/?id=5212>. Accessed November 18, 2021

16 The name ,Chasiklidika‘, Greek *Χασικλιδικά*, comes from the word *χασίσι*, English ,hashish‘.

17 *Τραγούδια της φυλακής*

18 The author of the novel, Minas Chamoudopoulos, explains the meaning of the word *ρεμπέτα*, defining it as «το σύνολον των νυκτοκλεπτών», English ,a group/gang of night burglars‘ (Chamoudopoulos 1871, 11).

already being used about forty times in the novel “Mysterious Night Burglars” by Minas Chamoudopoulos, published in 1871 in Smyrna (Gauntlett, forthcoming, 8). As Gauntlett also points out, the transition from *ρεμπέτα* denoting ‘a gang of burglars’ to *ρεμπέτης* denoting ‘an individual gang-burglar’ resembles the derivation *μάγγα/μάγκα* ‘a band of armed irregulars’ < *μάγκας* ‘wayward individual’ (Gauntlett, forthcoming, 8). The word *ρεμπέτα* appears in a range of other publications after 1871. For instance, it appears in several issues of the weekly satirical newspaper *Romios*¹⁹ and in the satirical verses of the Kefalonian poet Tzortzis Molfetas (Sarandakos 2013).

As already mentioned, there is a wide range of different hypotheses that strive to clarify the etymology of the word *rebetiko*. In his essay on *rebetiko*, Greek composer and songwriter Panos Savvopoulos suggests fifteen different etymologies of this word (Savvopoulos 2006, 13-14). A few popular statements concerning the etymology of this word will be presented. Some scholars associate it with Turkish *rebet asker* (English ‘irregular troops of the Ottoman army’) or Turkish *rebet* (English ‘out of law’). At the same time, these hypotheses find no support as the word **rebet* is not attested in Turkish (Gauntlett 1983, 89). Interestingly, the acknowledged *rebetiko* composer and singer Michalis Genitsaris uses the words *ρεμπέτης* ‘*rebetis*’, *ρεμπετασκέρι* (*ρεμπέτ ασκέρ*) ‘*rebet asker*’ and *ρεμπεσκές* ‘good-for-nothing’ in his autobiography. For example, Genitsaris mentions the fact that his mother used to call him a *rebeskes* ‘good-for-nothing’ because he did neither want to go to school, nor to work²⁰, and adds that it was synonymous to *rebetis*:

Και μ’ έβριζε η μάνα μου ρεμπεσκέ, δηλαδή ρεμπέτη. Γιατί ο ρεμπέτης ήταν βρισιά, αλλά τώρα το ‘χουνε έπαινο να τους λένε ρεμπέτες. (Genitsaris 1992, 6)

“My mom called me a *rebeskes* (a lazy good-for-nothing), that is to say a *rebetis*. Because *rebetis* was a swear word, but now it is a praise for them [musicians] to be called *rebetes*.”

Because of an affray with a policeman that, according to the words of Genitsaris, happened because the policeman broke his bouzouki, a 17-years old Genitsaris was imprisoned for six months. Regarding this incident, Genitsaris comments as follows:

Είχα γίνει και εγώ ένας ρεμπεσκές. (Genitsaris 1992, 14)

“I also became one of the *rebeskedes*.”

It seems that here Genitsaris links the word *rebeskes* to the life of an outlaw. Later in his autobiography he expresses the opinion that one cannot create *rebetiko* and be called *rebetis* without experiencing it, without living as a wayward and outlaw individual (Genitsaris 1992, 72-73). Genitsaris therefore observes *rebetiko* music and a specific way of living as inseparable from each other.

19 The newspaper *Ρωμηός* (*Romios*) existed between 1883 and 1918, and its creator and only author was the satirical poet Georgios Souris. The word *ρεμπέτα* appears in the issues published in 1889, 1894 and 1909 (Sarandakos 2013).

20 As Genitsaris points out several times in his autobiography, already as a child he was only interested in making music. According to Genitsaris, he spent his childhood in the coffeeshop of another prominent *rebetis*, namely Giorgos Batis, which was located opposite to his parents’ house in Piraeus, although his father was strongly against him going there. There, he also learned to play the bouzouki and baglamas, thus according to his own words, at the age of 10 he already learned to play the songs «Ντουντού» (“Doudou”) and «Μεμέτης» (“Memetis”) (Genitsaris 1992, 1-5).

At some point Genitsaris mentions that he and his colleague Giannis Papaioannou lost their jobs because their coffeeshop was closed by the police due to the presence of prostitutes there, and Papaioannou had to find another job to support his mother, whereas Genitsaris did not care about losing a job:

Εμένα δεν μ' ένοιαζε, ήμουν ρεμπετασκέρι. (Genitsaris 1992, 19)

“I did not care about it – I was a rebet asker”

Here, Genitsaris uses the word *ρεμπετασκέρι* ‚rebet asker‘ referring to himself as an individual, although this term is known to define a group of people and not a single person. Genitsaris also uses the word *ρεμπετασκέρι* ‚rebet asker‘ to define a specific way of living that denies traditional values such as building a family or having a permanent job.

Εγώ μόλις βγήκε ο δίσκος, παράτησα και τη δουλειά και άρχισα το ρεμπετασκέρι. (Genitsaris 1992, 16)

“As soon as the disc came out, I quit the job and began a rebet asker.”

Nikos Sarandakos makes the assumption that not *rebetis* was derived from *rebet asker*, but conversely that the term *ρεμπέτ ασκέρ* ‚rebet asker‘ comes from a fusion of *rebetis* and Turkish *asker* ‚soldier‘, and is a Greek word built according to the Turkish word formation pattern, that first appeared in an unspecified newspaper in 1916. The word *ρεμπεσκές* ‚good-for-nothing‘, in turn, is supposed to be a fusion of the pseudo-Turkish phrase *ρεμπέτ ασκέρ* ‚rebet asker‘ (Sarandakos 2013).

Besides linking the genre name *rebetiko* to the phrase ‚rebet asker‘, there also exists a hypothesis that the word *rebetiko* comes from the ancient Greek stem *ρεμβ-*, later *ρεμπ-* and its derivatives such as *ρέμβομαι/ρέμπομαι/ρεμβάζω* (English ‚roam, wander‘) (Gauntlett 1985a, 40). Another assumption is that it comes from Slavic sg. *rebenok* / pl. *rebiata* (English ‚child/lads‘). The plural form *rebiata* corresponds to the Greek *παλικάρι* (English ‚lad, young man‘) (Zelepos 2001, 17), (Gauntlett 1985a, 37-41). Stathis Gauntlett mentions another interesting hypothesis that was proposed by the British Turkologist Victor Louis Ménage, namely that the name of the music genre *rebetiko* comes from the Turkish *harābāti* (English ‚person who spends his/her time in meyhane²¹ / tavern‘) (Ayverdi 2005) that was loaned in turn from Arabic *khārāb* (English ‚ruin‘²²) (Zelepos 2001, 17), (Gauntlett 1985a, 38-39). At the same time, while the omission of the initial *ha-* appears in many Turkish loanwords to other languages, according to Ménage and Tietze²³ the vowel change from *-a* to *-e* usually happens with borrowings into Turkish, but not in reverse direction (Gauntlett 1985a, 39).

Thus, although there exist many hypotheses concerning the etymology of *rebetis*, none of them can be proven conclusively. Ioannis Zelepos points out that one may suppose that these assumptions were created on the basis of the present-day interpretation of *rebetis*. Thus, a reconstruction in the reverse direction was built, such that the words that were suggested as a starting point for the term *rebetis* have the same meaning even if from the

21 A Meyhane is a traditional tavern in Turkey, Iran, Azerbaijan and the Balkans that serves wine or other alcohol, and meze. The lexeme comes from Persian and is a compound of two words: *mey* (English ‚wine‘) and *khaneh* (English ‚house‘).

22 (Redhouse 1880, 274)

23 Austrian Turkologist Andreas Tietze

etymological point of view they are not connected to each other (Zelepou 2001, 17). A semantic change is often observed in diachronic linguistics and can not be excluded in the case of the history of the word *rebetiko*.

In contemporary research, a *rebetis* is usually seen as a roamer, vagabond, carefree person who does not follow the well-established way of living (Savvopoulos 2006, 13). It is also questionable whether it defined the rules of living that rebetes had, or it was an image created from the outside (Anagnostou 2018, 11). A conclusion that can be drawn from the assumptions mentioned above is that the name of the genre *rebetiko* is most likely connected to the way of living of rebetes, but is not connected to the musical characteristics of the genre (as mentioned in the autobiography of Genitsaris (Genitsaris 1992, 72-73)). At the same time, it seems to be important to mention that one cannot equate the representatives of the music genre with the figure of *rebetis* in song lyrics, which actually appears quite late, only in the mid-1930s. The fact that a lot of *rebetika* pick the harsh realities of life as a central theme does not mean that all representatives of the music genre belonged to the underworld and were drug-addicts. Gauntlett points out that Markos Vamvakaris, Stelios Keromitis, Anestis Delias, Giorgos Batis, Dimitris Gongos (Bayiaderas) and Giannis Papaioannou were frequent visitors of the hashish dens (Gauntlett 1985a, 89). At the same time, many of them came in touch with *rebetiko* when they still were children. For example, Markos Vamvakaris, a prominent representative of *rebetiko*, who is often called a patriarch of *rebetiko* due to his influence on the music genre, was drawn to *rebetiko* not because of the above-mentioned theme, but because he fell in love with the sound of the bouzouki at the market where he worked as a butcher²⁴ (even though later in his life he used to go to the hashish dens of Piraeus) (Kounadis 2003a, 27). Michalis Genitsaris who according to his own words spent all his childhood in the coffeeshop of Batis, learned to play his first *rebetika* when he was 10, although he admitted himself that after his first disc appeared, he took 'a bad path' spending all nights with friends playing bouzouki and drinking (Genitsaris 1992, 1-5, 16). Michalis Genitsaris, who was sent into exile on the island of Ios under Metaxas' regime because of being a 'danger to the public', as he himself called the reason for his sentence, also met there exiled Anestis Delias. Genitsaris mentions that he saw Delias taking drugs and reprimanded him openly, making Delias finally quit drugs (Genitsaris 1992, 39-46). Drug consumption, violent affrays and love affairs as well as prison life were part of *rebetiko* (Tragaki 2009, 49), which is testified by interviews with *rebetiko* exponents and their autobiographies.

At the same time, many exponents had no or very loose connections to the underworld. One of the most well-known exponents of *rebetiko* today, Vasilis Tsitsanis, who received violin lessons in Trikala and originally came to Athens to study at the law school, "denied any influence of the mangika (songs associated with the mangas underworld communities) and chasiklidika ('hashish-songs') on his own work" (Tragaki 2009, 62). As Tragaki also

24 According to his own words, Markos Vamvakaris started to write poetry at the age of nine or ten, and wrote his first song when he was fourteen or fifteen years old (Kounadis 2003a, 27). In 1917, when he was only twelve years old, he was forced to leave his home island Syros and go to Piraeus to look for work, due to the difficult financial situation of his family. At first, Vamvakaris worked as a load carrier and then as a butcher. His first encounter with the music happened even before he came to Piraeus. His father was a professional gaida (Southeastern bagpipe) player and also played the bouzouki a little bit. At the market in Piraeus he met a fisherman named Nikos Ayvaliotis, whom he considered to be his only bouzouki teacher. He was not a teacher in a general understanding of a word, because Vamvakaris was learning to play the instrument by hearing (Kounadis 2003a, 25-29).

points out, Tsitsanis avoided the term *rebetiko*, preferring instead the more general term *laiko tragoudi* when referring to his work (Tragaki 2009, 62).

Early records of rebetiko present the work of the professional musicians of café aman, or Asia Minor, style of rebetiko. These professional musicians performed as part of an ensemble (for example, *smyrneiki* or *elliniki estoudiantina*²⁵) in cafés amans, and not in hashish and gambling dens. Main instruments of such ensembles were the oud, the santour, the qanun, and the violin. Such ensembles were sometimes called *santouroviolia*, named therefore after their primary instruments santour and violin (Tragaki 2009, 49). Songs accompanied by the bouzouki²⁶ and baglamas existed since circa 1930 in a kind of closed community of *tekedes*, namely hashish and gambling dens. The professional musicians' ensembles such as *elliniki estoudiantina* were employed by the representatives of foreign record companies who visited the country for a short period of time and did not get entry into the closed world of *tekedes* music (Tragaki 2009, 3).²⁷ This fact means that many recordings, that some scholars consider part of rebetiko, were performed outside of the underworld context, and that some rebetiko exponents had little in common with the frequenters of the hashish dens. As Roderick Conway Morris points out: "Although the café aman was a favorite resort of prostitutes and ruffians and was frequented by the lower strata of society generally, there was a number of more strictly cabalistic criminal venues".²⁸

With the growing demand of record production, the exponents of rebetiko were not able to produce enough song lyrics in the given time and thus used the services of jobbing versifiers, or 'hack-versifiers' (Greek *λόγηδες*) (Gauntlett 1985a, 132-133, 163-164). These jobbing versifiers composed song lyrics not only for rebetiko, but also for other music genres, which resulted in reciprocal influences between the genres (Gauntlett 1985a, 133): the notable lyricist Kostas Manesis, for instance, created song lyrics for *laiko tragoudi* (,popular song'), *rebetiko* and *elaftro tragoudi* (,light (europeanised) song'). The lyricist Charalampos Vasileiadis, who among other things worked as a translator and spoke five languages, is also the author of a range of laiko and rebetiko songs. Some of the so-called *logides* had actually no or only little experience of underworld life with which rebetiko used to be connected. Although they had a great impact on rebetiko, *logides* "tended not to enjoy great esteem among the musicians they served" (Gauntlett, forthcoming). It may explain the designation they got: a neologism of that time *λόγηδες* ,hack-versifiers' < *λόγια* ,words' instead of prevalent *στιχουργ-*

25 The *Elliniki Estoudiantina*, or the Greek *Estoudiantina*, was an orchestra from Constantinople that was formed in the first decade of the twentieth century. *Estoudiantina* defined at that time a small orchestra in Asia Minor that consisted of three to eight members. According to Kounadis, the first *estoudiantina* appeared in Smyrna at the end of the nineteenth century. It was the *estoudiantina Ta politakia* established by Vasilis Sideris, a musician from Constantinople. Another well-known orchestra of this kind was *Smyrneiki Estoudiantina*, for which it is not clear whether it was a general name for such kind of orchestras at that time in Smyrna, or a definition of a specific orchestra whose main singer was Giorgos Tsanakas (Kounadis 2003a, 294-295, 298-299).

26 The first known recording of a Greek song accompanied by a bouzouki was made in 1917 in Görlitz. It is the song «Χήρα ν' αλλάξεις τ' όνομα» (English "Widow, Change Your Name"). According to Gerassimos Alexatos, the singer was Apostolos Papadiamantis, the nephew of the well-known Greek novelist Alexandros Papadiamantis, and the bouzouki player was Kostas Kalamaras from Syros (Alexatos 2018). They were soldiers of the IV Greek Army Corps that was besieged by the Bulgarian forces near Kavala during World War I. The corps was demobilised and brought to the German town of Görlitz by rail. The recording of this song as well as of about 70 other Greek records were made by the Royal Prussian Phonographic Commission (cylinder records). The song uses the rhythm of the Smyrneic song "Manes Bournovalia". Its recording can be found at <https://www.youtube.com/watch?v=JbNzHX9CY0w> (accessed January 24, 2022).

27 More information on café aman and the Greek café music tradition in general can be found in Morris, Roderick Conway. Greek Café Music. In (Bucuvalas 2018)

28 Morris, Roderick Conway. Greek Café Music. In (Bucuvalas 2018)

γοί ,song lyrics writers‘ < στίχος ,verse‘.²⁹ As Gauntlett points out, it is only around 1946 that the lyricists started to demand that their names also appear on the records. When dealing with the songs that were recorded until 1946 (Gauntlett 1985a, 132), it is often difficult to find the author of the text.³⁰

The renowned Greek journalist and ethnographer Konstantinos Faltaits, who is considered to be the first researcher of rebetiko, composed rebetika himself under the pseudonym K. Roumeliotis. His first work that touches the study of rebetiko is considered to be his Athenian novel *Άλλος κόσμος* (English *Another World*) that depicts the world of rebetes and was published in 1915 in the newspaper *Akropolis*. In 1929 in the magazine *Μπουκετο*, Faltaits published lyrics of a number of rebetiko songs under the title *Τραγούδια του μπαγλαμά* (English *Songs of Baglamas*), dividing them according to their thematic affiliation. As Gauntlett points out in this article, he ‘exhorted folklorists to take a serious interest in these “veritable masterpieces of poetic art”’ (Gauntlett, forthcoming).³¹ When it comes to rebetika composed by Faltaits, one should mention the song «Ο ρεμπέτης» (“Rebetis”) sung by Stellakis Perpiniadis (disc number: Columbia DG-2036, recorded in 1934).

The idea that it is not plausible to connect rebetiko directly with the criminals is exemplified in the paper by Anagnostou as follows: “Given the insufficient source material at our disposal, we are unable to substantiate or disprove the mythology that accompanied [...] [rebetiko]. The existence of delinquents and criminals who sang lyrics about hashish may be taken for granted, but we should not assume that they were the only ones or even the first ones to do so” (Anagnostou 2018, 11).

As already mentioned, the word *rebetis* can be interpreted from two different perspectives. Besides the definition as a representative of the music genre itself (as given above), a rebetis can also be seen as a fictional character of rebetiko song lyrics. Gauntlett defines rebetis in the role of a fictional character as follows: “the heroic figure of the bon vivant who, disdaining conventional thrift and seeking solace for the harsh realities of life turns to hashish (in pre-war compositions) or more prestigious intoxications (in post-war compositions)” (Gauntlett 1983, 86).

Earlier the figure of rebetis was introduced. At the same time, although the name of music genre is called rebetiko, rebetis appears quite seldom in song lyrics. According to Savvopoulos, the word rebetis is first mentioned in rebetika song lyrics in 1934 (Savvopoulos 2006, 30). It is possible that Savvopoulos meant the song “Rebetis” (Greek «Ο ρεμπέτης»), composed by K. Roumeliotis (Konstantinos Faltaits) and sung by Stellakis Perpiniadis, which was recorded in 1934 (disc number Columbia DG-2036). According to the data available in the corpus created in the scope of this master’s thesis, the word *rebetis*³² appears already in

29 This assumption is based on a commentary on *λόγηδες* (English ‘jobbing versifiers’) by Nikos Politis (personal communication, 7 December 2021).

30 One of the first well-known writers of rebetiko song lyrics is Nikos Mathesis, who is the author of many song texts that appeared in the 1930s (Gauntlett 1985a, 163).

31 More information about the impact of Faltaits on the research of rebetiko as well as about his other works can be found on the website of the *Faltaits’ Museum* at <https://faltaits.gr/el/>. Accessed December 18, 2021

32 Based on the corpus created in the scope of this master’s thesis, the female form for *rebetis*, namely *rebetissa*, appears first in the song “Mortissa (Tough Girl) of Kokkinia” (Greek «Η μόρτισσα της Κοκκινιάς») by Kostas Karipis (disc number HMV AO-2215). The word *rebetissa* appears in this song in the scope of an interjection addressed to the singer Roza Eskenazi.

1933 in the song “Stoner”³³ (Greek «Χαρμάνης»³⁴) by Markos Vamvakaris. This is a stanza from this song by Vamvakaris:

Σ' αυτόν τον ψεύτικο ντουινιά	In this fake world
ντερβίσης θ' αποθάνω,	I am going to die as a dervish,
σ' αυτόν τον ψεύτικο ντουινιά	In this fake world
ρεμπέτης θ' αποθάνω	I am going to die as a rebetis

One of the central figures of rebetiko is actually mangas (Anagnostou 2018, 3) who is closely connected to the figures of alanis, ntervisis/dhervisis (dervish, hashish smoker), mortis, vlamis, seretis and koutsavakis.³⁵ Mangas represents an outlaw person that is associated with a particular excessive appearance and behaviour, dress code as well as a slang idiom, the so-called mangika. In the corpus of rebetiko song lyrics the word mangas first appears in the song “My Sweet Mother” (Greek «Μανάκι μου») of unknown authorship and sung by Andonis Dalgas. Here is a stanza from this song³⁶, which was recorded in 1926:

Μανάκι μου και ντες και ντες	My sweet mother and des and des (exclamation)
βρε, δεν θα 'βρεις μάγκκα σαν κι εμέ	Hey, you won't find such a brave guy like me

According to Anagnostou, the figure of mangas as a person that belonged to a particular social group, the so-called *mangiko synafi* (the mangas community)³⁷, appeared in the first two decades of the twentieth century. Similar to the case with rebetis, it is also unclear if the group with characteristics typical for the mangas community was created by its representatives or from the outside. In the 1930s, the term mangika as songs of or about manges (and other relevant terms such as seretiko or antamiko) has appeared in the press (Anagnostou 2018, 3). According to Tragaki, the first appearance of the word *rebetis* only in the mid-1930s suggests that the genre name as well as the figure of *rebetis* itself came from the outside of Greece due to the commercialisation of the genre (Tragaki 2009, 24).

A frequency analysis of the usage of words mangas and rebetis as well as their derivatives in rebetiko song lyrics gives interesting results (the corpus created in the scope of this

33 Also according to Gauntlett, the song “Stoner” by Vamvakaris is the first to mention the term rebetiko in lyrics (Tragaki 2009, 42).

34 Other possible translations of the song title are ,addict longing for his fix‘ or a ,person dying for a smoke‘. The song appeared on the recording with disc number Odeon GA-1674.

35 As Savvopoulos points out, the same words were also used by the Greek novelist Alexandros Papadiamantis to describe the underworld of Greek cities (Savvopoulos 2006, 30). Interestingly, besides the synonyms of mangas mentioned above with a positive connotation there exist also the derivatives with a negative connotation: for example, *ψευτόμαγκας*, English ,fake mangas‘, or *πουστόμαγκας*, English ,fag mangas‘ (Tragaki 2009, 32).

36 Disc number: His Master’s Voice AO-164

37 More about the concept of mangiko synafi as well as the figure of mangas can be found in (Tragaki 2009, 26-37).

master's thesis contains 5,167 songs, see Section 3.2.3). A query for the stem *ρεμπέτ-/ρεμπετ-* results in 28 songs, while the one for the stem *μάγκ-/μαγκ-* results in 359³⁸ songs.³⁹

When speaking about the term *rebetiko*, it is important to also mention the fact that especially in the first two or almost three decades of the twentieth century, it was unusual to put the genre name on the gramophone record labels for songs that are considered to be *rebetika* nowadays. Instead of a genre affiliation, one can often find information such as vocal type (for example, soprano or tenor), instrumentation or the corresponding rhythm (dance) for a particular song. Many gramophone records that were produced before 1930 and are considered to be classical *rebetika*⁴⁰ have generic labels such as popular song, folk song or *δημώδες* (English 'demotic') or just Greek song. Especially the gramophone records made by the American record company Columbia often had labels like popular song or folk song. It is important to bear in mind that most of the Columbia records made before 1930⁴¹ were recorded in the USA (especially in New York and Chicago), and genre affiliation could help the record company to sell the product. Obviously, it is absurd to define the genre of a record solely based on the information of its label.⁴² Also, equivalent names for *rebetiko* such as *mortiko* or *tsachpiniko* were rare at that time in comparison with the more generic ones mentioned above. Figure 2.1 shows an example of a *rebetiko* record without any genre affiliation.

On this record one can find information about the composer (Dimitris Semsis) and singer (tenor Andonis Diamantidis (Dalgas)). The label also provides the information that the song was accompanied by the *laiki* 'popular' orchestra (Greek *λαϊκή ορχήστρα*) without specifying the included instruments. In general, the instrumentation or orchestration of the recordings cannot be taken for granted and be equated with the live *rebetiko* performance tradition of that time. Quite often the record companies employed professional ensembles that worked in multiple different musical genres (Tragaki 2009, 3).

In the context of the recording of the song "Tempress from Smyrna" (Greek «Σμυρνιακά καμωματού») it is also remarkable that even the nickname of its singer, namely Dalgas, is associated with the genre of *rebetiko*. Andonis Dalgas, who is one of the most well-known singers of the Smyrna school of *rebetiko*, was born in Constantinople. He got his nickname, which comes from Turkish *dalga* 'wave, ripple, dizziness, intoxication', due to unusual waves in

38 When searching for such a stem, it is important to keep in mind that some other words that are not connected with that of *mangas* can be included. This is, for example, the case with the word *magioros* (Greek *μαγκιώρος*) that comes from Italian *maggiore* < Latin *major*, comparative of *magnus* (English 'big'). It has therefore nothing to do with the word *mangas* and is usually used to refer to outstanding abilities of a person in some area, i.e. is similar to the meaning of 'master'. The number of such cases in the search result is very low, though, and does not influence the percentage ratio of the occurrence of the words *rebetis* (Greek *ρεμπέτης*) and *mangas* (Greek *μάγκας*) in the corpus of *rebetiko* song lyrics.

39 The search was based on the stems in order to get all the forms of the words *rebetis* (Greek *ρεμπέτης*) and *mangas* (Greek *μάγκας*) as well as their derivatives such as *mangaki* (Greek *μαγκάκι*) or *mangitis* (Greek *μαγκίτης*), for example.

40 It seems important to mention that the term 'classical *rebetiko*' (which mostly refers to *rebetika* songs that appeared between 1930 and 1940) may be problematic, as one might assume that *rebetika* that were created outside of this period would miss some characteristics of *rebetiko* to be defined as classical, which would be a mistake. Music genres in general are not stable and undergo certain transformations. More details on the term 'classical *rebetiko*' as well as 'pseudorebetiko' ('fake *rebetiko*') can be found in (Gauntlett 1985a, 22). *Rebetika* of the 1930s also missed some features of *rebetika* of the earlier period.

41 The first four decades of the 20th century are seen today as the heyday of *rebetiko*.

42 For example, out of 2,820 78rpm records presented in the *Virtual Kounadis Archive* only 54 records have *rebetiko* mentioned on their disc metadata ("Εικονικό Μουσείο Αρχείου Κουνάδη [Virtual Museum of the Kounadis Archive]" 2019).



Figure 2.1: Gramophone Record of the Song “Tempress from Smyrna”, Greek «Σμυρνια κάμοματού». Recorded in 1929 in Athens (disc number: Pathé X. 80039). Image Source: *Virtual Kounadis Archive* (URL: <https://www.vmrebetiko.gr/item/?id=5256>, accessed December 2, 2021)

his voice. The word *νταλγκάς/νταλκάς*, which also exists in Modern Greek and stands for ‘strong desire, passion’ (Manolis Triantafyllidis Foundation 1998), is a word used in rebetiko lyrics.⁴³ It is synonymous to *μεράκι*, which also stands for ‘strong desire’ and is more commonly used in rebetika than the previous one.⁴⁴

When it comes to the genre affiliation, Markos Vamvakaris, for example, during his lifetime was often seen as a composer of laiko music (Kounadis 2003a, 105). This is due to the fact that rebetiko until about the 1950s was seen as a part of laiko and only later on was recognised as an independent music genre.

Often, one and the same song was recorded by different record companies at almost the same point in time and got different genre affiliations. This was, for example, the case with the song «Τα λερωμένα τ’ άπλυτα», English “The Dirty and Unwashed”, of unknown authorship. In 1929, a record of this song, performed by Andonis Dalgas appeared (disc number: Pathé X. 80036, Athens). This gramophone record was labeled as rebetiko.⁴⁵ Four years later,

43 The word *νταλγκάς/νταλκάς* appears in 40 songs of the *Rebetiko Corpus*.

44 The word *μεράκι*, which is an Arabic word borrowed in Greek via Turkish (Greek *μεράκι* < Turkish *merak* < Arabic *maraq*), appears in 226 songs of the *Rebetiko Corpus*.

45 The record as well as the respective metadata and the label of the record can be found in the *Virtual Kounadis Archive* at <https://www.vmrebetiko.gr/item/?id=4479>. Accessed November 17, 2021

in 1933, the record of “The Dirty and Unwashed” sung by Rita Abatzi was published (disc number: Columbia D.G. 363, Athens).⁴⁶ This gramophone record was labeled as kleftiko (klephtic song)⁴⁷ although there are no differences on a musical or lyrics level to the first one.

An even more striking example of how relative the genre affiliation in the scope of rebetiko was, is the song “The Girl from Smyrna” (Greek «Σμυρνιά») of unknown authorship, which was recorded by two different companies in the USA in 1924, but sung by one and the same person, namely Marika Papangika. The first record in question was made by the Greek Record Company⁴⁸ in Chicago (disc number: Greek Record Company 511-A) and labeled as rebetiko in both English and Greek.⁴⁹ The second record of this song was made by the Columbia Record Company in New York and labeled as a folk song (disc number: Columbia USA 7011-F).⁵⁰ As in the case with the song “The Dirty and Unwashed”, these two records do not differ on a musical or lyrics level.

It is also interesting to look at the recording of the song “Mangas (Tough Guy, Outlaw)” (Greek «Μάγκας») of unknown authorship and sung by Andonis Dalgas. It was recorded in Athens in 1929 (disc number: Pathé X. 80039).⁵¹ The record is labeled as rebetiko, although as Leonardos Kounadis and Nikos Ordoulidis point out, it is based on a melody of a traditional (paradosiako) song of Greeks of Constantinople.⁵² This song appeared also under other names: “Chasapiko of Constantinople” (Greek «Χασάπικο πολίτικο») recorded in 1926 in Athens (disc number: His Master’s Voice AO-178), “The Vlamis” (Greek «Ο βλάμης») recorded in 1929 in Athens (disc number: His Master’s Voice AO-377), “The Mangas (Tough Guy)” (Greek «Ο μάγκας») recorded in 1927 in Athens (disc number: Odeon GA 1433) and “I don’t Want You anymore, Sweet Butcher” (Greek «Χασαπάκι δεν σε θέλω πια») recorded in 1931 in Athens (disc number: Polydor V 51083). The records of “The Vlamis” of 1929, “The Mangas (Tough Guy)” of 1927 and of “Chasapiko of Constantinople” of 1926 did not have any genre affiliation on the record label, while the record of the song “I don’t Want You anymore, Sweet Butcher” of 1931 was labeled as *δημώδες* (English ,demotic‘). This example shows one more time, how generic the formal genre affiliation in the case of rebetiko

46 The record as well as the respective metadata and the label of the record can be found in the *Virtual Kounadis Archive* at <https://www.vmrebetiko.gr/item/?id=9932>. Accessed November 17, 2021

47 As klephtic songs, or ballads (Greek *Κλέφτικο τραγούδι*), one defines songs of klephts (Greek sg. *κλέφτης* which means ,thief, brigand‘). Klephts were freedom-loving members of disobedient groups in Greece, that lived in the mountains and fought against the Ottoman rule. They made their living with raids and robberies. Such groups existed from the sixteenth century until the end of the Greek independence war of 1821-1829, when they merged with the other warriors. The klephtic songs dealt with the heroic deeds and the life of klephts in general (Manolis Triantafyllidis Foundation 1998) (Trypanis 1981, 592-594).

48 Interestingly, the song was accompanied by the orchestra of the Greek Record Company. As already mentioned, there can be substantial differences regarding the instrumentation in recordings and the instrumentation typical for live rebetiko concerts. Some record companies hired musical ensembles that performed a number of different music genres, others even owned their own musical ensembles (Tragaki 2009, 3), which is the case of the Greek Record Company.

49 The record produced by the Greek Record Company as well as the metadata concerning the record and the label of the record can be found in the *Virtual Kounadis Archive* at <https://www.vmrebetiko.gr/item/?id=4927>. Accessed November 17, 2021

50 The record produced by the Columbia Record Company as well as the metadata concerning the record and the label of the record can be found in the *Virtual Kounadis Archive* at <https://www.vmrebetiko.gr/item/?id=4609>. Accessed November 17, 2021

51 The recording as well the corresponding metadata can be found at <https://www.vmrebetiko.gr/item/?id=4484>. Accessed November 17, 2021

52 The commentary of Leonardos Kounadis and Nikos Ordoulidis on the song “Mangas” can be found at <https://www.vmrebetiko.gr/item/?id=4484>. Accessed November 18, 2021

and other Greek music genres was. This ambiguity also shows that the borders between the music genres are in most cases unclear. One finds many cases where popular rebetiko songs use the melodies or complete songs known from traditional (*paradosoaka*) songs. Kounadis and Ordoulidis make another interesting observation, namely that the melody used in the song “Mangas” appears in the songs of neighbouring music traditions: for example, in the Jewish songs “Mein Velvele”⁵³ by Aaron Lebedeff and “Malkele”⁵⁴ by Morris Goldstein as well as in the Bulgarian song “Да ви кажа братя големия си дерт”⁵⁵ (English ‘Let me Tell You, Brothers, my Biggest Grief’) by Constantin Bambov.⁵⁶

At this point it seems important to mention the fact that a ‘genre’ is, as Tragaki defines, a construction (Tragaki 2009, 23). In the case of rebetiko it is a construction that was to a certain extent influenced by the record companies. The first Greek song was recorded in the USA in 1896, and the first rebetiko songs were recorded outside of Greece. It still remains unclear, whether the first rebetika were recorded in Constantinople, Smyrna or the USA. Ole Smith even argues, that the term rebetiko was imported to Greece via American recordings (cited as in (Tragaki 2009, 24)). It is in this context of interest, how the record market influences a particular music genre and its song lyrics.

2.2 CHRONOLOGICAL BOUNDARIES OF REBETIKO

It is very difficult to define the chronological boundaries of rebetiko. There is a disagreement on both the date of its origin as well as its decline. Some researchers place the roots of rebetiko in the second half of the nineteenth century, other set the lower boundary to the year 1922 and the following mandatory exchange of population between Turkey and Greece and therefore the arrival of refugees to Greece (Anagnostou 2018, 15). The adherents of the point of view that rebetiko originates in 1922 claim that the roots of this musical genre lie in Asia Minor, and that it was brought to mainland Greece by the refugees from the Asia Minor. Zelepos, contradicting this hypothesis, mentions the fact that the word *rebetiko* had been documented outside the Asia Minor already before 1922 as well as that there are gramophone records from the USA dating back to the first two decades of the twentieth century which were performed by migrants from Greece.⁵⁷ Moreover, the study of rebetiko

53 The song “Mein Velvele” by Aaron Lebedeff was recorded in Yiddish in New York in 1926. More information at https://adp.library.ucsb.edu/index.php/matrix/detail/2000222538/E4033-E4034-Mein_Velvele. Accessed November 18, 2021

54 The song “Malkele” by Morris Goldstein was recorded in Yiddish in New York in 1923. More information at https://adp.library.ucsb.edu/index.php/matrix/detail/2000214981/OK_cat_14076-b-Malkele. Accessed November 18, 2021

55 The song “Да ви кажа братя големия си дерт” (English “Let me Tell You, Brothers, my Biggest Grief”) by Bambov was recorded in 1927. The recording of the song can be found at <https://youtu.be/IUgzFEPYQMA>. Accessed November 18, 2021

56 The commentary of Leonardos Kounadis and Nikos Ordoulidis on the melody used in the song “Mangas” as well as in the songs of the neighbouring music traditions can be found at <https://www.vmrebetiko.gr/item/?id=4484>. Accessed November 18, 2021

57 Between 1890 and 1924, about 500,000 migrants came to the USA from Greece, leaving aside the Greeks who came from the territories that were under the Ottoman/ (Turkish) control (Kounadis 2003a, 255). For comparison, the population of Greece in 1890 was about 2.2 million and in 1924 about 6.5 million (Kostis 2018, 162, 275), so the number of migrants was almost 10% of the population of Greece of 1924. Most Greeks were migrating due to economic reasons (high taxes, unlucky war with Turkey in 1897, currant crisis and finally the bankruptcy of Greece in 1893) in a hope of finding a better future. The USA though did not welcome the migrants in the way they expected. No one who was ill or a criminal was allowed to enter the USA, so all the refugees were checked by a doctor, whether they were healthy and only then they could enter the American soil. Most immigrants

and Greek music in general, without considering neighbouring music traditions, is doomed to failure because like every other music tradition the Greek music did not exist in a vacuum without being influenced by other genres and music traditions (Zeleps 2001, 23-24). Besides the influences of the Byzantine chant and Turkish music tradition, Greek popular music and rebetiko in particular experienced influence by Jewish⁵⁸, Spanish, Armenian and Russian music traditions (it was to a certain extent a reciprocal influence).⁵⁹ An interesting example of a song that exhibits influences of neighbouring music traditions is one of the most well-known amanedes⁶⁰ of the Greek population of the Asia Minor (Kounadis 2010, 8-9), namely “The Smyrna Amanes in Minor” that often also appeared under other names. Two of these records are the “Smyrna Amanes in Minor” (Greek «Σμυρνέικο μανέ μινόρε») recorded in 1911 and sung by Lefteris Menemenlis (disc number: Favourite Record 1-55017)⁶¹ and the “Smyrneic (amanes) in Minor” (Greek «Σμυρνάικο μινόρε») recorded in 1919 and sung by Marika Papangika (disc number: Columbia USA E-7151).⁶² As Nikos Ordoulidis points out, both amanedes have a conclusion that differs rhythmically from the main part⁶³ of the song, the waltz melody that is derived from the Old Russian Waltz “Ex-

were doing difficult physical jobs, for example, building the railways. Due to the harsh life realities, they were looking for a shelter in the songs of their home country, especially in demotic and rebetiko songs. This kind of music was performed in coffee shops where migrants gathered after a hard work day. From 1924 onwards the number of Greek migrants to the USA sank because the USA introduced stricter migration laws (Kounadis 2003a, 255-270).

- 58 During the first decades of the twentieth century, Constantinople and Smyrna had a large Jewish community. Thessaloniki also had a large Jewish community and was even nicknamed *Jerusalem of the Balkans* (in 1913, for example, the Jewish community constituted one third of the population of the city, thus being the largest ethnical group of the city until the beginning of the Second World War).
- 59 More details can be found in the text by Leonardos Kounadis and Nikos Ordoulidis at <https://www.vmrebetiko.gr/item/?id=4083>. Accessed November 17, 2021.
- 60 (*A*)*manes*, Plural *Amanedes*, a vocal improvisational solo that was usually performed as a concluding part of a song, is a firm component of Smyrna rebetika. At the same time, it presents an independent musical form (Zeleps 2001, 70, 110). It is similar to Turkish *gazel*. The choice of dromos/ makam was defined by the instrument that was playing the main melody, which in most cases until 1930 was a violin, or in rarer cases a politiki lyra, or Constantinopolitan lyra. As Tragaki points out, amanes was often accompanied by only one music instrument (Tragaki 2009, 15). The amanes usually consists of three parts: intro, main part and finale, and is sung by only one person (Kounadis 2003a, 363-365). The lyrics of amanes usually consist of couplets in fifteen-syllable metre. Although as an Oriental musical element amanes had a very strong influence on rebetiko, it started to disappear after the establishment of the dictatorship of Ioannis Metaxas due to the censorship under which all ‘alien’ elements were undesired in Greek music (Kounadis 2003a, 363-365). Other popular amanedes are “Tzivaeri”, “Tambachaniotikos” and “Antamaman”. The name of this music form comes from the word *aman*, which is a borrowing from the Ottoman Turkish and was often repeated in amanedes as a way to gain time to think of a next line of text. One can distinguish between the Oriental and the Western amanedes, the difference of which is often reflected in their names: for example, “Sambach mane” or “Rast mane” use “Oriental” melody types, i.e. dromoi/ makams, and “Amanes in Major” (Greek «Ματζόρε μανές») or “Amanes in Minor” (Greek «Μινόρε μανές») use the Western equal tempered scale and the Western classical music modes. At some point at the beginning of the 20th century, when the active recordings of amanedes began in Asia Minor, this music genre became popular in the bourgeoisie. This fact influenced the record production of amanedes and their westernisation. Nikos Politis makes an interesting remark about the amanes “Smyrna Amanes in Major” (Greek «Σμυρνέικο μανέ ματζόρε») sung by Lefteris Menemenlis, declaring that it was actually performed in hijaz and not in major (Politis 2006).
- 61 More details about the recording of “Smyrneic Amanes in Minor” by Menemenlis can be found at <https://rebetiko.sealabs.net/display.php?d=0&recid=17152>. Accessed November 17, 2021
- 62 More details about the recording of “Smyrneic (Amanes) in Minor” by Papangika can be found at <https://www.vmrebetiko.gr/item/?id=4602>. Accessed November 17, 2021
- 63 The musical structure of amanes can also be described in a more fine-grained manner than mentioned before (intro, main part and finale). Some amanedes were pure vocal compositions without instrumentation. As already mentioned, the lyrics of amanes in most cases consisted of couplets in fifteen-syllable metre. Those amanedes that were accompanied by one or more musical instruments had in most cases the following structure: intro -

pectation” by Herold Kittler. Ordoulidis suggests that this melody entered the Greek music repertoire through the Jewish music tradition.⁶⁴ Another example that shows a reciprocal influence between the Greek music tradition in the context of rebetiko and the neighbouring music repertoires was presented in the previous section in the case of the song “Mangas” sung by Dalgas. As it was shown, this song is connected to the Jewish and Bulgarian music traditions. One could find a number of other examples that testify the cross-cultural evolution of a particular melody or a song as a whole. It is important to bear in mind that the interaction between different music genres, either with those from the past or those existing simultaneously with the genre in question, has always existed.

Rebetiko was also influenced by Greek genres that were popular at the same time, such as *dimotiko tragoudi* ‘demotic (folk) song’ and *laiko tragoudi* ‘popular song’, *elafro tragoudi* ‘Greek light song’ and Greek revue theatre. Moreover, rebetiko exponents often crossed the genre borders and worked in several music genres simultaneously: for example, Roza Eskenazi sang rebetika, kleftika, dimotika and laika songs. In the previous paragraph, it was mentioned that rebetiko experienced influence by neighbouring music traditions such as Jewish, Armenian or Turkish. Due to the fact that many representatives of rebetiko came from such cosmopolitan cities back then, like Smyrna and Constantinople, their artistic work was a direct fusion of different music traditions: some exponents of rebetiko spoke not only Greek and Turkish, but also Armenian. Roza Eskenazy, who was born to a family of Sephardic Jews in Constantinople, sang in three languages, namely in Greek, Turkish and Armenian. The prominent music instrument players and singers Kostas Karipis and Andonis Dalgas, both originally from Constantinople, also sang in both Greek and Turkish. Some of their songs present a mixture of Greek and Turkish text: for example, the song «Τσιτφτετέλι (Θα σπάσω κούπες)» (English “Tsittfteteli (I will Break Cups)”) of unknown authorship and sung by Kostas Karipis (disc number: Polydor V-50254, recorded in 1928). As it was already mentioned, the so-called logides, or jobbing versifiers, had a great impact on rebetiko, and also brought into rebetiko elements from Greek music genres such as *elafro tragoudi* ‘Greek light song’. Some rebetiko exponents were also participants of the Greek revue theatre (*epitheorisi*) or Greek operetta: for example, Petros Kyriakos and Giannis Ioanidis. The influence of *epitheorisi* may be seen in the presence of extended dialogs in some rebetiko songs (as for example, in the song “Five Years in Prison (The Sound of a Hookah Pipe)” by Vangelis Papazoglou). These examples show one more time how flexible a genre border is.

We will now return to the question about the chronological boundaries of rebetiko and will try to present different hypotheses concerning this issue. According to Savvopoulos, the word *rebetiko* first appeared between 1910 and 1913 on the label of two recordings (Savvopoulos 2006, 26-27). The word from which the term *rebetiko* is supposed to be derived from, namely *ρεμπέτα* denoting ‘a gang of burglars’, appeared, as already mentioned in the previous section, much earlier, namely in the novel “Mysterious night burglars” by

eight-syllable hemistich (half-line) of the first line - short interlude/ break - seven-syllable hemistich of the first line - short interlude - refrain of the seven-syllable hemistich - interlude - second fifteen-syllable line without break - conclusion/ *gyrisma* (Greek *γύρισμα*), when the rhythm of the song changed (Politis 2006). For *gyrismata* (Plural) one often chose well-known waltz, polka or, for example, ball melodies, such that they were relaxing for the audience after intense amanes (Dragoumis 2003, 168).

⁶⁴ It remains unclear, whether this melody was originally Russian or Jewish. The text by Ordoulidis as well as more details about the recording of “Smyrneic (Amanes) in Minor” by Papangika can be found at <https://www.vmrebetiko.gr/item/?id=4602>. Accessed November 17, 2021

Minas Chamoudopoulos published in 1871 in Smyrna. Savvopoulos also points out that the term *rebetiko* coexisted with other genre names: for example, *koutsavakiko*, *mangiko*, *mortiko*, *tsachpiniko*, and *mourmouriko*.

One reason why it is difficult to trace the origins of *rebetiko* is because, as already mentioned, for a long time the term *rebetiko* was not commonly used to define this music genre. According to the data available today, the term *rebetiko* referring to a specific music genre first appears on the labels of records made in Constantinople and Smyrna between 1910 and 1913 (Anagnostou 2018, 4). One of them was most probably published in 1912 in Constantinople by the record company *Orfeon Record*. This disc (see Figure 2.2) has on one side the song “Aponia” interpreted by the orchestra *Elliniki Estudiantina* (Savvopoulos 2006, 14-15).⁶⁵ As one can also see on the record, the genre name *rebetiko* is put in brackets. The peculiarity of using the word *rebetiko* was also noted by Stathis Gauntlett, who mentions the fact that this position was usually reserved for the description of the dance that accompanied the song (Gauntlett 1985a, 32). According to Savvopoulos, the first time that the word *rebetiko* appeared in a printed publication was in 1917 in the magazine *Σφιγξ* (English *Sphinx*) in Cairo (Savvopoulos 2006, 19).



Figure 2.2: First Known Gramophone Record with the Genre Name Mentioned on a Label. Image Source: Cover of the Book by Savvopoulos

⁶⁵ According to Savvopoulos, the song “Aponia” came from the theatrical tradition of revue and was very popular in Smyrna (Savvopoulos:15).

Another reason why at the record labels the term *rebetiko* came into use so late, might be that until 1930 most record companies were located outside Greece and were not Greek. These companies often put only the English names of the genres on the label. For example, folk song, popular song or Greek bum song.⁶⁶ One reasonable hypothesis is that often the records targeted not only the Greeks and thus the title had to be more understandable to others (e.g. records produced in the USA by the Columbia record company).⁶⁷ In the first and second decade of the twentieth century even more often the records of Greek music had no information about genre affiliation at all. Besides the name of the song written in Greek and Latin letters and the disc number it more often provided information about instrumentation, involved artists, vocal characteristics and rhythm.⁶⁸

At the same time the fact that such music was not called *rebetiko* does not mean that it did not belong to the *rebetiko* genre. *Rebetiko* due to the similarities in its origins is often defined as Greek blues. Blues also faced similar discussions regarding its origins. As Lynn Abbott and Doug Seroff underline, “the era of popular blues music was not suddenly set into motion by Mammie Smith’s 1920 recording of “Crazy Blues”” (Abbott/Seroff 1996, 402). It began much earlier as the first blues composition, namely Antonio Maggio’s “I Got the Blues”, was published in 1908 and goes back to the 1860s.

Before the first song with genre name *rebetiko* on the label appeared in 1912, one finds records of songs that, even if we might not call them *rebetika*, influenced the genre to a such an extent that almost every well-known *rebetiko* musician reinterpreted them in his or her own way. As a prime example, the song “Smyrna Song in Minor” (Greek «Σμυρνέϊκο μινόρε») could be considered, which is the most well-known *amanes*. This *amanes* is known under a number of different names, such as «Μινόρε της αυγής» (English “Dawn Song in Minor”), «Μινόρε μανές» (English “(A)manes in Minor”), «Σμυρνέϊκος μανές» (English “Smyrna (A)manes”) or just «Μινόρε» (English “Song in Minor”). It was recorded as well as reinterpreted many times (according to Nikos Ordoulidis⁶⁹, it was recorded more than 60 times). Kounadis points out that as a melody the “Smyrna Song in Minor” had the largest influence on *rebetiko* as a whole (Kounadis 2003a, 368). The first documented recording of this song dates back to 1907 and has the title «Μινόρε της αυγής» (English “Dawn Song in Minor”). According to Kounadis, it was already well-known at the end of nineteenth century in the Asia Minor (Kounadis 2003a, 368).⁷⁰ This *amanes* has elements that are today associated with classical *rebetika*, namely the broken voice of a singer, references of a singer to the composer or the music performer of a song (extended spoken interjections) as well as interjections such as *aman*. Below an excerpt of the text of “(A)manes in Minor” («Μινόρε μανές») recorded in 1909 either in Smyrna or Thessaloniki, is presented:

66 On the term *Greek bum song* used for labelling *rebetiko* records see (Gauntlett 1985a, 32).

67 *Rebetiko* songs recorded by Columbia often had labels such as folk song or popular song, for instance, songs like “The Sweet Island Girl” (Greek «Νησιωτοπούλα») of unknown authorship and sung by Stavros Kaloumenos (recorded in New York in 1932, disc number: Columbia USA 56306-F) and “Enamoured Girl” (Greek «Η σεβνταλού») of unknown authorship and sung by Dimitris Atraidis (recorded in 1931, disc number: Columbia USA 56303-F) (Source: *Virtual Kounadis Archive* <https://www.vmrebetiko.gr>, accessed January 4, 2022).

68 The *Virtual Kounadis Archive*, gr. *To Αρχείο Κουνάδη*, contains 2,820 78 rpm records performed by the Greeks between 1900 and 1990 and provides the disc record images as well as a short description including metadata to every record which facilitates further research on *rebetiko*. The lyrics texts are not provided. One further value of the *Virtual Kounadis Archive* is that it includes many rare recordings from the first two decades of the twentieth century, that are not available elsewhere on the Web.

69 More information is available at <https://www.vmrebetiko.gr/item/?id=4602>, accessed November 16, 2021.

70 More information about “Smyrna Song in Minor” as well as about the genre of *amanes* in general can be found at <https://www.vmrebetiko.gr/item/?id=4424>. Accessed November 16, 2021.

- Αμάααν!
Ενώ, σκληρά, δεν μ' αγαπάς, γιατί πλέον δεν
παύεις,
μόνο μου δίνεις βάσανα και το κορμί μου κού-
εις.

- Aman!
If you don't love me, you cruel girl, why
don't you leave,
You only torture me and burn my body.

This amanes was sung by Giannis Tsanakas (disc number: Odeon No-54728). The hero of the “(A)manes in Minor” is rejected by the ‘heartless’ beloved, who continues to torture him. Love, and in most cases unrequited love, is one of the central topics not only of amanedes but also rebetika. The presented amanes is an example of a metric form typical for amanedes: namely, a couplet in fifteen-syllable metre.

As Gauntlett points out, some rebetiko song texts include folk couplets recorded by the folklorists in the mid-19th century. As he mentions, the couplet presented below appears in the song «Από κάτω απ’ τις ντομάτες» (English “Under the Tomatoes”) of unknown authorship and sung by Giannis Ioannidis (disc number: Columbia USA 56137-F, recorded in 1928 in New York) as well as in the anthology⁷¹ of Greek Folk songs by Arnold Passow, published in 1860 (Gauntlett 1985a, 173) (Gauntlett 1985b, 222-223).

Βάρα με με το στιλέτο
κι όσο αίμα βγάλω πιά το

Stab me with a dagger
And as much blood as comes out, drink it.

The next couplet appears, as Gauntlett states, in the *Anthology of the demotic songs of Epirus* by Panayiotis Aravantinos published in 1880⁷² (Gauntlett 1985a, 172):

Το μαντήλι σου διπλώνεις
και θαρρώ πως με μαλώνεις

You are folding your handkerchief,
And I think that you are scolding me.

This couplet also appears in the above-mentioned song «Από κάτω απ’ τις ντομάτες» (English “Under the Tomatoes”) as well as in the song «Ελενάκι»⁷³ (English “Elenaki (Sweet Eleni)”).⁷⁴ Almost all available records of the song «Ελενάκι» do not have a genre affiliation, except for the recording by Marika Papangika that is categorised as a *folksong & dance* (disc number: Columbia USA 56034-F, recorded in New York in 1926), and a recording by Elliniki Estoudiantina that is categorised as a *chorus* (disc number: Concert Record Gramophone,

71 The anthology “Τραγούδια Ρωμαίικα: Popularia Carmina Graeciae Recentioris” by Arnold Passow includes eight categories of songs, among which are the klephtic and Vlach songs, erotic songs and couplets. The above-mentioned couplet is categorised as an erotic couplet (Passow 1860, 497).

72 Aravantinos, Panayiotis. 1880. *Συλλογή Δημοδών Ασμάτων της Ηπείρου [Anthology of the Demotic Songs of Epirus]*. Rethymno: University of Crete.

73 The song «Ελενάκι» (English “Elenaki”) was recorded at least seven times either in its original or in a reinterpreted way. Its first known record was made in Smyrna in 1911 in a mixture of Turkish and Greek (disc number: Concert Record Gramophone, recorded in Smyrna in 1911). The couplet presented above can be found, for instance, in the records sung by Marika Papangika (disc number: Columbia USA 56034-F, recorded in New York in 1926) and Giorgos Vidalis (disc number: Odeon GA-1153, recorded in Athens in 1926). More information about the available records of «Ελενάκι» as well as its recordings can be found at <https://www.vmrebetiko.gr/item/?id=4355>. Accessed December 20, 2021

74 Elenaki is a diminutive form on *-aki* from the personal name *Eleni*.

recorded in Smyrna in 1911). The song «Από κάτω απ' τις ντομάτες» (English “Under the Tomatoes”) is categorised on a record label in Greek as *ρεμπέτικο ζείμπέκικο* (*rebetiko zeibekiko*) and in English as a *bum song*.⁷⁵

This at the same time does not mean that rebetika “can be traced back to 1842⁷⁶, 1860⁷⁷, 1871⁷⁸, and 1880⁷⁹, does not necessarily mean that they were ab initio rebetika and that the tradition of the whole genre dates from that time (or beyond)” (Gauntlett 1985a, 172). As Gauntlett also points out, these couplets appear not only in rebetiko, and may have been borrowed from the early music tradition (Gauntlett 1985a, 173).

While trying to trace the roots of rebetiko, Gauntlett suggests, that it could be helpful to look at the tradition of songs touching such themes as hashish consumption and gambling that are usually associated with rebetiko. Gauntlett makes therefore an assumption that rebetiko could appear soon after the formation of the activities that they depict (Gauntlett 1985a, 174). When it comes to the custom of hashish smoking, hitherto no proof has been found indicating that it was a native Greek tradition. Most scholars come to the conclusion that it came to Greece via the Muslim cultures (Gauntlett 1985a, 175). According to Michail Stringaris⁸⁰, “hashish-smoking was introduced to Greece from the Eastern Mediterranean in the first half of the 19th century” most probably via Syros, which was an important port of Greece back then (Gauntlett 1985a, 176-177). As main disseminators of hashish smoking in the 19th century he observes sailors and refugees who practiced this custom mainly in seaports and prisons. As Gauntlett points out, a socioeconomic model similar to that of Syros could also be found in Piraeus⁸¹ of the late 19th century, Smyrna, Constantinople, Thessaloniki and other seaport cities of the Ottoman Empire (Gauntlett 1985a, 178-180). In Constantinople, the hashish consumption was attested by Evliya Çelebi⁸² already in 1634, although the nationality of hashish eaters was not mentioned. In the work of Evliya there is also evidence of songs (without specifying the language) performed by the drunkards of the Galata district of Constantinople, that was one of the centres of Greek low life and where the main dock of Constantinople was located. As Gauntlett points out, “if [the language of songs] were Greek, then these could be the earliest extant ‘Rebetika’” (Gauntlett 1985a, 182), moving the origins of the music genre back to at least as early as the mid-17th century.

To sum it up, one can say that although the word *rebetiko* in the context of a music genre was first documented in 1912, the genre appeared much earlier, namely in the second half of the nineteenth century or even earlier. Even if one uses another definition for this kind of music, it does not change the contents of this music genre. Many songs appeared as records

75 The recording of the song «Από κάτω απ' τις ντομάτες» (English “Under the Tomatoes”) can be found at <https://www.vmrebetiko.gr/item/?id=4097>. Accessed December 20, 2021

76 In 1842 the anthology of songs by Tommaseo was published. See Tommaseo, Niccolò. 1842. *Canti Popolari Toscani, Corsi, Illirici, Greci*. Venice.

77 1860 is the year of publication of the anthology “Τραγούδια ρωμαίικα: Popularia carmina Graeciae recentioris” by Arnold Passow.

78 This date refers to the great conflagration of 1870 in Constantinople, which may be depicted in the song «Μάθατε τι έχει γίνει» (English “Did You Get what has Happened?”) performed by Giorgos Mouflouzelis and recorded in 1971 (Gauntlett 1985a, 173-174) (Gauntlett 1985b, 212).

79 Gauntlett could refer to the Realist prose writers of the 1890s, who depicted koutsavakides of the late 19th century Athens that practiced hashish smoking and used to perform underworld songs (Gauntlett 1985a, 184).

80 Stringaris, Michail. 1937. *Χαοίς. Ψυχοπαθολογική, Κλινική, Κοινωνιολογική Μελέτη επί των Σύννεπειών του Κανναβισμού [Hashish. Psychopathological, Clinical and Sociological Study on the Effects of Cannabis]* Athens.

81 According to Stringaris, hashish smoking was introduced to Piraeus in c. 1880 (Gauntlett 1985a, 178).

82 Evliya Efendi. 1834. *Narrative of Travels in Europe, Asia, and Africa in the Seventeenth Century* (translated from the Turkish by the Ritter Joseph von Hammer). London: Oriental Translation Fund.

without genre affiliation or were defined as demotic, although they had a certain author, but the idea of intellectual property was not developed until the 1930s.

The upper boundary is, as already mentioned, also unclear (Gauntlett 1985a, 24-25). There is a debate on whether rebetiko stopped to exist in the mid-1950s or continued its evolution afterwards. For Manos Hatzidakis, who was one of the first to underline the value of rebetiko in his lecture given in 1949, “rebetiko song vanished along with its cultural context, that is, the culture of underworld people; [...] He viewed contemporary rebetiko [with its increasing commercialisation] as a phase of corruption” (Tragaki 2009, 102). Hatzidakis, like many other rebetologists, defined the songs of the first generation of rebetiko exponents as authentic, in contrast to the contemporary ones. Dinos Christianopoulos on the contrary saw rebetiko as a music genre ‘still under formation’ and observed the rebetiko associated with marginal social groups only as starting point for ‘proper’ rebetiko to come (Tragaki 2009, 102). In the following paragraph, both points of view will be presented in more detail.

The adherents of the point of view that rebetiko came to an end in the 1950s, often align it with the process of transition of *rebetiko* into *laiko*⁸³, which became very popular around that time (Tragaki 2009, 25). One can see a range of reasons that contributed to the decline of rebetiko. A large part of representatives of rebetiko died during the period of German occupation. Strict censorship that had only a short break in 1946, influenced the music production, and rebetiko in particular, heavily (Kounadis 2003b, 300). The repressive political regime that was accompanied by persecutions and exiles caused a new wave of refugees, this time to Germany, Australia, Canada and less to the USA. During Metaxas’ regime with its strict censorship and also during the occupation period with the suspension of record production, rebetika depicting the underworld continued to exist only in rebetiko venues. After World War II, the persecution of the underworld became even harsher (Gauntlett 1985a, 187). Many rebetiko haunts had been closed which also contributed to the decline of rebetiko.

According to Kounadis, the drastic change of the population of Greek cities played an important role in the marginalisation of rebetiko. Although in the 1940s many Athenians died due to famine and other misfortunes of war, the population remained constant at around 1,100,000⁸⁴, due to migration.⁸⁵

When speaking about the possible demise of rebetiko in the 1950s, it is essential to bear in mind that a transition of one genre into another is a long process. Zelepos, when introducing the different phases in the evolution of rebetiko, defines the third and last period that lasted from 1936/1940 to the 1950s as the *laiko* period of rebetiko, if one assumes that rebetiko vanished in the early/mid-1950s (Zelepos 2001, 24-25). With Tsitsanis, who was the most prominent representative of this generation of rebetiko exponents, rebetiko experienced further popularisation and disassociation from the songs of low life. Vasilis Tsitsanis,

83 The term *laiko* refers in this context to the form it took from 1950s to 1980, when *contemporary laiko* appeared.

84 According to the census of Greece of 1940 the population of Athens with suburbs was 1,117,792 and according to the next census held in 1951 the population of the Greek capital with suburbs was 1,376,202 (Kotzamanis 1997, 12).

85 According to Kostis, the estimations of famine deaths in Greece during the Axis occupation (1941-1944) lie around 300,000, but the numbers vary, since many victims were from lower-income groups that did not have the support of a family (Kostis 2018, 301). At the same time Kounadis, while speaking about the reasons for the decline of rebetiko, notes that around 500,000 Athenians died due to famine and other misfortunes of war (Kounadis 2003b, 302). According to Kostis though, 500,000 is the number of victims of war in Greece, and not only in Athens, during the entire period of occupation (the whole population of Greece at that time was 7,500,000, so the number of war casualties was almost 7% of the population (Kostis 2018, 302).

who worked in both rebetiko and laiko music genres, is therefore linked to the transition from rebetiko to laiko (Zeleps 2001, 73) (Tragaki 2009, 61). In general, the boundary between rebetiko and laiko is often unclear: some researchers even classify rebetiko as *laiko*, Greek popular song⁸⁶ (Gauntlett 1983, 80).⁸⁷ Besides Tsitsanis, many other representatives of rebetiko also worked simultaneously in the laiko music genre, and rebetika that are seen today as classical often appeared as recordings under the name of laika.

Rebetiko made its way from the music of marginal groups to a music that touched a large part of the Greek society. In this change some researchers see its transition to laiko. There are several reasons for this change of social context. The dictatorship of Metaxas and strict censorship influenced the song lyrics, and songs with underworld connotation were banned by the censor (Gauntlett 1985a, 101-102). Most rebetika of that period had an erotic theme (Gauntlett 1985a, 164-165). As a result of the war, themes depicted in rebetika became more familiar to larger parts of the population. Furthermore, the topics of post-war rebetiko became more diverse (Gauntlett 1985a, 165), resulting in a rise in popularity also among the upper classes of society. Rebetiko exponents were reacting to the changing social profile of the audience, such that soon after the war a new style of rebetiko, the so-called archondorebetiko ('posh/(noble) rebetiko')⁸⁸ appeared (Gauntlett 1985a, 129-130). This new style of rebetiko is considered to be a confluence of *elafro tragoudi* (Greek light music) and rebetiko (Tragaki 2009, 66). Due to the rapid change of music trends in post-war Greece, archondorebetiko⁸⁹ reached its demise very soon, namely in the beginning of the 1950s (Gauntlett 1985a, 130,148), but it "increased the popularity of rebetiko style of entertainment amongst wealthy urbanites" (Tragaki 2009, 66).

The debates on rebetiko and its value as part of Greek culture, which started in the late 1940s with the article of Anoyianakis and the lecture of Hatzidakis, only intensified in the 1960s, when it was still questioned whether it was a "'proper' Greek folk music" (Tragaki 2009, 96-100). These debates increased the interest in the works of the first generation of rebetiko exponents as well as in rebetiko songs that were strongly connected to the tradition of demotic song, as they were considered to be more 'authentic' or 'pure' than the contemporary rebetika. Many old rebetika have been rerecorded either in their original or in reinterpreted form. Finally, these debates resulted in the so-called revival of rebetiko in the late 1960s⁹⁰ and especially in the 1970s (Tragaki 2009, 93-113), although as Tragaki mentions,

86 Remarkably, many records of songs, which are nowadays considered to be rebetika, include on their labels such English terms as 'folk song' or 'popular song' although in Greek the genre affiliation was missing.

87 Other researchers categorise rebetiko as urban folk song (*αστικό δημοτικό τραγούδι*). This debate is also complicated by the fact that there is no uniform usage of the terms *dimotiko tragoudi* and *laiko tragoudi*, to a such extent that sometimes they are seen as synonyms. Gauntlett, portraying the point of view of Anoyianakis, who classified rebetiko as *laiko*, claims that *dimotiko tragoudi* can be seen as a 'folk song' transmitted in oral tradition, and *laiko* as a "more recent (post-1821) tradition of popular song whose form is fixed" (Gauntlett 1983, 80,91). According to another point of view, *laiko* is 'urban song', whilst *dimotiko*, or demotic song, is 'rural song' (Gauntlett 1983, 80,91).

88 The protagonist of archondorebetiko leads a bohemian life and is a frequenter of luxurious revelries (Gauntlett 1985a, 129-130) (Tragaki 2009, 66).

89 The songs «Κολωνάκι - Τζιτζιφιές» ("Kolonaki - Tzitzifities") composed by Markos Vamvakaris and sung by Vamvakaris himself and Soula Kalfopoulou (disc number: Odeon GA-7490, recorded in 1949) and «Αρχοντορεμπέτισσα» ("Archondorebetissa") composed by Giorgos Mitsakis and sung by Mitsakis himself, Rena Dalia and Yiannis Tatasopoulos (disc number: HMV AO-5008, recorded in 1951) are examples of archondorebetiko.

90 Vamvakaris, like many other representatives of the first generation of rebetiko exponents, started in the 1960s his 'second career', after the oblivion of the 1950s, when he had to tour to remote provinces to get some work as a rebetiko musician (Gauntlett 1985a, 150). The reason for the demise of popularity of Vamvakaris and other exponents of his generation in the 1950s was also the fact, that in circa 1952 Manolis Chiotis added a fourth

the term ‘revival’ is problematic because “it presupposes that rebetiko was at some point ‘dead’” (Tragaki 2009, 124). Going to bouzoukia nightclubs and other rebetiko venues became a popular type of entertainment (Tragaki 2009, 113). As Giannis Papaioannou points out, the expression *πάμε στα μπουζούκια*⁹¹ (English ‘let’s go to bouzoukia’) became a sign of freedom (Gauntlett 1985a, 131). Rebetiko music and music instruments that were associated with it, namely bouzouki and baglamas, were perceived as a protest against the Junta regime (Tragaki 2009, 113-116).

Starting in the early 1980s, researchers speak about the second revival of rebetiko that lasts until the present day (Tragaki 2009, 113-116). According to Tragaki, one distinguishes it from the first revival due to the change of the ways rebetiko music was observed and practiced (Tragaki 2009, 124). In 1982, the Australian documentary film *Rembetika: The Blues of Greece* narrated by Anthony Quinn was released. In the following year, the film *Rembetiko*, directed by Costas Ferris, was released. It is in this period, that a range of biographies and autobiographies of important figures of rebetiko were published.

It is remarkable that a considerable amount of rebetika, which were originally composed between the wars and afterwards during the civil war, were recorded only in the 1970s after the fall of the Junta. Also due to the dictatorship of Metaxas and the Junta, a lot of songs of that period remained unpublished. Thus, the issue concerning the upper boundary in the 1950s seems to be questionable. For example, Vamvakaris has never managed to record his song «Χαϊδάρι» (“Haidari”)⁹² that he composed, but had never written down during the occupation period (Kounadis 2003a, 32). According to Stelios Vamvakaris, son of Markos Vamvakaris, who also devoted his life to rebetiko, the lyrics of this song survived, but not the music.⁹³ In 1980 Giorgos Dalaras released an album *Ρεμπέτικα της κατοχής* (English *Rebetika of the Occupation*), which included the song “Haidari”. The music to this song was written by Stelios Vamvakaris. One can find a range of such examples, so in my mind the question on whether rebetiko vanished to exist in the 1950s or in the 1970s when several prominent exponents of rebetiko died, or continues its existence on the stages of today’s *stekia*, taverns, *koutoukia*, *rebetadika* and other rebetiko venues depends on the observer’s approach and point of view on the history of rebetiko.

2.3 PERIODISATION OF REBETIKO

Before starting to define a periodisation of rebetiko, it is important to mention that such an analysis as well as setting the boundaries of the existence of a particular music genre are to a certain extent artificial. Tragaki makes a very important observation, namely that rebetiko was for a long time regarded as a particular musical style⁹⁴, so that all the changes that

pair of strings to the bouzouki. Such bouzoukia soon became popular, and music ensembles employed only the players of a four-course bouzouki (tetrachordo bouzouki), and not the formerly more popular three-course bouzouki (trichordo bouzouki) (Tragaki 2009, 72,112).

91 See the song «Πάμε στα μπουζούκια» (English “Let’s Go to Bouzoukia”) composed by Nikos Gounaris and sung by Marika Ninou (disc number: Liberty Αμερικής L 202 A, recorded in 1955).

92 The name of the song refers to the Haidari concentration camp operated by the German SS from September 1943 to September 1944 in the suburbs of Athens.

93 According to Stelios Vamvakaris’ words at Vamvakaris, Stelios. «Τα στερνά τιμούν τα πρώτα...». 04 December 2012, Θέατρο Αυλαία, Thessaloniki.

94 Here Tragaki refers the reader to the etymology of the word ‘style’, that comes from the Ancient Greek *στυλος* (English ‘column, pillar’), literally meaning therefore “something consolidated and static” (cited as in (Tragaki 2009, 68)).

occurred over time were seen as alien in the context of the so-called ‘original’ rebetiko (Tragaki 2009, 48). According to Tragaki, it is for this reason, that the periodisation of rebetiko was seen as follows: “The transformations of rebetiko music were [...] examined in the same way as the evolution of biological organisms: the resulting schema involves successive eras described as the ‘genesis’, ‘culmination’, ‘decay’ and ‘demise’ of rebetiko, perceived as concrete and non-interactive stages of a predetermined course of tradition” (cited as in (Tragaki 2009, 48)).

At the same time, as Tragaki also mentions, the introduction of a periodisation can facilitate the process of understanding the history of rebetiko (Tragaki 2009, 48).

Attempting a periodisation of rebetiko is a difficult task. As already mentioned, the roots of rebetika are often seen in the second half of the nineteenth century and its demise in the 1950s. At the same time it is clear that in such an area like music, the formation of a new music genre takes a long period of time and this new genre is based on elements of the old ones, so that it is very often difficult to define not only chronological boundaries but also clear boundaries between different music genres, as they do not exist in a vacuum and the idea of a ‘clear’ music genre is also quite naive.

As mentioned above, there are multiple views on the periodisation of rebetiko, and a number of rebetologists, like Elias Petropoulos, Panagiotis Kounadis, Stathis Gauntlett and Ioannis Zelepos have discussed this issue and presented their attempts on a periodisation.

Kounadis defines three periods of rebetiko, namely, the first one lasting until 1922, the second one taking place between 1922 and 1936/1940, and the third one between 1945 and circa 1955 (Kounadis 2003a, 252).

According to him, the first period started in the 1850s and lasted until 1922, so it is bounded by the Asia Minor Disaster and the following mandatory exchange of population between Greece and Turkey. This period is characterised by rebetika of anonymous authorship that had a strong connection to demotic song and were usually products of oral transmission (Kounadis 2003a, 36). For this period, Kounadis defines two types of rebetika: those that were performed in Greece and those performed in Asia Minor. The main differences between these two types of rebetika were their social context and the instrumentation. Rebetika performed in Greece during that period were orchestrated by tambouras, bouzouki and baglamas, and their typical venues were closed communities of hashish dens (‘tekedes’) and prisons.⁹⁵ Primary music instruments of rebetika performed in Asia Minor as well as in some other areas such as Egypt were violin⁹⁶, oud, qanun (kanonaki), santour and guitar. According to Kounadis, the typical venues of this type of rebetika were the streets of low-life districts as well as laika (‘folk’) centres (Kounadis 2003a, 286). At the same time, despite the performance among the economically deprived social strata, rebetika of this period were also part of the repertoire of the café aman ensembles where they were removed from their “low-life context [...] [and] were performed on purpose, to satisfy the customers’ musical preferences” (Tragaki 2009, 49-50). Among the rebetiko exponents

⁹⁵ The songs of that type of rebetiko did not draw attention of foreign record companies due to its connection to criminals and other individuals of the underworld. Moreover, the venues in which they were performed, were not open to foreigners. The first song accompanied by bouzouki was recorded in Görlitz in 1917, and although it attests this music instrument, the song belongs to the Asia Minor style of rebetiko. The first known recording of rebetiko in Greece that was accompanied by bouzouki, was made in 1932.

⁹⁶ According to Kounadis, in Constantinople instead of a violin a politiki lyra, or Constantinopolitan lyra, was popular (Kounadis 2003a, 286).

of that time one can highlight Giorgos Katsaros⁹⁷, Kyriaki Antonopoulou known as Mme. Coula⁹⁸, Marika Papangika⁹⁹ and Giorgos Tsanakas. Most of the rebetiko records of the period before 1922 were produced in New York, Smyrna and Constantinople. Furthermore, many records of that time were made by professional music ensembles called *estoudiantines*. Taking into account the fact that almost half a million immigrants came to the USA between 1890 and the early 1920s, including many prominent rebetiko exponents, America can be seen as the third center of rebetiko together with Asia Minor and Greece, as there rebetiko was performed in a different social context and actively absorbed the elements from other music genres, that were not only Greek. The majority of Greeks, about 90%, that arrived in the USA between circa 1890 and early 1920s were male, and did not plan to stay in America for a long time.¹⁰⁰ The Music, which they practiced and experienced in *kafeneia* (coffee shops), for them was a way to forget the harsh realities of life. The rebetiko exponents in the USA worked in several different music genres maybe even more often than those in Greece or the Ottoman Empire, making the definition of genre borders even more difficult.

The second period, according to Kounadis, takes place between 1922 and 1936/1941. It therefore includes the heyday of both styles of rebetiko: namely, Smyrna style, or café aman style, that prevailed till circa 1933, and the Piraeus style of rebetiko that prevailed in the 1930s. According to Kounadis, these two styles coexisted during the period between 1922 and 1936/1941 with the Piraeus style dominating from 1932 onwards (Kounadis 2003a, 287). Furthermore, although the café aman tradition with its characteristic music repertoire emerged in Greece long before 1922¹⁰¹, the arrival of refugees from the Asia Minor increased the popularity of the Smyrna style of rebetiko drastically. The upper boundary of the second period varies between 1936, the year when the dictatorship of Metaxas was established, and 1941, the year when Greece was occupied by Germany and the record production was stopped.

The third period takes place between 1936/1941 and about 1955 (Kounadis 2003a, 252, 274). At that time many representatives of the Smyrna style of rebetiko had already passed away¹⁰², and the strict censorship of Metaxas motivated the appearance of the new generation of rebetiko exponents including Vasilis Tsitsanis, Manolis Chiotis, Giannis Papaioannou and Giorgos Mitsakis. It is during this period that a four-course bouzouki replaced a three-course bouzouki, and instead of dromoi, or makamia, the preference was given to the Western well-tempered system (Kounadis 2003a, 286). Strict censors of the Metaxas regime did not favour the oriental influences on rebetiko, so that the position of the café aman songs as well as of its exponents became even more difficult compared to the beginning of the 1930s, when the Piraeus style started to dominate rebetiko. With the establishment of

97 Giorgos Katsaros, who also started his career in the United States, performed not only rebetika, but also demotic and elafra songs as well as songs of the revue theatre ('epitheorisiaka tragoudia'). He toured the world as a musician visiting Australia, Latin America, India and Northern Africa. In 1927 Katsaros also visited Greece and made a number of records together with the exponents of the Asia Minor, or Smyrnic, style of rebetiko. (Kounadis 2003a, 223).

98 In 1916 and 1917 Mme. Coula, Greek *Κα. Κούλα*, who performed not only rebetika, recorded 34 songs in Greek and Turkish in New York (see Bucuvalas, Tina/Frangos Stavros K. Overview of Greek Music in America. 1880-1918. In (Bucuvalas 2018)).

99 In 1925 Marika Panangika and her husband Gus (Kostas) Papangikas opened a nightclub in New York on West Thirty-Fourth Street near Eighth Avenue. In this club she used to sing not only in Greek, but also in Armenian (see Bucuvalas, Tina/Frangos Stavros K. Overview of Greek Music in America. 1918-1933. In (Bucuvalas 2018)).

100 See Bucuvalas, Tina/Frangos Stavros K. Overview of Greek Music in America. 1880-1918. In (Bucuvalas 2018)).

101 The first café aman was opened in Athens in 1873.

102 Among them are Marika Papangika, Andonis Dalgas and Vangelis Papazoglou.

the dictatorship and the following implementation of the censorship, amanedes, an inalienable part of the café aman repertoire, were banned.¹⁰³

Below, some fragments of newspaper articles on amanes published under Metaxas' regime can be found.

“Είναι το πιο ασύδοτο έργο που βγήκε ποτέ από ομάδα ανθρώπων. [...] Ο διωγμός του αμανέ είναι μόνο μια λεπτομέρεια. Ο πόλεμος πρέπει να είναι γενικότερος. Κατά του δίσκου!”¹⁰⁴

“[Amanes] is the most unaccountable work ever made by a group of people. [...] The expulsion of amanes is almost nothing. The war must be more general. Against the record production!”

“Ήταν ντροπή στον αιώνα του αεροπλάνου και του ραδιοφώνου η ύπαρξις του αμανέ. Επιτέλωσ είμαστε ένας λαός με εξέλιξη και πολιτισμό. Δεν μπορούμε να ενθυμούμεθα κάθε τόσο, ότι υπήρξαμε επί μίαν σειράν σκλάβοι.”¹⁰⁵

“It was a shame in the age of the airplane and the radio for amanes to exist. Finally, we are a developed folk with rich culture. We can not remember every now and then that we were for some time slaves.”

The censors attacked not only the rebetiko songs of Smyrna style and amanedes for their oriental elements, but also all rebetika referring to drug consumption as well as other elements of the underworld context.¹⁰⁶ The exponents of rebetiko had to practice self-censorship before submitting their songs for approval.¹⁰⁷ In an interview, Markos Vamvakaris while talking about the law against drugs that was introduced in 1920, but actually came into force only in 1936, even calls it the ‘law against bouzoukia’. As he points out, the bouzouki players were persecuted by the police and had to play in secret (Kounadis 2003a, 29). Some exponents of rebetiko were imprisoned or sent to exile: for example, Anestis Delias and Michalis Genitsaris were sent to exile on the island of Ios. The influence of censorship on the topics and the language of rebetiko was large, and most of the recorded songs of the period of Metaxas' dictatorship were erotic (Gauntlett 1985a, 106, 164-165). All these factors, as already mentioned, motivated the appearance of the new generation of rebetiko exponents who stimulated the process of europeanisation of rebetiko (Gauntlett 1985a, 110).

Alternatively, Kounadis proposes to set the year 1941 as the beginning of the third period, but with the Axis Occupation in 1941 the record production in Greece was suspended.

103 Remarkably, two years before the implementation of censorship in Greece, in 1934 the Kemalist government banned amanedes and Ottoman music practice in general in Turkey (Pennanen 2004, 13) (Liavas 2009, 238-240).

104 Papantoniou, Zacharias. «Ο αμανές εν διωγμώ!..» [*Amanes in expulsion!..*]. Ελεύθερον Βήμα. 03/07/1938. Athens. Cited as in (Liavas 2009, 238, 240)

105 Marakis, Nikos. «Ο αμανές» [*Amanes*]. Σημεία. 30/11/1937. Piraeus. Cited as in (Liavas 2009, 248)

106 One of the first songs that were banned is the song «Βαρβάρα» (“Varvara”) with double entendre composed by Panagiotis Toundas and sung by Stellakis Perpiniadis (disc number: Columbia DG 6159). It was first recorded in February 1936, shortly before the 4th of August Regime was established. Soon after the introduction of censorship in 1936, the song that had become a bestseller was banned. One of the reasons for the ban is considered to be the rumours that by the character of Varvara the daughter of the dictator, Loukia, was represented. The song was so popular, that it was recorded in New York in November of the same year (disc number: Orthophonic ORS-359, singer: Tetos Dimitriadis). More information about the history of this song can be found on the website of Panos Savvopoulos, at http://www.panossavvopoulos.gr/p/blog-page_96.html (accessed December 23, 2021).

107 Vamvakaris, for example, speaking about the horror of self-censorship, admits that he had to change his way of writing songs (Liavas 2009, 249).

Although the occupation ended in 1944, the actual record production started again only in 1946, so one could place the third period of rebetiko between 1946 and 1955. At the same time, this exclusion of the years of German occupation of Greece may be questionable, because although there was no record production, representatives of rebetiko continued to write songs and the genre therefore continued its evolution.

Ioannis Zelepos also sees the origins of rebetiko in the 19th century, but he divides the time span between 1922 and 1936/1940 into two separate periods: into a period that was dominated by the Smyrna, or café aman style of rebetiko, and a period that was dominated by the Piraeus school of rebetiko. The first one lasted from 1922, shortly before almost a million refugees arrived in Greece¹⁰⁸, until 1933. The reason for the rising popularity of the Piraeus style may lie in the appearance of recordings of songs accompanied by bouzouki, which along with baglamas belongs to the primary instruments of the Piraeus style of rebetiko. Although the first rebetiko accompanied by bouzouki was recorded in Görlitz in 1917¹⁰⁹, this recording remained unknown for a long time and was found only in the twentieth century in the German archives¹¹⁰ (see (Alexatos 2018)). In America, the first known recording of a Greek song accompanied by bouzouki was made in 1926: it was the song «Τρεις λυγαριές» (“Three Chaste Trees”).¹¹¹ The American record market had a strong influence on the Greek record industry since the invention of the phonograph in 1877 and the first Greek song being recorded in the USA in 1896. As Gail Holst-Warhaft pointed out in her book on rebetiko: “The history of rebetika, both as recorded and live genre, has always been closely linked to the émigré communities in the United States”.¹¹² Although the first known recording of a Greek song accompanied by bouzouki was made in the USA already in 1926 and the first known recording of rebetiko accompanied by bouzouki was made in Greece in 1931¹¹³, the American recordings with bouzouki that followed in 1932 and 1933 had a much larger impact on the Greek recording market. Among the records that had a significant influence on the change of rebetiko music tradition in Greece are instrumental compositions that appeared on two discs recorded in New York: «Το μυστήριο» (“Mystery”) and «Μινόρε του τεκέ» (“Minor of Tekes”) recorded in 1932 (disc number: Columbia USA 56294-F) as well as «Ραστ του τεκέ» (“Rast of Teke”) and «Μουρμούρικο» (“Mourmouriko”) recorded in 1933 (disc number: Columbia USA 56327-F).¹¹⁴ The formerly stigmatised bouzouki suddenly became an instrument in high demand by the record companies as something that can please the audience due to the great success of the American

108 Many refugees arrived to Greece already before the mandatory exchange of population in 1923.

109 The song «Χήρα ν' αλλάξεις τ' όνομα» (“Widow, Change Your Name”) sung by Apostolos Diamantis (disc number: Royal Prussian Phonographic Commission PK-1004).

110 Although this recording of «Χήρα ν' αλλάξεις τ' όνομα» (“Widow, Change Your Name”) became known only recently, there exist other recordings of this song. For example, it was recorded in 1928 under the name «Τζιβαέρι μανές (Χήρα να αλλάξεις όνομα)» (“Tzivaeri Manes (Widow, Change Your Name)”) sung by Kostas Karipis (disc number: Polydor V-50208).

111 The song «Τρεις λυγαριές» (“Three Chaste Trees”) is accompanied by bouzouki (Kostas Nokotis) and accordion (Giannis Sfondilias) (Maniatis 2001).

112 Cited as in Bucuvalas, Tina/Frangos Stavros K. Overview of Greek Music in America. In (Bucuvalas 2018)

113 The song «Τα δίστιχα του μάγκα» (“The Couplets of Mangas”) of unknown authorship and sung by Giannis Spachanis (disc number: Columbia DG-147) is considered to be the first known recording of a song accompanied by bouzouki made in Greece, namely in Athens.

114 More information about the influence of American record production on the Greek market as well as the first bouzouki recordings in the USA and Greece can be found at Bucuvalas, Tina/Frangos Stavros K. Overview of Greek Music in America. 1918-1933. In (Bucuvalas 2018) as well as at <https://www.vmrebetiko.gr/item/?id=4141> (Accessed January 4, 2022).

recordings with bouzouki in Greece. The absence of bouzouki in earlier rebetiko recordings at the same time does not mean it was not present in live rebetiko performances. Until the beginning of the 1930s, the presence of bouzouki and baglamas were mostly limited to the close communities of hashish dens and prisons.¹¹⁵ The appearance of the first records accompanied by bouzouki in Greece marked the beginning of the rapidly growing popularity of the Piraeus style of rebetiko, which soon prevailed over the Smyrna style. The *Legendary Quartet of Piraeus*, founded in 1934 and headed by Markos Vamvakaris, soon became very popular and had a large impact on the further evolution of rebetiko and the growing popularity of the Piraeus style of rebetiko. The third period in the categorisation of Zelepos conforms with the proposal by Kounadis, namely between 1936/1941 and the 1950s. This period of rebetiko is usually tied with the transition to *laiko* and sometimes even defined as the *laiko* period of rebetiko. As Zelepos points out, rebetika of this period lost their strong connection to the underworld and broadened their range of topics (Zelepos 2001, 71-72). As already mentioned, in this period one observes a westernisation of rebetiko that was caused not only by the persecution of oriental music elements under the dictatorship of Metaxas, but also by the general change of tastes. Dromoi, or makamia, were replaced by the Western well-tempered system. In this period, one often observes an appearance of the second voice. While in the Piraeus style of rebetiko of the 1930s, the female artists were less present than in rebetika of the Smyrna style, the number of female representatives of the music genre grew again after the war (Zelepos 2001, 80). Furthermore, in this period a distribution of roles within the orchestra of rebetiko performers appears, and the figure of a music star in the context of rebetiko evolves. While earlier rebetiko exponents carried out several tasks simultaneously, from the end of the 1940s onwards the rebetiko artists were mostly responsible for one particular task (Zelepos 2001, 92-93). Markos Vamvakaris, for example, was often a composer and a songwriter who during the performance often sang and played bouzouki simultaneously. Stelios Kiromitis, who was also a member of the quartet of Piraeus, often had the roles of a composer, songwriter, instrument player or even orchestra conductor simultaneously (Kounadis 2003a, 49). It is also notable that during the *laiko* period of rebetiko, to be more precise from 1946 onwards, the Greek lyricists started to demand their work to be acknowledged on the labels of the records, whereas until 1946 the author of a rebetiko song text often remained unclear (Gauntlett 1985a, 163). The main venues of rebetiko of that period were no more hashish dens and cafés amans, but taverns and entertainment centres, or *kosmika kentra*. As already mentioned, due to the suspension of record production in Greece from the beginning of the Axis Occupation in 1941 until 1946, when the record production was resumed, many rebetiko songs written during the war were recorded only after 1946 or even 1974, when the Junta fell. Due to the absence of recordings for the period between 1941 and 1946, the main source on rebetiko of that period are personal interviews and (auto)biographies of the rebetiko exponents. One of the most well-known rebetika of the occupation period is the song «Θα σαλπάρω» (English “I’ll Leap”) that is considered to be written by Michalis Genitsaris in 1942 (Genitsaris 1992, 56). Although Genitsaris claimed to have composed it during the Occupation, the first known recording of this song with his participation appeared only in 1976 (disc number: Polydor

115 One of the first references of the songs accompanied by baglamas was made by the ethnographer and journalist Kostas Faltaitis in the article “How sing imprisoned, thieves and hashish smokers” published in the magazine *Mpouketo* in February 1929. In his article, Faltaitis stressed the importance of further research of these songs and listed a number of couplets of the songs of baglamas (Liavas 2009, 245-247).

2421077), with three recordings that appeared earlier. Its first recording under the name «Οι σαλταδόροι» (English “The Leapers”) was performed by Giorgos Katsaros in the USA in 1947 (disc number: Standard F-9025) and the second one under the name «Θα με κάνουν να σαλτάρω» (English “They’ll Make me Leap”) was performed by Spyros Zagoraios and Zoi Zagoraiou in Athens in 1969 (disc number: Sonata 120-B). Only the recording of 1974 as well as its later recordings contain the lyrics that are known today in Greece and give a heroic representation of a group of lads who jumped on the German trucks to steal the goods that were later sold on the black market (Zelepos 2001, 76-78) (Gauntlett 1985a, 119-120). The appearance of such a song in the repertoire of Genitsaris is to be expected: as he points out himself, he and his colleagues had to cater to the tastes of the audience of centres where they performed, and the mentioned *saltadoroi* were the frequenters of rebetiko venues during the war (Genitsaris 1992, 56) (Gauntlett 1985a, 117). After the Second World War the audience of rebetiko had changed drastically. While earlier rebetiko was associated with marginal social groups, after the war and the following Greek Civil War the themes raised in rebetika referred to much broader social classes. The number of representatives of rebetiko that did not have any or only a cursory connection to the underworld also grew: for example, Vasilis Tsitsanis, who is often seen as the main representative of the *laiko* period of rebetiko and who also preferred to use the term *laiko* instead of *rebetiko* in reference to his works, received musical classes and had only a weak connection to the underworld. From the music of the underworld rebetiko evolved into a kind of music that was close and understandable to a large part of the population of Greece. Zelepos also mentions that some rebetologists, who connect the 1950s with the demise of rebetiko, isolate one more period, namely a period of archondorebetiko, or noble rebetiko, that lasted until circa 1960 (Zelepos 2001, 25).

Some scholars see rebetiko as a popular music genre of the interwar period, to be more precise the music that existed between 1922 and 1955 (Anagnostou 2018, 2). One of the reasons for such boundaries is that the research on rebetiko is to a large extent based on the studio records of rebetiko and not on its pervasive culture in a specific social context. Although the first recordings that were labeled as rebetika appeared already in the second decade of the twentieth century, most rebetika that are considered to be classical nowadays were recorded during the interwar period. Actually, when placing the era of rebetiko between 1922 and 1955, most researchers bear in mind the heyday of rebetiko and do not define strict boundaries of the music tradition in these thirty years. Similarly to Ioannis Zelepos and Panagiotis Kounadis, Elias Petropoulos defines three periods of the heyday of rebetiko: namely, the Smyrna period from 1922 to 1932, the Piraeus period from 1932 to 1940 and the *laiko* period from 1940 until 1952 (Petropoulos 1991, 16-17). Moreover, Petropoulos was among the rebetologists who describe the demise of rebetiko in the 1950s, although a number of songs in his anthology *Rebetika Tragoudia* was composed afterwards (Gauntlett 1985a, 9).

In his doctoral thesis, Stathis Gauntlett proposed a more precise description of the evolution of rebetiko during seven different stages: non-commercial rebetika of oral tradition performed in hashish dens, gambling dives and prisons; first ‘commercial’ rebetika that were performed in cafés amans; commercial recordings and personal compositions before 1936; rebetika of the period between 1936 and April 1941; rebetika of the period of suspension of record production in Greece between 1941 and 1946; rebetika composed after 1946; revival of rebetika in the 1960s-1970s (Gauntlett 1985a, 52-53). What is specific about the

periodisation suggested by Gauntlett is that it proposes to study different styles of rebetiko throughout the whole history of rebetiko and not only of a specific period. Thus, rebetika of the café aman style are observed not only in the context of rebetika that were performed until 1933, when the Piraeus style of rebetiko started to dominate over the Smyrna style, but from the appearance of rebetika in the repertoire of cafés amans at the end of the nineteenth century up to the period after the World War II. Although the Smyrna style of rebetiko, starting from the mid-1930s, did not return to its former popularity and merged to a certain extent with the prevailing Piraeus style of rebetiko, it continued its existence in the works of rebetiko exponents from Asia Minor such as Giorgos Rovertakis, Roza Eskenazi, Rita Abatzi, and Vangelis and Angeliki Papazoglou. When it comes to the oral tradition, it is usually connected to the origins of rebetiko going back to the nineteenth century. At the same time, as Gauntlett points out, it continued to exist also afterwards, especially on the islands, where many representatives of the underworld were sent to into exile (Gauntlett 1985a, 116).

In his doctoral thesis, submitted in 1978, Gauntlett describes a revival of rebetiko in the 1960s and 1970s as a period when many old rebetika were rerecorded (Gauntlett 1985a, 154-155). Due to the censorship, some of these rebetika were rerecorded in an altered fashion: for instance, the song «Το χαρέμι στο χαμάμ» (English “The Harem in the Hamam”) composed by Anestis Delias and reinterpreted by Stratos Pagioumtzis appeared under the name «Μες της πόλης το χαμάμ», English “In the City’s Hamam”.¹¹⁶ The first recording in question appeared in 1935 (disc number: Columbia DG 6165) and the second one in 1961 (disc number: Phillips 7739). In the version of 1961 the couplets referring to the smoking of hashish were replaced by neutral ones. During the period of the Junta, rebetiko was seen as a protest against the regime: as Tragaki points out, playing a bouzouki was opposed to a classical musical education at a conservatoire (Tragaki 2009, 114, 116).¹¹⁷ One of the rebetiko exponents who was active in the period of first revival and reinterpreted many old rebetika is Giorgos Mouflouzelis. During this period, a number of (auto)biographies of rebetiko representatives was published: for example, those of Giorgos Rovertakis (1973), Kostas Roukounas (1974), and Markos Vamvakaris (1978). First sizeable anthologies of rebetiko songs also appear during this period: for instance, the *Rebetika Tragoudia* by Elias Petropoulos (1968) and *Rebetiki Anthologia* by Tasos Schorelis (1977-1981). Petropoulos, who ignored the censorship regulation, was even sentenced to five months in prison due to the supposedly pornographic illustrations in his book (see (Petropoulos 1991, 63)), and his anthology was banned (Tragaki 2009, 109-110). In this period, first foreign works on rebetiko appear, among them *Road to Rembetika* by Gail Holst (1975) and *Rebetika. Songs from the Old Greek Underworld* by Katharine Butterworth and Sara Schneider (1975).

As already mentioned in the previous section, starting in the early 1980s researchers speak about the second rebetiko revival that lasts until today. What distinguishes the second revival from the first one is among others the fact that the rebetiko tradition of that period does not have anything to do with censorship. A range of rebetika that were created during the period of the censorship, which lasted almost forty years until circa 1974, or

116 The title of the song probably refers to the city of Constantinople, as it is often can be found in Greek sources in its shortened version Πόλη (English ‘City’, capitalised) instead of Κωνσταντινούπολη (English *Constantinople*). In this case it is difficult to determine, whether it should be capitalised due to the fact that on the label of the record the song title is written in capital letters.

117 The Athens Conservatoire started to teach the bouzouki as late as 2014 (see <https://polykandriotis.gr>, accessed January 7, 2022.)

during the record suspension in the Occupation period, were first recorded in their original form or due to the lack of sheet music and written down lyrics in a form close to the way they were performed in rebetiko venues of the time of censorship. For instance, the album of Giorgos Dalaras *Ρεμπέτικα της κατοχής* (English *Rebetika of the Occupation*), released in 1980, contains some early, previously unrecorded songs, including «Επιδρομή στον Πειραιά» (English “Raid on Piraeus”) composed by Genitsaris and «Χαϊδάρι» (English “Chaidari”) composed by Vamvakaris. The Greek Socialist party PASOK, founded in 1974, decided to support rebetiko, which may be seen as a political strategy to form a closer connection to the people (Tragaki 2009, 131): in 1981, PASOK founder and newly elected prime-minister Andreas Papandreou was depicted dancing zeibekiko to the song of Tsitsanis in a tavern (Zelevos 2001, 178). In 1983 the film *Rembetiko*, directed by Costas Ferris, was released, which soon became very popular. During this period, a range of further (auto)biographies of rebetiko exponents were published: for example those of Roza Eskenazi (1982), Michalis Genitsaris (1992) and Giorgos Mitsakis (1995). During the second rebetiko revival also museums such as the Museum of Vamvakaris on the island of Syros and the Museum of Tsitsanis in Trikala were opened. With the growing importance of the internet, rebetiko also invaded this medium. For instance, in 1996 the website *Ρεμπέτικο Φόρουμ* (English *Rebetiko forum*)¹¹⁸ was launched, and it remains a lively platform for exchange of information and communication about rebetiko to this day, where one can find information about probably all aspects of rebetiko including bouzouki lessons, musical theory, discussions on the history of rebetiko and neighbouring music genres as well as the promotion of rebetiko venues and scheduled live meetings of the forum. Another digital source on rebetiko, including a large collection of rebetiko songs, is the website *rebetiko.sealabs.net*.¹¹⁹ In 2019 the website of the *Virtual Kounadis Archive*¹²⁰ was released. At the moment, it contains 2,820 digitised 78 rpm records (most of them are rebetika), sheet music, postcards, photographs and many other data related to rebetiko as well as some other Greek music genres. Among the rebetiko exponents of the second revival who were or are still active, one should mention Stelios Vamvakaris, Giorgos Dalaras, Agathonas Iakovidis, Giannis Lebesis and Thessalonikian female singer Mario (Maria Konstantinidou). Many rebetiko songs were reinterpreted by such popular Greek musicians as Glykeria, Haris Alexiou and Anna Vissi. There exist a range of *kompanies*, or bands, that perform rebetiko: for instance, Trio Tekke, Megla, *Ρεμπετιέν* (Rebetien), *Γιαβάς* (Giavas) and *Νίκελ* (Nikel). Today, rebetiko is performed in such places as *stekia* frequented by the young audience, *koutoukia* (small taverns usually with a cellar), taverns, and more glamorous nightclubs called *rebetadika* and *ellinadika* (Tragaki 2009, 134-135).¹²¹

So, although there are different views on the periodisation of rebetiko, the tendencies are similar. Some periods are marked by historical events like, for example, the Asia Minor Disaster of 1922¹²² and the establishment of the dictatorship of Metaxas in 1936, others by

118 The website of *Ρεμπέτικο Φόρουμ* can be found at <https://rebetiko.gr> (accessed January 8, 2022). According to Tragaki, the establishment of the forum was organised by Costas Ferris (Tragaki 2009, 128).

119 Available at <https://rebetiko.sealabs.net> (Accessed January 8, 2022)

120 The website of the *Virtual Kounadis Archive* can be found at <https://www.vmrebetiko.gr/> (Accessed January 8, 2022).

121 As an example of rebetiko venues active today in Greece one can mention the following ones: a tavern *Εφήμερον* (Efimeron), a tavern *Μεταξού* (Metaxou), a coffee shop *Τυφλόμυγα* (Tyflomyga) and a rebetadiko *Αργώ* (Argo) in Athens; *Φαγοποιεϊόν στον Θωμά* (a small plates restaurant *Stou Thoma*) in Piraeus; a tavern *Χατζή Μπαχτσέ* (Chatzi Bachtse), a coffee shop *Το Ανατολικόν* (To Anatolikon) and a night club *Πριγκηπέσσα* (Pringipessa) in Thessaloniki.

122 As the Asia Minor Disaster in Greek historiography are defined the aftermaths of the Greco-Turkish war 1919-1922 that ended with the compulsory exchange of population between Greece and Turkey in 1923 based not

the evolution of the music genre itself (for example, the transition from the Smyrna style to the prevailing Piraeus style in rebetiko in circa 1933). The disagreements concerning periodisation occur, as already mentioned, both with regard to the origins as well as the upper boundary of rebetiko. It seems more plausible to set the lower boundary to approximately the 1850s, when the tradition of rebetiko started to develop both in Asia Minor and in mainland Greece. Even though the refugees from Asia Minor almost doubled the population of Athens, Piraeus and Thessaloniki and therefore influenced to a large extent the music and everyday life of large Greek harbour cities, the development of rebetiko in Greece goes back to the 1850s (Kounadis 2003a, 392). The upper boundary is often set around the mid-1950s, marking the transition to the laiko genre. At the same time there are scholars and musicians who hold the view that rebetiko survived until today. Despite the fact of a certain marginalisation of the genre in the 1950s and growing popularity of other music genres (for example, laiko, including its period of indomania in the 1950s-1960s¹²³), it seems questionable to define the endpoint of the history of a genre in that way. Due to a range of historical events like the censorship that was prevalent in Greece from 1936, with a short interruption in 1946, until 1974, when the Regime of the Colonels fell, and the Axis occupation when record production was stopped, many rebetiko songs of that period existed in a transformed, censored version and were recorded for the first time only or re-recorded in the 1970s. Markos Vamvakaris, one of the most well-known representatives of rebetiko, said in an interview in 1971, one year before his death, that he was still writing rebetiko songs (Kounadis 2003a, 25). The opinion that all songs that have the aspiration to be rebetika but were recorded or published in written form after 1955 are not ‘real’ rebetika, is not convincing in the very least due to the fact that many rebetika of the period before 1955 were first recorded in the 1970s (for example, as already mentioned in the previous section, the song “Haidari” by Vamvakaris that was written during the Occupation, but first recorded only in 1980 by Dalaras). Another point for discussion is whether 1933 or 1936 should be chosen as the beginning of the second period of rebetiko. The year 1933 refers to the appearance of the first recordings of bouzouki from the USA, which became very popular in Greece, as well as to the establishment of the *Legendary Quartet of Piraeus* headed by Markos Vamvakaris in 1934 that signalled the transition from the Smyrna style to the Piraeus style in rebetiko. 1936 is the year in which Ioannis Metaxas came to power and established a dictatorship that resulted amongst others in the establishment of strict censorship in Greece. Zelepos defines the second period from 1933 to 1936/1940. Regardless of different exact periodisations, the 1930s are seen as a time of classical rebetika, a time when the Smyrna style merged with the Piraeus style and the latter one shaped the further stages of the development of rebetiko. This was not an entirely natural process and was caused not only by a general change of musical tastes of the audience. As already mentioned, oriental elements in music were not desirable after Metaxas came to power. The Smyrna school of rebetiko was associated with the image of an enemy and was not seen as a part of Greek culture and history at that time. Moreover, the recording capacity of 78 rpm records was limited

on a national but on religious identity, so that about 1,100,000 Greek Orthodox had to leave Asia Minor and Thrace and approximately 650,000 Muslims had to leave to Turkey.

123 Due to a large interest in Indian culture, which was caused by the widespread popularity of Indian films in the period between the 1950s and the mid-1960s, many Indian songs were adapted by the laiko exponents and provided with Greek lyrics (Zelepos 2001, 100-102). For example, the well-known Greek song «Καρδιά μου καημένη» (English “My Poor Heart”) sung in 1960 by Stratos Dionisiou and Veatriki Kali is a reinterpretation of Babis Bakalis of the Indian song from the film *Mother India* (1957).

to three minutes, which also contributed to the gradual disappearance of such established components of Smyrna style of rebetiko as *amanes* and *taximi*, that were performed at the end of a song (Zelepos 2001, 57, 110). The period of transition to *laiko* and the possible demise of rebetiko in the mid-1950s is often connected with the introduction of the four-course bouzouki by Manolis Chiotis instead of a three-course bouzouki that was popular before. The first known recording with a four-course bouzouki dates back to 1956 with the song «Θεσσαλονίκη μου» (English “My Thessaloniki”) composed by Manolis Chiotis and sung by Stelios Kazantzidis (disc number: Columbia DG 7229, Athens) (Pennanen 1997, 95). As Genitsaris points out, after World War II a new style of rebetika, the so-called *archondorebetika*, or noble rebetika, appeared. According to Genitsaris, the record companies and music venues at that time only employed those who played a four-course bouzouki, such that the old generation of bouzouki players had to look for jobs outside of the Greek capital (Genitsaris 1992, 67-68). The *baglamas*, the second primary instrument of the Piraeus style of rebetiko, almost disappeared from the rebetika of the post-war period (Zelepos 2001, 170). In the context of the four-course bouzouki, it is also important to mention that it was not invented in the 1950s, and the first four-course bouzoukia are considered to be produced already in the nineteenth century (Zelepos 2001, 97).¹²⁴ Notably, a four-course bouzouki is listed in the catalogue of 1913 of a musical instrument manufacturer ran by Anastasios Stathopoulos in New York (Moustakas 2019).

As mentioned above, the period between 1922 and 1933 was dominated by *Smyrneïko rebetiko*, or the Smyrna style of rebetiko, and the following period that ended either in 1936 or 1940/1941 was dominated by *Pireotiko*, or the Piraeus style. These are two main schools of rebetiko. In the following paragraphs, these two styles will be presented in more detail.

The *Smyrneïko rebetiko*, or songs in *café aman* style, had actually two main centres, namely Smyrna and Constantinople¹²⁵, which is the reason for a debate whether it is plausible to define a style that was typical for the Asia Minor in general by the name of a particular city, namely Smyrna. Anagnostou assumes that such a definition may be caused by the fact that Constantinople was associated with conservative traditions and ecclesiastic music, whereas Smyrna was seen as a more Europeanised city that was sometimes called the *Paris of Orient* (Anagnostou 2018, 6-9). Risto Pekka Pennanen argues that Constantinople was also an important cultural centre and a cosmopolitan city at that time. While stressing the Hellenocentrism of the term *Smyrneïko*, Pennanen suggests to call this musical style an Ottoman-Greek café music, “a term based on the ethnic background of the musicians and the main performance milieu” (Pennanen 2004, 2-4). Therefore, due to the ambiguity of the term *Smyrneïko rebetiko*, a range of alternative names was suggested. For instance, besides the already mentioned term Ottoman-Greek café music, also the term *Asia Minor rebetiko*, or *Mikrasiatiko rebetiko* was proposed. In this context, it is noteworthy that the term *Smyrneïko* could be seen not only in the context of hellenocentric views, but also as a

124 Some researchers connect the bouzouki with the ancient Greek *pandura* and Byzantine *tambouras* (Liavas 2009, 25) (Tragaki 2009, 94). Interestingly, bouzouki and *tambouras* are testified as music instruments that accompanied songs of struggle during the Greek War of Independence in the 1820s. For instance, Nikolaos Kasomoulis, an important participant of the Greek Revolution, used to play bouzouki, as he documented himself in his work “Military Memoirs of the Revolution of the Greeks 1821-1833” (Liavas 2009, 33).

125 In the present work the name Constantinople is used due to the fact that the city was renamed to Istanbul only in 1923 when the Republic of Turkey was established. The official change of name took place only in 1930. The present research involves mainly the period in the history of the city before 1923.

term expressing counter-nationalism. As Anagnostou points out, in the 1970s “[it] marked a counter-nationalism, a revolt against the dominant image of the Greek and an effort to include the refugees in the interpretative narratives of twentieth-century Greek history. The term highlights internal divisions within Greek society concerning the definition of Greekness [...]” (Anagnostou 2018, 8-9).

According to Kounadis, Smyrneïko as a music category was formed around the 1830s in the Asia Minor with two main centres, namely Smyrna and Constantinople (Kounadis 2003a, 393). As already mentioned, one of the main differences between the Piraeus and Smyrna style of rebetiko is the instrumentation. Although rebetiko is usually associated with music instruments like the bouzouki and the baglamas, the Smyrna school of rebetiko originally did not use them. The bouzouki was first recorded in 1917 in Görlitz, but became a popular instrument in rebetiko studio recordings only at the beginning of the 1930s. Primary instruments of the Smyrna style of rebetiko were violin, oud, santour, qanun (or kanonaki, as it is called in Greek) or politiki lyra (or Constantinopolitan lyra). Another difference between these two styles concerns the figure of the singer: while Smyrneïka rebetika often included a female singer, the Piraeus rebetika were male dominated, with a characteristically rough voice, which one can hear, for example, in the songs of Vamvakaris (Zeleps 2001, 99) (Kounadis 2003a, 36). The Asia Minor Disaster of 1922 and the wave of refugees to Greece resulted in a change of themes raised in Smyrneïka due to the harsh realities of life. As mentioned above the period of rebetiko from 1922 to 1933 was dominated by the Smyrna style. From 1935 onwards, one observes the assimilation of Smyrneïko into the prevailing Pireotiko rebetiko (Kounadis 2003a, 394-395). Due to differences in performance of Piraeus and Smyrna rebetiko, some scholars put Smyrneïko into a separate category. When considering Smyrneïka as part of rebetika, the question arises from which point in time one could include them into rebetiko. As Kounadis points out, Smyrneïko as an urban popular song appeared around 1830 in Asia Minor. Therefore, it is unclear, whether one should include Smyrneïka from at least 1830 onwards, from 1873, when the first café aman with a corresponding repertoire was attested in Athens, or only from 1922, a year before refugees from Asia Minor came to Greece, which resulted in a boom of the Smyrna style of rebetiko in Greece. As Anagnostou points out, “[the musical category Smyrneïko] is a part of rebetiko but at the same time distinct [...] [and] is located at the blurred frontier between demotic song and rebetiko” (Anagnostou 2018, 9).

In mainland Greece, one can also define more than one centre of rebetiko. For example, besides Piraeus rebetika one can observe the rebetiko tradition of Thessaloniki. Despite the strict censorship imposed in 1936, the Thessalonikian police, and especially its chief Nikolaos Mouschoundis, showed a protective attitude towards rebetiko exponents, so that many of them moved to Thessaloniki or performed there regularly (Tragaki 2009, 58-64). Among them were rebetiko representatives such as Markos Vamvakaris, Giorgos Batis, Giannis Papaioannou, Stratos Pagioumtzis and Vasilis Tsitsanis, and most of them knew Mouschoundis, the fan of bouzouki, personally. Vasilis Tsitsanis, who lived in Thessaloniki during the war, even opened an ouzeri, or ouzo bar, there, in which he regularly performed rebetika. Due to the flourishing of rebetiko in Thessaloniki from 1936 onwards, some rebetiko researchers put the Thessalonikian school of rebetiko into a separate category: among them, for example, the Thessalonikian poet and rebetiko expert Dinos Christianopoulos (Tragaki 2009, 65).

To summarise, one can observe more than two styles of rebetiko, namely Smyrna and Piraeus styles. Moreover, although the period from 1922 until circa 1933 was dominated by Smyrneika, and the period afterwards by Pireotika, these two styles coexisted to a certain extent. Due to the fact that one deals mainly with the output of record labels, and rebetika accompanied by bouzouki and baglamas started to be actively recorded only in the 1930s, existing before mainly in the close communities of hashish dens and prisons, the songs that one has at its disposal today as a researcher are mainly Smyrneika rebetika before circa 1934, when Markos Vamvakaris made his first recordings. A similar problem arises for Smyrneika rebetika, which are often synonymously called *café aman rebetika*, as *café amans* were not the only venues in the Ottoman Empire and probably Greece, where such songs were performed. The songs performed in *café amans* to a certain extent depicted a stylisation and romanticisation of a low-life context that were created to satisfy the audience (Tragaki 2009, 65) (Gauntlett 1985a, 70, 73). Therefore one observes a certain similarity in the origins of both styles of rebetiko that were performed in the low-life haunts and spread into another sort of venues: Piraeus rebetika from hashish dens and prisons to taverns, and Smyrneika to *café amans*. Moreover, even though Piraeus rebetiko started to dominate rebetiko from 1934 onwards and the Smyrna style merged with it to a certain extent, *café aman rebetika* continued to exist.

In this section different hypotheses concerning the periodisation of rebetiko were presented, with the heyday of rebetiko tradition between 1922 and 1955. The rebetiko tradition that goes back to at least as far as the mid-19th century, was rediscovered in the 1960s, at a time of debates about a ‘proper’ form of a Greek music. Today, in the period of the second revival that started around 1980, rebetiko continues to be practiced at the stages of numerous music venues. Also, the ambiguity and the characteristics of two well-known styles of rebetiko were discussed. The landscape of rebetiko is certainly much broader than Piraeus and Smyrna, even though they present an immense part of the rebetiko that has survived to this day. It also includes Athens, Thessaloniki, Constantinople/Istanbul and other parts of Greece and the Ottoman Empire/Turkey, and even America and Australia, which have large Greek diasporas.

2.4 THE QUESTION OF MARGINALITY OF REBETIKO

«Βγήκα από το Πανεπιστήμιο του πεζοδρομίου...»

— Markos Vamvakaris^a

^a As cited in (Kounadis 2003a, 25)

Rebetiko is often defined as the music of the urban underworld. Especially at the beginning of its existence this music genre had a bad reputation due to the fact that it was associated with marginal social groups. The growing interest towards rebetiko from the 1960s onwards, when the first rebetiko revival began, or even earlier, namely at the end of the 1940s, when one of the first works on the cultural value of rebetiko appeared, lead to a more diverse perception of rebetiko. As Gauntlett points out, at one extreme there are those who connect rebetiko with the urban criminal underworld, and at the opposite those “who minimize underworld involvement”. There are also those in the middle who associate

the origins exclusively with the underworld from which it departed to its ‘genuineness’ (Gauntlett 1985a, 23-24).

The association of rebetiko with the underworld is caused by the fact that many representatives of rebetiko music came from the poor social classes and that early rebetika mainly depicted underworld topics. At the same time, due to a number of historical events such as Greece’s unsuccessful war with the Ottoman Empire in 1897, a difficult economic situation at the end of the 19th century (including the bankruptcy of the Greek state in 1893), the Balkan wars of 1912-1913, the first World war, and the mandatory exchange of population in 1923 between Greece and Turkey after another bloody war, it is difficult to determine whether these groups were actually as marginal as they have been presented for a long time. Frequent topics of rebetiko songs were difficulties of life such as poverty, illnesses, sorrow, despair, unrequited love, unfairness of life, gambling, drug consumption, and criminality. Taking into account the conditions of life in Greece at that time, we come to the conclusion that the issues covered in rebetiko songs mattered for much larger parts of the society than only rebetiko exponents. Moreover, one cannot equate the rebetiko exponents with the figures depicted in rebetiko song lyrics. The well-known Greek scholar and researcher of rebetiko music Panagiotis Kounadis mentioned in an interview given to the *Stavros Niarchos Foundation* about Markos Vamvakaris, who is one of the main representatives of rebetiko, that “[Vamvakaris] never wrote anything more than his own life stories. His life was the same as for millions of other people [in Greece]”.¹²⁶

The words mentioned in the epigraph belong to Markos Vamvakaris and can be translated as “I came out of the street university”. The quote seems to reflect the nature of rebetiko well. As already mentioned, rebetika for a long time have been considered to be of lesser value, compared to other music in Greece. It was labeled as the music of the underworld, drug addicts, prostitutes and other marginals. Representatives of rebetiko were criticised for using oriental music elements¹²⁷ that did not suit the ideology of Greece at that time, which saw everything that reminded of the Ottoman Empire as hostile and tried to reconstruct the idea of continuity of the Greek state from the antiquity to the present day. At the same time, music, like every other form of art, also exists outside of the national ideologies. Every music genre experiences the influence of its predecessors and is in a constant exchange with neighbouring music traditions. Rebetiko was no exception. It has absorbed elements of the music tradition of ancient Greece, Byzantine church music, and the Turkish and Arabic music tradition in general. Many melodies of rebetika songs are based on the melodies of Greek traditional music, the so called *paradosiaka*. Rebetiko represents both in music and in text a valuable source of Greece’s history.

Even if one considers rebetiko songs that depict hashish consumption, in Greece this was not something connected exclusively with rebetiko. Hashish as well as some other drugs were for a long time seen as a medicine. Throughout the nineteenth century drugs were widespread among scientific and literary circles. Jacques-Joseph Moreau, the author of the book *Hashish and Mental Alienation*, which is considered to be the first scientific work about

126 <https://www.snf.org/el/grafeio-typou/lista-neon/2019/03/arheio-kounadi-%E2%80%93-mia-nota-istorias-rempetikou/> (accessed January 24, 2022)

127 For example typical oriental instrumentation or oriental maqams by melody construction, although it is questionable to call them purely oriental as well as purely Greek, as it was a mixture of different traditions, that was being building over centuries. The division into Oriental and originally Greek tradition is getting even more problematic when we take into account the fact that the Greeks were living for many centuries between the Orient and Occident.

the side effects of drug consumption, was a hashish consumer himself and a member of the Parisian *Club des Hashischins*.¹²⁸ Furthermore, cannabis cultivated in Greece was an important export good at the end of nineteenth and the beginning of twentieth century (see (Gotsinas 2021)). The prohibition of cannabis in Egypt in the 1880s only increased cannabis cultivation in Greece. Although a law prohibiting the public use of hashish was introduced in 1891, and a law concerning the prohibition of cultivation, trade and consumption of cannabis was introduced in 1920, it only came into force in 1936 after the censorship of Metaxas was established.

In his widely known lecture on rebetiko, Chatzidakis made the following comment: “Well, I don’t think that this snobbishness regarding the rebetiko song can stand in our way of appreciating its worth and loving it for the truth and the power it contains. These songs are familiar and intimate to us in so many ways that nothing comes even remotely close to them” (cited as in (Liavas 2009, 255)). Many representatives of rebetiko were refugees from the Asia Minor who came to Greece as a result of the mandatory exchange of population between Greece and Turkey in 1923. More than a million refugees arrived in Greece, which corresponded to about 20% percent of the population of Greece at that time (Kostis 2018, 275). They were forced to leave behind almost everything in Turkey and had nothing in Greece. Difficult economic situations and lack of jobs also caused active inner migration in Greece at the end of the nineteenth and the beginning twentieth century: for instance, a twelve-year-old Markos Vamvakaris and his father were forced to leave Syros and went to Piraeus in search of a job. These harsh realities of life found their depiction in rebetika, which referred not only to rebetiko exponents, but to a much larger number of Greeks. In the course of time, the number of Greeks which could identify with or relate to the themes covered in rebetiko constantly grew, which was caused by cataclysms such as World War II and the Civil War afterwards.

Furthermore, throughout its history rebetiko involved a large number of individuals that were not related to the underworld. As it was already mentioned in the Section 2.1, as early as in the 1930s the jobbing versifiers, or *logides*, were used by the rebetiko exponents, who otherwise were not able to keep up with the demand for new lyrics by the record industry. Many of them were well-educated and had no connection to the underworld, for instance, the lyricist Charalambos Vasileiadis who also worked as a translator and spoke five languages. The Greek journalist and ethnographer Konstantinos Faltaits, who may be considered the first to stress the artistic merit of rebetiko songs and the importance of their research, composed rebetika himself under the pseudonym K. Roumeliotis.

In summary, it would be false to deny the connection between rebetiko and marginal groups. As rebetiko exponents testified themselves in interviews and (auto)biographies, outlaw life characterised the early period of rebetiko. For instance, Michalis Genitsaris mentioned in his autobiography numerous violent affrays and criminal affairs in which he participated himself and that excessive alcohol and hashish consumption was widespread among many rebetiko exponents. At the same time, a number of rebetiko exponents including, for instance, Vasilis Tsitsanis had little or no connection to the underworld. Moreover, in the course of time, songs that were a romanticised stylisation of the underworld became popular outside of their social context. This romanticisation of the figure of *mangas*, a common protagonist of rebetika, started quite early – at least as soon as rebetika entered the

128 Among the members and visitors of the *Club des Hashischins* were Alexandre Dumas, Charles Baudelaire, Honoré de Balzac and Victor Hugo (Levinthal 2012, 167-172).

repertoire of the *cafés amans* (Gauntlett 1985a, 70, 73). Songs depicting low life were further stylised by the second generation of bouzouki players and reached their climax in the period of the rebetiko revival. In general, the recordings, which are the main source of knowledge on rebetiko today, have always included a certain degree of stylisation. The censorship that was imposed in Greece in 1936 and existed with a short interruption in 1946 until the fall of the Junta in 1974, also had a great impact on the context in which rebetika were performed as well as on a broadening of topics of rebetika and its removal from the underworld context. The history of rebetiko and its exponents is therefore a complex one, and should not be reduced to its link to the Greek underworld, and terms such as *underworld* and *marginal groups* deserve a deeper discussion.

2.5 FROM THE ORAL TRADITION TO THE COMMERCIALISATION OF REBETIKO

Sometimes representatives of rebetika were criticised for taking old traditional Greek songs and presenting them as their own creation (Kounadis 2003a, 121). First of all, it is important to mention that the boundary between the reinterpretation of an old work and plagiarism is not that clear in music, especially in popular music that in the beginning was transmitted in the oral tradition. Secondly, the boundary between Greek traditional music and rebetika is also not that clear. The creation of a new music genre or a transition of one music genre into another is a long process, thus most musicians work in several music genres simultaneously.¹²⁹ Another important detail is that rebetiko existed at the beginning in the oral tradition¹³⁰, so that the transition from oral to a written tradition with specified authorships of songs caused that one and the same song was attributed to several musicians. In the context of rebetiko, the concept of intellectual property became important only after the beginning of the commercialisation of rebetiko, i.e. the beginning of massive record production that started in the second half of the 1920s and the early 1930s. The establishment of the first Greek organisation for the protection of intellectual property¹³¹ in 1930 also contributed to the strengthening of the idea of intellectual property (Kounadis 2003b, 289).

One of the characteristics of rebetiko song lyrics are spoken interjections, or *tsakismata*¹³², like «Γειά σου Μάρκο με το μπαγλαμά σου!»¹³³ (“Health to you, Markos, with your baglamas!”), «Γειά σου Μάρκο μου ντερβίση!»¹³⁴ (“Health to you, Markos, you dervish!”), «Γεια σου, Ρί-

129 For example, Markos Vamvakaris, one of the most well-known representatives of rebetiko, performed some traditional songs or used the melody of traditional songs in his own creations and at the same time presented a large variety of his own creations. The ethnomusicologist Lambros Liavas observes Vamvakaris creations as a connecting link between *dimotika*, or demotic (‘folk’) songs, and rebetika (Liavas 2009, 16).

130 Rebetika continue to exist in oral tradition also after the beginning of record production: for instance, in the works of rebetiko exponents of the Metaxas period who were sent into exile on the Aegean islands or during the Occupation period when record production was suspended (Gauntlett 1985a, 186-187).

131 *Ανώνυμη Εταιρεία Προστασίας της Πνευματικής Ιδιοκτησίας (ΑΕΠΙ)*

132 Remarkably, there exists a website that provides a *tsakismata* generator that randomly selects interjections from a pool of 11,688,200,277,601 different interjections. It can be found at <http://tsakismata.rebetiko.org> (accessed January 12, 2022).

133 For more details see the song «Όταν πλύνω τουμπεκάκι» (English “When I Clean the Hookah Tobacco”) by Markos Vamvakaris (disc number: His Master’s Voice AO-2065, recorded in 1933).

134 For more details see the song «Φραγκοσυριανή» (English “Catholic Girl from Syros”) by Markos Vamvakaris (disc number: His Master’s Voice AO-2280, recorded in 1935).

τα!»¹³⁵ (“Health to you, Rita!”), «Γεια σου Μιχάλη Γενίτσαρη, λεβέντη!»¹³⁶ (“Health to you, Michalis Genitsaris, you lad!”), «Γεια σου Καρίπη αθάνατε, γεια σου!»¹³⁷ (“Health to you, Karipis, you immortal, health to you!”), «Να μου ζήσεις Ογδοντάκη μου, γεια σου!» (“Long live my Ogdontakis, health to you!”) and «Γεια σου Τζοβενάκη με το σαντούρι σου!»¹³⁸ (“Health to you, Tzovenakis, with your santouri!”), «Γεια σου Ογδοντάκη με το βιολί σου!»¹³⁹ (“Health to you, Ogdontakis, with your violin!”), «Γεια σου Λάμπρο μου με το κανονάκι σου!»¹⁴⁰ (“Health to you, Lambros, with your qanun!”).

The use of such interjections can be interpreted as a way of highlighting or specifying the artist, as the concept of intellectual property appeared in the context of rebetiko only with the beginning of its commercialisation. This hypothesis was also proposed by Stathis Gauntlett, who also mentions that such shout-outs were used in a live performance for coordination between musicians on the stage (Gauntlett 1985a, 92).

The inclusion of spoken interjections in live performances is something widely used, but their appearance in studio recordings is not expected. Their function in the context of rebetiko is multidimensional. The role of spoken interjections in rebetiko was thoroughly studied in a paper by Michael Kaloyanides.¹⁴¹ He defines four forms of interjections that are used in rebetika: the so-called kefi-call¹⁴² (a short exclamation like *Opa!* or *Aman!*), the *yiassou* shout-out (a personalised toast or a praising phrase like those mentioned in the previous paragraph), keepin’ it real phrases that create an atmosphere of imagined performance for the exponents themselves and the audience¹⁴³, and the extended dialogue¹⁴⁴ that appears at the beginning of the recording and resembles the dialogues of the Greek revue theatre (*epitheorisi*).

As already mentioned, one can interpret the use of interjections in rebetiko studio recordings in different ways. The *yiassou* interjections always contain the phrase *γεια σου* which literally means ‘to your health’, as well as the name of the addressee, being therefore a way to establish ownership of the song. The last element of the *yiassou* interjections, a qualifier,

135 For more details see the song «Λειβαδιά» (English “Leivadia”) of unknown authorship that has been recorded more than 15 times. The recording in question was released in 1933 and sung by Rita Abatzi (disc number: Columbia DG-471).

136 For more details see the song «Εγώ μάγκας φαινόμενα» (English “I Seemed a Tough Guy”) by Michalis Genitsaris (disc number: Columbia DG-6312, recorded in 1938).

137 For more details see the song «Οι μπαγλαμάδες» (English “Baglamades”) of unknown authorship from the year 1927 (disc number: Odeon GA-1249).

138 For more details see the song «Οι ψαράδες» (English “The Fishermen”) of unknown authorship from the year 1928 sung by Karipis (disc number: Polydor V-50202).

139 For more details see the song «Οι νέοι χασικλήδες» (English “The Young Hashish Smokers”) of unknown authorship from the year 1928 sung by Karipis (disc number: Odeon GA-1346).

140 For more details see the song «Η Αρμενίτσα» (English “The Armenian Girl”) by Kostas Karipis from the year 1934 sung by Rita Abatzi (disc number: His Master’s Voice AO-2167).

141 Kaloyanides, Michael. “‘Health to You, Marko, with Your Bouzouki!’: The Role of Spoken Interjection in Greek Musicians’ Imagined Performance World in Historical Recordings Made in America and Abroad.” in (Bucurvalas 2018)

142 The Greek word *κέφι* was borrowed via Turkish from Arabic and stands for ‘a sense of high spirits’.

143 The song «Το παιχνίδι του Αμερικάνου» (English “The Game of an American”) composed by Kostas Skarvelis and sung by Rita Abatzi (disc number: His Master’s Voice AO-2305) from the year 1936 includes three forms of interjections: a kefi-call (*ώπα*, English ‘hop’), *yiassou* shout-outs («Γεια σου Σαλονικιέ μου!», English ‘Health to you, my Salonikios’), and a keepin’ it real phrase («Ψοφάω για δολάρια!», English ‘I am Crazy about Dollars’).

144 The song «Η φωνή του αργιλέ» (English “The Sound of a Hookah Pipe”) composed by Vangelis Papazoglou and sung by Stellakis Perpiniadis, in which the bubbling of a waterpipe can be heard as part of an accompaniment, includes an extended dialogue in which an arghileh, or a waterpipe, is also the topic of the conversation (disc number: Columbia DG-6066, recorded in 1935).

as Kaloyanides defines it, differs. As it is shown above, it often defines the instrument the musician is playing, which is a very valuable source of information about the orchestra, as the rebetiko music performers were in most cases not specified on the labels of the records or in the catalogs and archives of the record companies. Instead, one usually finds phrases such as *δημώδης ορχήστρα* or *λαϊκή ορχήστρα* which can be translated as ‘demotic orchestra’ or ‘popular orchestra’ respectively.¹⁴⁵ Interestingly, in some cases the music instrument, for which the musicians are praised in the interjections, did not correspond to the ones they were actually playing. As Kaloyanides points out, it could be either by mistake or a way ‘to authenticate the recording as part of a particular genre with characteristic defining instruments’.¹⁴⁶ As Kaloyanides mentions, Spyros Peristeris sometimes is praised as a bouzouki player, despite actually playing a guitar. Some notice that some song recordings that include interjections praising Spyros Peristeris as a bouzouki player, include only a guitar. Others claim that Peristeris used to play a bouzouki with three single strings (instead of a traditional three double course bouzouki, i.e. containing three double strings tuned in unison).¹⁴⁷

Besides such qualifiers as music instruments, *yiassou* interjection often include qualifiers such as, for instance, *mangas*, *dervish* and *levendis* (English “lad”), which are used to “establish the musician’s street cred” (see examples of interjections presented above).¹⁴⁸

The *Rebetiko Corpus* contains 448 songs that include the so-called *yiassou* spoken interjections, out of 3,772 songs with provided song lyrics.¹⁴⁹ Remarkably, in these interjections, that were very popular in rebetika until the end of 1930s¹⁵⁰, some exponents of rebetiko are referred to by their given name (for example, Markos Vamvakaris), others by their surname (for example, Vasilis Tsitsanis) or by both given name and surname (for example, Michalis Genitsaris). This could be explained by the fact that during the history of rebetiko the role of the rebetiko exponent has changed, and music was becoming more and more a commercial product that was also connected with the idea of the intellectual property. In the beginning, the rebetes performed while being very close to the audience in a hashish den, a gambling dive or a café aman, then they moved into a tavern, and finally to the stages of local en-

145 If the recording was made in the United States, the phrase ‘Greek popular orchestra’ is also common.

146 Cited as in Kaloyanides, Michael. “‘Health to You, Marko, with Your Bouzouki!’: The Role of Spoken Interjection in Greek Musicians’ Imagined Performance World in Historical Recordings Made in America and Abroad.” in (Bucuvalas 2018)

147 The song «*Τεκετζής*» (English “The Hash-den Owner”) composed by Spyros Peristeris and sung by Zacharias Kasimatis is considered to be orchestrated by a guitar and a bouzouki with single strings played by Peristeris (disc number: Parlophone B-21707, recorded in 1933-1934 in Athens). This assumption is made due to the presence of interjection praising Peristeris as a bouzouki player: namely, «Γεια σου ρε Σπυράκι με το μουζουκάκι σου!» (English “Health to you, Spyraiki, with your sweet bouzouki!”). The recording of the song can be found at <https://www.youtube.com/watch?v=hjnpzHEU3rg> (accessed January 12, 2022). Another example of a song about which instrumentation is being argued, is the song «*Μυστήριο ζεϊμπέκικο*» (English “Mysterious Zeibekiko”) of unknown authorship (disc number: Columbia DG 275, recorded in 1932-1933 in Athens). According to Leonardos Kounadis and Nikos Ordoulidis, although the bouzouki is mentioned on the label of the record, it is not included in the instrumentation. The record and the commentary of Leonardos Kounadis and Nikos Ordoulidis can be found at <https://www.vmrebetiko.gr/item/?id=9872> (accessed January 12, 2022).

148 Cited as in Kaloyanides, Michael. “‘Health to You, Marko, with Your Bouzouki!’: The Role of Spoken Interjection in Greek Musicians’ Imagined Performance World in Historical Recordings Made in America and Abroad.” in (Bucuvalas 2018)

149 It is also important to bear in mind that such phrases were often not included in the song text by song lyrics collectors at <https://rebetiko.sealabs.net> (accessed January 12, 2022), and it was not possible to verify the song text by hearing the song record for all the data included in the corpus. This means that the real number of such interjections in rebetiko recordings is even higher.

150 This observation was made during the work with the corpus of rebetiko song lyrics.

tertainment centres (*kosmika kentra*), films, radio and title pages of the magazines (Liavas 2009, 225). Markos Vamvakaris belongs to the first group of rebetiko exponents and had a drastic influence on the rebetiko of the 1930s, while Vasilis Tsitsanis shaped to a large extent the rebetiko after the second world war, when the idea of a rebetiko musician as a popular star was already formed. This could explain why Markos Vamvakaris was addressed in the songs always only by his given name and Vasilis Tsitsanis by his surname.

Interestingly, in this context Tragaki points out, that due to the fact that in professional music making coordination was needed, the elements that were considered to be “the original improvised and unshaped oral tradition gradually became standardised and fixed” (cited as in (Tragaki 2009, 49). It is possible that this was also the case with the interjections. Kaloyanides also points out, that the early rebetiko studio records did not differ much from the live performances.¹⁵¹ The norms for a studio record were formed later, so that some elements typical for live performances were kept even in later studio recordings to become a distinguishing feature of rebetiko, emphasising its ‘authenticity’. The inclusion of interjections in studio records could also go back to the oral tradition of early rebetiko, being therefore again a special characteristic emphasising the ‘authenticity’ of a particular song’s recording as being part of a particular genre.

To summarise, interjections may not only be a way of perpetuating the work of an artist, but also the way of creating a virtual atmosphere of a live rebetiko performance for the musicians themselves as well as the future audience.

In the context of rebetika, we also often deal with the fact that a song was given as a present to another musician, thus it is often very difficult to ascertain the author of a particular rebetiko song. For instance, Vangelis Papazoglou is considered to have given some of his songs as a present to other composers and singers (Kounadis 2003a, 41). Giannis Kyriazis, another rebetiko exponent, claimed to be the singer of about 100 songs, of most of which he has written the lyrics himself, but which he had given as a present to his friends (Kounadis 2003a, 61). Correctly assigning credit to lyricists generally is a problem in the context of rebetiko, due to the fact that before circa 1946 lyricists were not specified on the labels of the records and therefore the authorship of song lyrics in this period remains often unclear (Gauntlett 1985a, 163).

Due to the fact that early rebetika were passed on only in oral tradition, some of them were then reinterpreted and recorded in many variations. For instance, there exist at least 15 variations of the song “Leivadia”, which itself is a variation of “Ali Pasha”, a well-known song from Ioannina (*Gianniotiko*).¹⁵² In the case of “Leivadia”, and a range of other songs recorded especially during the first two decades of the twentieth century, it is unclear whether they could be considered as belonging to the rebetiko genre. Some researchers would probably refer to it as a demotic or a traditional song (*paradosiako*).

While working with the corpus, another peculiarity of rebetiko was discovered, namely that many rebetika have similar titles, although in terms of content they are different. One of the words that was often used to build a new title from the old one is the adjective ‘new,

151 Cited as in Kaloyanides, Michael. “‘Health to You, Marko, with Your Bouzouki!’: The Role of Spoken Interjection in Greek Musicians’ Imagined Performance World in Historical Recordings Made in America and Abroad.” in (Bucuvalas 2018)

152 According to Aristomenis Kalyviotis, the song “Ali Pasha”, which was devoted to the Ottoman Albanian ruler Ali Pasha of Ioannina, was recorded 34 times in the period between 1900 and 1926. More information about the songs “Ali Pasha” and “Leivadia” can be found at <https://www.vmrebetiko.gr/item/?id=5344> and <https://www.vmrebetiko.gr/item/?id=4409> (accessed January 13, 2022).

young’: for instance, «Μερακλού» (English “Devotee”) and «Νέα μερακλού» (“Young Devotee”), «Σμυρνιά» (“Girl from Smyrna”) and «Νέα Σμυρνιά» (“Young Girl from Smyrna”), «Η χήρα» (“The Widow”) and «Νέα χήρα» (“Young Widow”), «Ο μάγκας» (“The Mangas”) and «Ο νέος μάγκας» (“The Young Mangas”). Although these songs have similar names, the songs containing the adjective ‘new, young’ in their title are not reinterpretations of the older songs, but independent creations. The large number of such songs may be evidence of a particular fashion of re-using older titles in the Greek record industry of that time. Another reason of such a phenomenon may be the large success of the songs with the original title, such that the use of the same title with the addition ‘new, young’ could help promote a new song.

In conclusion, due to the oral tradition of early rebetiko, many songs were later on recorded by multiple different artists. Many of these *adespota* (‘ownerless, without author’) rebetika were recorded in the first two decades of the twentieth century, when the record production began, as well as from the 1960s onwards, when the first rebetiko revival began and rebetiko enthusiasts were in search of ‘real’ rebetika. Another peculiarity in the context of rebetiko authorship is the use of spoken interjections, which is usually not typical for studio records. Moreover, many rebetika have similar titles, but this is not automatically a sign of plagiarism or reinterpretation. Many of them are similar neither in a musical aspect, nor in text. Finally, even though oral tradition is usually associated with early rebetika of the second half of the nineteenth century, or even earlier, and the first two decades of the twentieth century, rebetika continued to be passed on orally also after the beginning of active record production in the second half of the 1920s and especially the 1930s. The censorship by Metaxas, which resulted in massive exiles of rebetiko exponents to the Aegean islands, and the suspension of record production during the Occupation period, contributed to the continuation of oral tradition also after the beginning of the commercialisation of rebetiko.

2.6 EXAMPLES OF REBETIKO SONGS

This section presents and analyses lyrics of several rebetiko songs that were recorded at different points in time in the course of the evolution of the music genre. However, the inclusion of particular songs in this section does not mean, that they should be considered the most characteristic to a particular period of rebetiko existence. These examples were selected with the intention to create a general picture of how rebetiko songs look like. The number of presented songs was limited to three, and they represent only a limited spectrum of themes raised in rebetika.

First, the song «Οι μπαγλαμάδες» (“Baglamades”) is presented. This recording, which dates back to 1928, is of unknown authorship and sung by Marika Papangika (disc number: Columbia USA 56138-F, New York).¹⁵³

Στους απάνω μαχαλάδες (3x)

διώξανε δυό ντερβισάδες

είχανε αργιλεδάκι

και φουμάρανε μαυράκι

In the upper mahallas (3x)

two dervishes were chased away

They had an arghileh

and smoked black weed

¹⁵³ The recording of the song «Οι μπαγλαμάδες» (“Baglamades”) can be found at <https://www.vmbetiko.gr/item/?id=4099> (accessed January 8, 2022).

παίζαν το μπαγλαμαδάκι για να σπάσουνε μεράκι.	They played bagmadaki to whoop it up
Γύρω-γύρω μπαγλαμάδες (3x) και στη μέση ντερβισάδες	Bagmades were all around (3x) and dervishes were in the middle
και φουμάρουν χασισάκι για να σπάσουνε μεράκι	And they smoke hash to whoop it up
παίζουν τα μπαγλαμαδάκια και χορεύουν ντερβισάκια.	Dervishes play baglamadakia and dance
Ο λουλάς και το καλάμι (3x) μ' έφεραν σ' αυτό το χάλι	A pipe and an arghileh straw (3x) brought me to this mess
ο λουλάς και το νεφέσι, μ' έφεραν σ' αυτή τη θέση	A pipe and joy from arghileh brought me to this place
ο λουλάς και το χασίσι, μ' έφεραν σ' αυτή την κρίση.	A pipe and hash brought me to this bout.

The song is accompanied by an orchestra consisting of a violin, a cello (Markos Sifnios) and a cimbalom (Kostas Papangikas) and is an example of the zeibekiko rhythm, which is the most popular rebetiko rhythm and dance. As already mentioned, the primary instrumentation of the Asia Minor (Smyrna) style were violin, qanun (kanonaki), santour and guitar, which is common in popular ensembles named *santouroviolia* that performed in cafés amans. At the same time, the instrumentation in the recordings of the Asia Minor style of rebetiko varied due to the fact that record companies employed well-known musicians that often performed songs in different music genres: cello and cimbalom, for instance, appear in a range of rebetika of that period. As it is often the case with rebetika, this song was rerecorded and reinterpreted many times: nine recordings of this song are known, with three recordings being sung by Marika Papangika, and its first known recording being sung by Mme. Coula in 1919 or 1920 in New York (disc number: Panhellenion 208-B).¹⁵⁴ It is an example of *adespoto chasiklidiko rebetiko* from the Asia Minor (Smyrneïko), or a rebetiko song of the period of oral tradition with the main theme of hashish consumption. Underworld activities, which among others include hashish consumption, prison life and violence, are considered to be the main theme of rebetika of the oral tradition (Gauntlett 1985a, 57-58). The presented song consists of iambic octosyllabic couplets, which, together with the trochaic octosyllabic and iambic decapentasyllabic (Political) verses, is one of three most frequently used metrical forms in the rebetika of the period before active commercialisation in the 1930s began (Gauntlett 1985a, 71). It is noteworthy that the text of the song «Οι μπαγλαμάδες» (“Baglamades”) contains a number of Turkish loanwords: for instance, *μαχαλάς* (English ,mahalla, district‘), *ντερβίσης* (English ,dervish‘), *μεράκι* (English ,passion, strong desire‘), *λουλάς* (English ,arghileh pipe/ bowl‘) and *νεφέσι* (English ,joy from smoking arghileh‘). The reason for this is the Asia Minor origin of the song. Many rebetiko exponents, who originally came from the Asia Minor and other parts of the Ottoman Empire, including Marika Papangika¹⁵⁵, performed songs in both Greek and Turkish. The number

154 All nine known recordings of the song «Οι μπαγλαμάδες» (“Baglamades”) can be found at <https://www.vmrebetiko.gr/item/?id=10451> (accessed 10 January 2022).

155 Marika Papangika was originally from the island of Kos which was under the Ottoman rule till 1912.

of Turkish loanwords decreases with the prominence of the Piraeus style of rebetiko in the early 1930s and the censorship imposed in 1936 that persecuted references to Greece's Ottoman past.

This song also demonstrated another peculiarity of improvised rebetika of the oral tradition and their recordings in the 1920s and early 1930s, namely that they rarely present a coherent narrative, instead often using a desultory narrative with independent themes in each couplet. As Gauntlett points out, this fits with the pattern of improvisation (Gauntlett 1985a, 60-61). In this context it is remarkable, that the narrative perspective within the song "Baglamades" changes. At first, the audience is told that two dervishes were expelled from some place located in the upper districts (mahallas) of a city. Then dervishes are depicted smoking arghileh, playing baglamades and dancing supposedly inside a tekes (hashish den). In conclusion, the narration changes from the third-person perspective to the first-person: the protagonist laments that arghileh and hashish smoking brought him into a difficult situation.¹⁵⁶ The audience can only speculate, which misfortune fell upon the protagonist. Dervishes are often figures of rebetika, and the word is used synonymously with *mangas* ('brave guy, outlaw'). The haunts in which they were smoking arghileh and played music had a mystical quality. Thus, they were called tekedes, which is originally a name for dervish monasteries.

The second song that will be presented is «Αντιλαλούν οι φυλακές» ("The Prisons Resound"), which was composed by Markos Vamvakaris. The first recording of the song was made in 1935/1936 and was sung by Markos Vamvakaris himself (disc number: Odeon GA-1918, Athens).

Αντιλαλούν οι φυλακές (3x) τό Μπούρτζι και ο Γεντί Κουλές	The prisons resound (3x) Bourtzi and Yedi Kule
Αντιλαλούν δυό σύρματα, Συγγρού και Παραπήγματα	Two prisons' wires resound, Syngrou and Parapigmata (barracks) ¹⁵⁷
Αν είσαι μάνα και πονείς (3x) Έλα στην Δίκη να με δεις	If you are my mom, and it hurts you come to the Court to see me
- Γεια σου ρε Μάρκο βασάνη!	- Health to you, Markos, you sufferer!
Έλα πριν με δικάσουνε Κλάψε να μ' απαλλάξουνε	Come before I am sentenced, cry for me to be released
- Γεια σου Αρτέμη με το μπαγλαμά σου!	- Health to you, Artemis, with your baglamas!
Τό σκότος καί ή φυλακή (3x) ειναι μεγάλο λακριντί	The darkness and prison (3x), They are a big wail of pain.
Αντιλαλούν οι φυλακές, (3x) το Ρίο και ο Γεντί Κουλές.	The prisons resound (3x) Rio and Yedi Kule

The song is performed by the *Legendary Quartet of Piraeus* and is accompanied by bouzouki (Markos Vamvakaris), guitar (Kostas Skarvelis) and baglamas (Anestis Delias, or Artemis).

156 "Baglamades" is an interesting example of a song in which the protagonist blames arghileh for making him troubles, as it usually presented as a way out from the harsh realities of life and not backwards.

157 Parapigmata were barracks of the Athens Military Prisons on the Vouliagmenis Avenue.

In the interjections, the voice of Stratos Pagioumtzis, the fourth member of the quartet, can also be heard. It is also an example of a 9/8 zeibekiko rhythm. The song «Αντιλαλούν οι φυλακές» (“The Prisons Resound”), like the previous one, consists of iambic octosyllabic couplets. In comparison to the song “Baglamades” presented earlier, the instrumentation of “The Prisons Resound” by Vamvakaris is very different. In the recording, bouzouki and baglamas can be heard, which until circa 1933 were mainly popular in the close communities of low-life haunts and prisons. Due to its instrumentation, this song, in contrast to “Baglamades”, is an example of the Piraeus style of rebetiko. At the same time, it uses the old Asia Minor melody that appears in the song «Τα ούλα σου» (“Everything in You”), which was recorded in 1919 in New York (disc number: Columbia USA E-5193, sung by Marika Papangika).¹⁵⁸ The same melody was also used in the songs «Μια χήρα μεσ’ την Κοκκινιά» (“A Widow from Kokkinia”) reinterpreted by Kostas Tzovenos in 1935 (disc number: Odeon GA-1807, sung by Marika Kanaropoulou) and «Με κάποια στα Πετράλωνα» (“With some Woman in Petralona”) reinterpreted by Iakovos Montanaris in 1936 (disc number: His Master’s Voice AO 2327, sung by Stellakis Perpiniadis). This demonstrates a typical characteristic of rebetiko, namely that many rebetiko exponents in their composition used old melodies of the period of the oral tradition of rebetiko before commercialisation began. It refers not only to the musical aspects, but also to the song text. For instance, some couplets of the song “The Prisons Resound” already appeared in the nineteenth century.¹⁵⁹

The central topic of the song “The Prisons Resound” by Vamvakaris is imprisonment, which is a frequent theme in rebetiko songs. In this song, five prisons are mentioned: the prison in Bourtzi castle located in Nafplio harbour; Yedi Kule, or Heptapyrgion prison in Thessaloniki; Syngrou prison in Athens; the buildings of the Athens Military Prison on the Vouliagmenis Avenue, and the prison in the Rio Castle near Patras.¹⁶⁰ Considering the first and the last couplet, it is difficult to imagine the prison in Bourtzi castle near Nafplio and Yedi Kule as well as the prison in Rio castle near Patras and Yedi Kule to resound. The rebetiko exponent Takis Binis suggested that Markos Vamvakaris changed the names of prisons in the couplets that circulated before in oral tradition. Thus, for example, he suggests that instead of Yedi Kule in Thessaloniki, originally Its Kale (*Ιτς Καλέ*) was mentioned. Its Kale, or Akronafplia, which was also used as a prison, is located opposite of Bourtzi castle in Nafplio harbour, which is also a former prison, such that both prisons could resound. Interestingly, the version of the song of 1935 includes the line *αντιλαλούν δύο σύρματα* (“two wires resound”) in the second couplet, whereas in the version of 1960, sung by Grigoris Bithikotsis, it is replaced by *αντιλαλούν δύο σήμαντρα*, which can be translated as “the two prisons’ bells resound”. The line in Bithikotsis’ song version is reminiscent of a church bell chime. It is at the same time possible, that Markos Vamvakaris used the verb *αντιλαλώ* (“resound, re-echo”) figuratively: namely, that the prisoners stayed in connection with each other even if they were in different prisons. This connection can be two-fold: they could be sentenced for the same reason (for instance, political prisoners) or they could still stay

158 The record of the song «Τα ούλα σου» (“Everything in You”) sung by Marika Papangika in 1919 can be found at <https://www.vmrebetiko.gr/item/?id=4179> (accessed January 11, 2022). The second known recording of this song appeared in 1920 also in New York (disc number: Panhellenion 5020, sung by Mme. Coula), although the song text differs from its first recording. The recording of 1920 with Mme. Coula can be found at <https://www.youtube.com/watch?v=PTQzJLz0tnQ> (accessed January 11, 2022).

159 More about this melody can be found at <https://www.vmrebetiko.gr/item/?id=4251> (accessed January 11, 2022).

160 Yedi Kule and the Syngrou prisons are frequently mentioned in the context of rebetiko.

in contact with each other and with the outside world. The word *σύρμα*, 'wire', besides the meaning as a "connecting wire between the prison and the outside world", in Greek argot can also stand for prison. Then, the line can be translated just as "two prisons resound".¹⁶¹ Moreover, the word *σύρμα* may also define hashish in Greek argot. Thus, it is possible that Vamvakaris refers to a sound produced when smoking a hookah, and hashish smoking was widespread in the Greek prisons of that time. If we interpret the word *σύρμα* as *hashish*, then the line *αντιλαλούν δύο σύρματα* may stand for 'two hookahs resound'. Such an interpretation is strengthened by the explicit use of the word *σύρμα* meaning 'black (hash) of high quality' in five songs of the *Rebetiko Corpus* including two other songs by Vamvakaris. For instance, in the song «Ο χαρμάνης» ("The Stoner")¹⁶² Vamvakaris refers to hashish using the phrase *σύρμα μαύρο*, 'black of high quality'. When it comes to the sound of a hookah in rebetiko, the song "Five Years in Prison (The Sound of a Hookah Pipe)" by Vangelis Papazoglou even includes the sound of a hookah pipe in the studio recording.¹⁶³ In any case, the song "The Prisons Resound" talks about the connection between the prisons either as a connection by means of sound or other linking ground. Next, the protagonist begs his mother to come and cry for him, so that he will be released. The figure of the mother appears frequently in rebetiko songs, and while the beloved one is usually depicted as a cruel and heartless woman, the mother is shown as a kind person who can rescue the protagonist from his troubles.

The final song that will be presented is «Μποέμισσα» ("Bohemian Girl"), composed by Vasilis Tsitsanis. The recording was made in 1938 with Stellakis Perpiniadis and Stratos Pagioumtzis as singers (disc number: HMV AO-2463, Athens).

Μποέμισσα ξανθιά γαλανομάτα γόησσα, αφράτη, ζηλευτή, μ' άναψες καημούς πολύ μεγάλους, είναι μαρτύριο για μένα η ζωή.	Blonde bohemian girl with blue eyes, plump and enviable charmer, you have put me in great sorrows, my life is a torture
Ένας μαγνήτης είναι η ματιά σου, ζαλίζει, ξεμυαλίζει σαν κοιτάς, μα εμένα χρόνια τώρα βασανίζεις, με κοροϊδεύεις κι όλο λες πως μ' αγαπάς.	Your look is like a magnet, people get dizzy and lose their mind when you look at them, but you're torturing me for years now, you mock me and you're always saying that you love me.
Μες στη καρδιά χρυσό παλάτι σ' έχω, σκλάβο μ' έκανες, τι θα γινώ; Μποέμισσα, στο λέγω, δεν αντέχω, τα γαλανά σου μάτια πάντα λαχταρώ.	I have you as a golden palace in my heart, you made me a slave, what am I going to be? Bohemian girl, I'm telling you, I can't stand it I'm always longing for your blue eyes. ¹⁶⁴

The song "Bohemian Girl" is accompanied by bouzouki (Vasilis Tsitsanis). In this song, the appearance of a second voice is remarkable, which is not typical for earlier rebetika, but

161 See the meaning of the word *σύρμα* (wire) at http://georgakas.lit.auth.gr/dictionaries/index.php?option=com_chronoforms5&chronoform=ShowLima&limaID=15269 (accessed January 11, 2022).

162 The song «Ο χαρμάνης» ("The Stoner") composed and sung by Markos Vamvakaris was recorded in 1933 (disc number: Odeon GA-1674).

163 The song «Η φωνή του αργιλέ (Πέντε χρόνια δικασμένος)» ("Five Years in Prison (The Sound of a Hookah Pipe)") by Vangelis Papazoglou was recorded in 1935 (disc number: Columbia DG-6066).

164 The translation of the song «Μποέμισσα» ("Bohemian Girl") was taken almost unchanged from the following webpage <https://lyricstranslate.com/en/mpoemissa-μ.ποέμισσα-bohemian-girl.html> (accessed January 11, 2022).

becomes widespread at the end of the 1930s and even more so after World War II. The song consists of three quatrains with irregular metre. In general, in rebetika of the late 1930s and afterwards, one can observe more diverse metrical forms with couplets or quatrains including lines of different length (Gauntlett 1985a, 122, 125). Moreover, while early rebetika mostly present a desultory narrative, rebetiko songs from the 1930s onwards depict a more coherent narrative, as one can for example see in the song “Bohemian Girl”.

The censorship imposed in 1936 had influenced rebetika drastically, such that songs referring to the low life as well as typical vocabulary connected to the underworld disappeared almost completely from the song texts. Most rebetika of that period were love songs (Gauntlett 1985a, 105-107). This also stimulated the appearance of a new generation of bouzouki players, to whom also Tsitsanis belonged. Vasilis Tsitsanis, who in general used the term *rebetiko* in reference to his works reluctantly, preferring the term *laiko* instead, even called his compositions *kantades*, meaning serenade-like love songs (Tragaki 2009, 61-62). The song «Μποέμισσα» (“Bohemian Girl”) also has love as its central topic. Its protagonist is delighted by the beauty of his beloved one, whom he calls a bohemian girl.¹⁶⁵ At the same time, as it is typical for erotic rebetika, the beloved woman is heartless, and the protagonist has to suffer from unrequited love.

As already mentioned at the beginning of this section, these three songs cannot cover the whole spectrum of themes and characteristics of rebetika. Besides rebetika depicting hashish consumption, prison life and sufferings of a protagonist from unrequited love, one can find, for example, numerous rebetika referring to different historical events (or phenomena)¹⁶⁶, glendi (party)¹⁶⁷, gambling¹⁶⁸ and exotic worlds¹⁶⁹, as well as songs that are dedicated to particular people, including exponents of rebetiko.¹⁷⁰ Further analyses of rebetiko song lyrics will be presented in the next two chapters, in the context of the corpus that was created in the course of the work on this thesis.

165 In the context of rebetika the word *μποέμισσα* (English ‘bohemian girl’) is synonymous to *μάγκισσα* (English ‘guttersnipe, brave girl’) or *μόρτισσα* (English ‘bum’).

166 See, for instance, such songs as «Κακούργα πεθερά» (“Villain Mother-in-law”) by Iakovos Mondanaris (disc number: Parlophone B-21598, recorded in 1931), «Τον Βενιζέλο χάσαμε» (“We Lost Venizelos”) by Stavros Pandelidis (disc number: Columbia DG-6209, recorded in 1936), «Οι σαλταδόροι (Θα σαλτάρω)» (“The Leapers (I’ll Leap)”) by Giorgos Katsaros/Michalis Genitsaris (disc number: Standard F-9025, first recorded in 1947) and «Ρεμπέτικη Κυβέρνηση» (“Rebetiko Government”) by Marinos Gavriil (Vinyl, 1976)

167 See, for instance, the song «Πέντε Έλληνες στον Άδη» (“Five Greeks in Hell”) by Giannis Papagioannou (disc number: His Master Voice AO-2739, recorded in 1947)

168 See, for instance, the song «Με ζουρνάδες με νταούλια (Ο μπακαράς)» (“With Zurnas and Davuls (Baccarat Player)”) by Panagiotis Toundas (disc number: Columbia DG-498, first recorded in 1934)

169 See, for instance, the song «Αραπίνες και πασάδες» (“Arab Women and Pashas”) by Giannis Papagioannou (disc numbers: Nina N/609, recorded in 1953)

170 See, for instance, such songs as «Ροζικά» (“Rozika (Sweet Roza)”) by Andonis Dalgas, which is devoted to Roza Eskenazi, a prominent rebetiko singer (disc numbers: Odeon GA-1691, HMV AO-2078, recorded in 1933), and «Μάρκος, ο Συριανός» (“Markos from Syros”) by Markos Vamvakaris devoted to Vamvakaris himself (disc number: His Masters Voice AO-2065, recorded in 1933)

In this chapter, the *Rebetiko Corpus* will be introduced. First, a range of digital and analog sources of rebetiko song lyrics and metadata is presented. The main focus is on the process of the creation of the corpus, and specifically on the challenges that had to be solved in this process. Finally, key data on the corpus is presented and its properties are discussed.

3.1 EXISTING REBETIKO SONG LYRICS AND METADATA SOURCES

Before going into the details of the *Rebetiko Corpus*, this section discusses sources with rebetiko song lyrics and metadata that are already available. Two main types of sources of rebetiko song texts and metadata can be distinguished: print sources and web-based sources. An additional source of rebetiko lyrics are the actual recordings, but these are not as easy to work with.

First, some of the main print sources will be presented. The landscape of print sources containing rebetiko song texts and metadata includes anthologies of songs and (auto-)biographies of rebetiko exponents, newspaper/ magazine articles and a range of scholarly and popular texts on rebetika in general that include collections of rebetiko songs of varying sizes. The two most sizeable collections of rebetiko songs are the anthologies by Elias Petropoulos and by Tasos Schorelis.

The anthology *Rembetika Tragoudia* by Elias Petropoulos was first published in 1968, at a time of strict censorship in Greece. His book was banned in the same year and Petropoulos was imprisoned for five months ‘for pornography’ (Tragaki 2009, 109). Later on an extended version of the book has been published, which was illustrated by Alekos Fasianos, and includes over 1,400 song texts divided into twenty categories such as love songs, hashish songs, songs of the army and the war, and songs about prison life (Petropoulos 1991). Petropoulos is considered to be the first to classify rebetiko songs according to their lyrical themes. His anthology also includes a brief history of rebetiko, an introduction to main properties of this music genre as well as numerous texts of different authors on rebetiko and its social context. It also contains a large collection of photographs related to rebetiko. Many songs are provided with information about their rhythm, the composer, the year of recording and their origins. For some songs, also information about the lyricist, singer, dromos (makam) as well as further recordings or reinterpretations of a particular song is included. Metadata such as instrumentation are very rare, and the information about the disc number is missing. Remarkably, although Petropoulos considered that rebetiko had vanished in the 1950s (Gauntlett 1985a, 9), in his anthology he has included songs that were recorded in the 1960s and the 1970s (see, for instance, (Petropoulos 1991, 74-76)). In general, the anthology of Petropoulos contains not only songs that appeared as recordings, but also those that appeared in other media. Thus, his anthology contains songs from the end of the nineteenth century, including couplets of koutsavakides, to the 1970s (Petropoulos 1991, 46, 112).

The anthology by Tasos Schorelis called *Rebetiki Anthologia* was first published in 1977. It contains about 480 rebetiko songs composed until 1952, including songs of the period of oral tradition, amanedes, as well as numerous song lyrics of particular rebetiko composers, which are organised in separate sections (Schorelis 1988). These separate sections are provided with short introductory texts, like the biography of a particular rebetiko composer or the history of amanedes. Moreover, the anthology of Schorelis contains a rebetiko glossary, numerous photographs of rebetiko exponents as well as sheet music for some of the songs. Regarding the metadata, most song lyrics are provided only with the name of the composer. Some songs also include the year in which a recording was made, as well as the name of the singer. Further information about existing records such as disc number, lyricist or instrumentation is not available.

Some of the most valuable web-based sources for rebetiko lyrics and metadata for songs were already mentioned in the previous chapter. For instance, rebetiko song lyrics and metadata are available on the following websites: *rebetiko.sealabs.net*¹, *Εικονικό Μουσείο Αρχείου Κουνάδη (Kounadis Archive Virtual Museum)*², *Ρεμπέτικο φόρουμ (Rebetiko Forum)*³, *kithara.to*⁴ (which provides information about the chords as well as the tonality of the song in question and a corresponding recording), *stixoi.info*⁵, *Art and the City*⁶, *Greeklyrics*⁷, *Rebet*⁸, *Discogs*⁹, *Music Archive of the library of Technological Educational Institute of Epirus*¹⁰ as well as *YouTube*¹¹ and *elTube*.¹² Two of the most important sources, the *Kounadis Archive Virtual Museum* and the webpage *rebetiko.sealabs.net*, will be presented in more detail.

The project of *rebetiko.sealabs.net* includes a wiki-based and collaboratively edited collection of over 3,500 songs that were recorded between 1919 and 1992 (mostly rebetika and dimotika). This collection contains not only the song lyrics, but also short biographies of rebetiko exponents and metadata for songs such as release date, names of the artists involved and the disc record number. Besides the song collection hosted on MediaWiki, *rebetiko.sealabs.net* also includes a forum with discussions about anything related to rebetiko. Moreover, the collection of songs hosted at *rebetiko.sealabs.net* is published under a license that allows further re-publication of this data for non-commercial purposes with reference to this license.¹³

The Kounadis Archive Virtual Museum was created in 2019 and has as its main goal the digitisation of more than 6,000 analog objects of the Kounadis Archive that have been collected since the 1960s (Chatzigrigoriou 2019). At the time of publication of this work, the virtual archive consists of 2,820 digitised 78 rpm records of almost exclusively Greek songs from the time span between 1904 and 1952 (mainly rebetika). Many of these recordings are

1 Available at <https://rebetiko.sealabs.net> (accessed January 19, 2022).

2 Available at <https://www.vmrebetiko.gr> (accessed January 19, 2022)

3 Available at <https://rembetiko.gr> (accessed January 19, 2022)

4 Available at <https://kithara.to> (accessed January 19, 2022)

5 Available at <https://www.stixoi.info> (accessed January 19, 2022)

6 Available at <https://www.artsandthecity.gr> (accessed January 19, 2022).

7 Available at <https://www.greeklyrics.gr> (accessed January 19, 2022).

8 Available at <https://www.rebet.gr> (accessed January 19, 2022)

9 Available at <https://www.discogs.com> (accessed January 19, 2022).

10 Available at <https://libsearch.teiep.gr> (accessed January 19, 2022)

11 Available at <https://www.youtube.com> (accessed January 19, 2022)

12 Available at <https://www.eltube.gr> (accessed January 19, 2022)

13 The lyrics data are published by *rebetiko.sealabs.net* under the license CC BY-NC-SA 4.0, which means that the data can be used for corpus creation and this corpus can be made freely available in the future (under specification CC BY-NC-SA 4.0).

rare and date back to the first two decades of the twentieth century, which makes the *Virtual Kounadis Archive* an even more valuable source. It also includes sheet music, concert programmes and post cards referring mainly to rebetiko, 3D models of music instruments, sound recording and reproduction machines, and photographs of rebetiko exponents. Furthermore, it includes a virtual tour through the Kounadis archive, games and an educational portal which at the moment includes the first pilot lecture and is aimed at presenting the history of Modern Greece through the lens of rebetiko. The recordings are provided with metadata in a format proposed by the Library of Congress (Chatzigrigoriou 2019). To be more precise, this includes information such as an image of the label of the recording, the type of the original recording, names of the artists involved, date and place of a recording, the language of a song and its length, the disc record number, the source and the license. Many records also contain information about instrumentation and rhythm. Moreover, numerous songs include extended commentaries on the history of a particular recording. The song lyrics are not provided. The user interface of the website of the Virtual Archive is very user-friendly. The items included in the archive are provided with tags, so that the data is easily searchable.

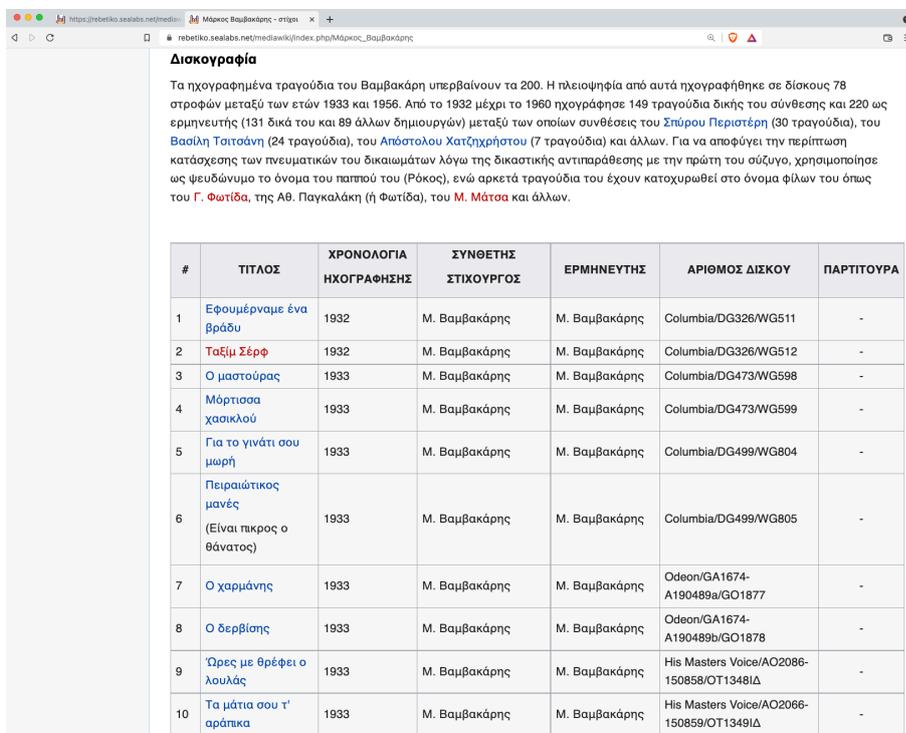
3.2 CREATION OF THE REBETIKO CORPUS

One of the main goals of the present work was not only to create a corpus of rebetiko song lyrics and use it for some case studies, but to publish it and make it freely available for further use. For this reason, the use of existing anthologies of rebetiko songs as source data is problematic, as the copyright has not expired. Moreover, these anthologies are not digitised, so the digitisation would cause additional complications and be quite time-consuming. Thus, the preference was given to a data source that already contains digitised data published under a license that allows future publication of this data for non-commercial purposes. The website *rebetiko.sealabs.net*, which has been presented above, fulfils this property and offers lyrics and metadata in relatively high quality under the license CC BY-NC-SA 4.0 (both the song lyrics and the metadata are in Modern Greek). It therefore became the main data source for the corpus. It is important to stress that the data available at *rebetiko.sealabs.net* is based on songs that were recorded, which means that it may miss some early rebetika or rebetika of the period of censorship that have never been recorded. Due to the fact that the *Rebetiko Corpus* is based on the collection of songs available at *rebetiko.sealabs.net*, these songs are thus also not included in the corpus. Further sources of rebetiko song lyrics, which are mentioned in Section 3.1, were used to verify the raw data as well as to add missing metadata. Next, the process of data extraction from *rebetiko.sealabs.net* will be presented.

3.2.1 Automatic Data Extraction

Since a fully manual preparation of a large corpus is very time-consuming, an attempt was made to at least partially automate the creation of the corpus. Rather than downloading each song manually, the programming language Python was used to write scripts that accomplish this task automatically. These scripts read the text of the website and parse it in order to then store certain data in a structured way.

The data is extracted from the source code of the individual pages of the rebetiko exponents at *rebetiko.sealabs.net*, which can be accessed via a special export function.¹⁴ To illustrate this, Figure 3.1 shows a part of the table of the works by Vamvakaris, and Figure 3.2 shows the corresponding source code, which contains all the information we would like to extract.

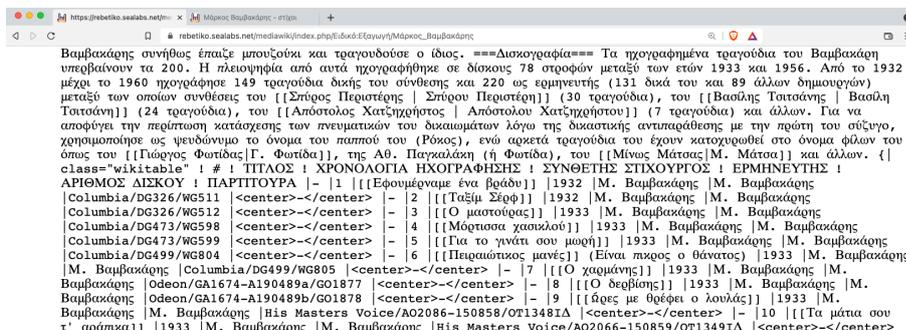


Δισκογραφία

Τα ηχογραφημένα τραγούδια του Βαμβακάρη υπερβαίνουν τα 200. Η πλειοψηφία από αυτά ηχογραφήθηκε σε δίσκους 78 στροφών μεταξύ των ετών 1933 και 1956. Από το 1932 μέχρι το 1960 ηχογράφησε 149 τραγούδια δικής του σύνθεσης και 220 ως ερμηνευτής (131 δικά του και 89 άλλων δημιουργών) μεταξύ των οποίων συνθέσεις του **Σπύρου Περιστέρη** (30 τραγούδια), του **Βασιλή Ταϊτσάνη** (24 τραγούδια), του **Απόστολου Χατζηχρήστου** (7 τραγούδια) και άλλων. Για να αποφύγει την περίπτωση κατάσχεσης των πνευματικών του δικαιωμάτων λόγω της δικαστικής αντιπαράθεσης με την πρώτη του σύζυγο, χρησιμοποίησε ως ψευδώνυμο το όνομα του παππού του (Ρόκος), ενώ αρκετά τραγούδια του έχουν κατοχυρωθεί στο όνομα φίλων του όπως του **Γ. Φωτίδα**, της **Αθ. Παγκαλάκη** (ή **Φωτίδα**), του **Μ. Μάτσα** και άλλων.

#	ΤΙΤΛΟΣ	ΧΡΟΝΟΛΟΓΙΑ ΗΧΟΓΡΑΦΗΣΗΣ	ΣΥΝΘΕΤΗΣ ΣΤΙΧΟΥΡΓΟΣ	ΕΡΜΗΝΕΥΤΗΣ	ΑΡΙΘΜΟΣ ΔΙΣΚΟΥ	ΠΑΡΤΙΤΟΥΡΑ
1	Εφουμέριναμ ένα βράδυ	1932	Μ. Βαμβακάρης	Μ. Βαμβακάρης	Columbia/DG326/WG511	-
2	Ταξιμ Σέρφ	1932	Μ. Βαμβακάρης	Μ. Βαμβακάρης	Columbia/DG326/WG512	-
3	Ο μαστούρας	1933	Μ. Βαμβακάρης	Μ. Βαμβακάρης	Columbia/DG473/WG598	-
4	Μόρτισσα χασικλού	1933	Μ. Βαμβακάρης	Μ. Βαμβακάρης	Columbia/DG473/WG599	-
5	Για το γινάτι σου μωρή	1933	Μ. Βαμβακάρης	Μ. Βαμβακάρης	Columbia/DG499/WG804	-
6	Πειραιώτικος μανές (Είναι πικρός ο θάνατος)	1933	Μ. Βαμβακάρης	Μ. Βαμβακάρης	Columbia/DG499/WG805	-
7	Ο χαρμάνης	1933	Μ. Βαμβακάρης	Μ. Βαμβακάρης	Odeon/GA1674-A190489a/GO1877	-
8	Ο δερβίσης	1933	Μ. Βαμβακάρης	Μ. Βαμβακάρης	Odeon/GA1674-A190489b/GO1878	-
9	Ώρες με θρέφει ο λουλάς	1933	Μ. Βαμβακάρης	Μ. Βαμβακάρης	His Masters Voice/AO2086-150858/OT13481Δ	-
10	Τα μάτια σου τ' αράπικα	1933	Μ. Βαμβακάρης	Μ. Βαμβακάρης	His Masters Voice/AO2066-150859/OT13491Δ	-

Figure 3.1: Part of the List of Works on the Page about Vamvakaris at *rebetiko.sealabs.net*



```

Βαμβακάρης συνήθως έπαιζε μπουζούκι και τραγουδούσε ο ίδιος. ===Δισκογραφία=== Τα ηχογραφημένα τραγούδια του Βαμβακάρη υπερβαίνουν τα 200. Η πλειοψηφία από αυτά ηχογραφήθηκε σε δίσκους 78 στροφών μεταξύ των ετών 1933 και 1956. Από το 1932 μέχρι το 1960 ηχογράφησε 149 τραγούδια δικής του σύνθεσης και 220 ως ερμηνευτής (131 δικά του και 89 άλλων δημιουργών) μεταξύ των οποίων συνθέσεις του [[Σπύρος Περιστέρη|Σπύρου Περιστέρη]] (30 τραγούδια), του [[Βασιλής Ταϊτσάνης|Βασιλή Ταϊτσάνη]] (24 τραγούδια), του [[Απόστολος Χατζηχρήστου|Απόστολου Χατζηχρήστου]] (7 τραγούδια) και άλλων. Για να αποφύγει την περίπτωση κατάσχεσης των πνευματικών του δικαιωμάτων λόγω της δικαστικής αντιπαράθεσης με την πρώτη του σύζυγο, χρησιμοποίησε ως ψευδώνυμο το όνομα του παππού του (Ρόκος), ενώ αρκετά τραγούδια του έχουν κατοχυρωθεί στο όνομα φίλων του όπως του [[Γιώργος Φωτίδας|Γ. Φωτίδα]], της Αθ. Παγκαλάκη (ή Φωτίδα), του [[Μίνος Μάτσα|Μ. Μάτσα]] και άλλων. {| class="wikitable" | # | ΤΙΤΛΟΣ | ΧΡΟΝΟΛΟΓΙΑ ΗΧΟΓΡΑΦΗΣΗΣ | ΣΥΝΘΕΤΗΣ ΣΤΙΧΟΥΡΓΟΣ | ΕΡΜΗΝΕΥΤΗΣ | ΑΡΙΘΜΟΣ ΔΙΣΚΟΥ | ΠΑΡΤΙΤΟΥΡΑ | - | 1 | [[Εφουμέριναμ ένα βράδυ]] | 1932 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | Columbia/DG326/WG511 | <center>-</center> | - | 2 | [[Ταξιμ Σέρφ]] | 1932 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | Columbia/DG326/WG512 | <center>-</center> | - | 3 | [[Ο μαστούρας]] | 1933 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | Columbia/DG473/WG598 | <center>-</center> | - | 4 | [[Μόρτισσα χασικλού]] | 1933 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | Columbia/DG473/WG599 | <center>-</center> | - | 5 | [[Για το γινάτι σου μωρή]] | 1933 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | Columbia/DG499/WG804 | <center>-</center> | - | 6 | [[Πειραιώτικος μανές]] (Είναι πικρός ο θάνατος) | 1933 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | Columbia/DG499/WG805 | <center>-</center> | - | 7 | [[Ο χαρμάνης]] | 1933 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | Odeon/GA1674-A190489a/GO1877 | <center>-</center> | - | 8 | [[Ο δερβίσης]] | 1933 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | Odeon/GA1674-A190489b/GO1878 | <center>-</center> | - | 9 | [[Ώρες με θρέφει ο λουλάς]] | 1933 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | His Masters Voice/AO2086-150858/OT13481Δ | <center>-</center> | - | 10 | [[Τα μάτια σου τ' αράπικα]] | 1933 | Μ. Βαμβακάρης | Μ. Βαμβακάρης | His Masters Voice/AO2066-150859/OT13491Δ | <center>-</center>

```

Figure 3.2: Source Code Corresponding to the Excerpt Shown in Figure 3.1 in the Wikimedia markup language (wikitext)

To give an example on how the data extraction works in practice, we will have a brief look at the code that was used to retrieve and store the works of Markos Vamvakaris. Listings 3.1 and 3.2 together present the code that was used to achieve this. Listing 3.1 is a general parser for the pages containing the lyrics and can be used for all artists on the website, while

¹⁴ Available at <https://www.mediawiki.org/wiki/Special:Export> (accessed January 22, 2022)

Listing 3.2 reads the overview tables or lists of works and has to be adapted for each artist.¹⁵ Retrieving the songs and metadata, including the lyrics, is a three part process. First, the list of works is parsed, which is specific for each artist and in this case happens in the function *parse_overview_page* (see Listing 3.2, lines 9 to 37). This function returns a list of entries, which needs to be further processed to extract the metadata for each song, e.g. the title, release year, disc number, and also the link to the page containing the lyrics. This is done in the function *parse_overview_page_entry* (see Listing 3.2, lines 39 to 49). Next, the source code of the corresponding lyrics pages is retrieved and the lyrics are extracted, which is done in the function *load_raw_lyrics* (see Listing 3.1, lines 13 to 30). Finally, some cleanup on the lyrics is performed, e.g. the title, some comments, and superfluous whitespace characters are removed from the lyrics. The result of this process is a collection of 178 songs by Vamvakaris with attached metadata and lyrics. These data were stored in a JSON format (see Listing 3.3 for an exemplary entry in this format).

3.2.2 Challenges of the Corpus Creation

The creation of the rebetiko corpus posed some challenges concerning the data preparation, which will be discussed in more detail in this section.

The wiki-based project *rebetiko.sealabs.net* offers a sizeable collection of rebetiko data. Nonetheless, as mentioned previously, for some songs metadata and/or song lyrics are missing or contain mistakes. For this reason, after extracting the raw data from *rebetiko.sealabs.net*, a verification process was performed. Since directly working with the JSON format, in which the data was stored, is cumbersome for humans, the database management tool *Navicat* was used to simplify this process. *Navicat* is a user-friendly interface for databases and database management systems and provides convenient ways for working with SQL queries and doing data modeling tasks. In the scope of the creation of the rebetiko corpus, it was mainly used for data editing as well as for writing and performing SQL queries in order to e.g. find duplicates. Figure 3.3 gives an impression on how the data is presented in this tool. It is neatly arranged in the form of a table, which is more suitable for the task of data verification than the pure JSON format. The table header illustrates which metadata are included in the corpus: namely, the song index, song name, release year, composer, performance metadata (information about orchestration, music performers and when given also conductor), lyricist, singer, disc record number, location of recording and the dance (rhythm) to a particular song. The corpus also includes metadata such as the source file name and the source page name, which were not included in the Figure 3.3 due to the lack of space. The *Navicat* table also includes the song lyrics that can be edited in an extra editor field, which can be also seen in Figure 3.3.

Next, we will look in more detail at the information which had to be verified. The lyrics and the metadata were verified manually, using a range of sources listed in Section 3.1.

As already mentioned earlier, the data on *rebetiko.sealabs.net* is based on songs which were recorded. This means that the publication date refers to the release year of a recording. Thus, a song might have already appeared at an earlier time. This fact should be kept in mind when making generalisations concerning the song lyrics: for instance, the analysis of thematic change over time or the prevalence of a particular topic in a particular time

¹⁵ The *rebetiko.sealabs.net* wiki-based collaboratively edited collection does not present data in a uniform format, thus the parser has to be adapted for each artist.

```

1 class Parser:
2     def __init__(self):
3         self.special_export_link = "http://rebetiko.sealabs.net/mediawiki/index.php/%CE%95%CE%B9%CE%B4%CE%B9%CE%BA%CF%8C
         :%CE%95%CE%BE%CE%B1%CE%B3%CF%89%CE%B3%CE%AE/"
4         self.metadata = dict()
5         self.songs = dict()
6
7         self.lyrics_raw = None
8
9     def process(self):
10        self.parse_overview_page()
11        self.load_raw_lyrics()
12        self.process_lyrics()
13
14    def load_raw_lyrics(self):
15        self.lyrics_raw = dict()
16
17        for key in tqdm(self.songs):
18
19            song = self.songs[key]
20            lyrics_page = song["page_name"]
21            lyrics_page = lyrics_page.replace("&quot;", "\"\"")
22            lyrics_page = lyrics_page.replace("&amp;", "&")
23
24            link = self.special_export_link + urllib.parse.quote(lyrics_page)
25            lyrics_page = urllib.request.urlopen(
26                link
27            )
28            lyrics_page_content = lyrics_page.read().decode("utf-8")
29
30            self.lyrics_raw[key] = lyrics_page_content
31
32    def process_lyrics(self):
33        for key in tqdm(self.lyrics_raw):
34
35            # if page does not contain any lyrics, set it to ""
36            if (
37                self.lyrics_raw[key].find("<page>") == -1
38                or self.lyrics_raw[key].find('<text xml:space="preserve" bytes="0" />')
39                != -1
40            ):
41                self.songs[key]["lyrics"] = ""
42                continue
43
44            lyrics = self.lyrics_raw[key]
45
46            # find basic start and stop, refine later
47            index_start = re.search(
48                r"<text xml:space=\"preserve\" bytes=\".*\">", lyrics
49            ).end()
50            index_stop = lyrics.find("[", index_start)
51
52            lyrics = lyrics[index_start:index_stop]
53            lyrics = html.unescape(lyrics)
54
55            # remove lyrics headline
56            lyrics_word = re.search(r"Στίχοι====", lyrics)
57            if lyrics_word is not None:
58                lyrics = lyrics[lyrics_word.end() :]
59
60            # remove melody
61            pattern = re.search(r"Μελωδία====", lyrics)
62            if pattern is not None:
63                lyrics = lyrics[: pattern.start()]
64
65            # remove text tag end
66            pattern = re.search(r"</text>", lyrics)
67            if pattern is not None:
68                lyrics = lyrics[: pattern.start()]
69
70            # strip leading newlines
71            lyrics = lyrics.strip("\n")
72
73            lyrics = lyrics.replace("\n \n", "\n\n")
74            lyrics = lyrics.replace("\n\n", "\n")
75
76            self.songs[key]["lyrics"] = lyrics

```

Listing 3.1: General Parser for Pages Containing Song Lyrics

```

1 class VamvakarisParser(Parser):
2     def __init__(self, identifier, identifier_english=None):
3         super().__init__()
4
5         self.metadata["id"] = identifier
6         self.metadata["id_encoded"] = urllib.parse.quote(identifier)
7         self.metadata["name_english"] = identifier_english
8
9     def parse_overview_page(self):
10        overview_page = urllib.request.urlopen(
11            self.special_export_link + self.metadata["id_encoded"]
12        )
13        content = overview_page.read().decode("utf-8")
14        index_start = content.find("{| class=&quot;wikitable&quot;}") + 1
15        index_stop = content.find("<", index_start)
16
17        content = content[index_start:index_stop]
18
19        counter = 1
20        index = 0
21
22        while counter < 500:
23            old_index = index
24            index = content.find("|" + str(counter) + "\n", old_index)
25            if index == -1:
26                index = old_index
27                counter += 1
28                continue
29
30            index += 1
31            index_stop = content.find("|-\n", index)
32
33            entry = content[index:index_stop]
34
35            self.parse_overview_page_entry(entry)
36
37            counter += 1
38
39    def parse_overview_page_entry(self, entry):
40        entry_parts = entry.split("\n")
41
42        identifier = int(entry_parts[0])
43        self.songs[identifier] = dict()
44
45        self.songs[identifier]["name"] = entry_parts[1]
46        self.songs[identifier]["name"] = re.search(
47            "\[([.*]*)\]", self.songs[identifier]["name"]
48        )
49        self.songs[identifier]["name"] = self.songs[identifier]["name"].group(0)[2:-2]
50
51        if self.songs[identifier]["name"].find("|") != -1:
52            self.songs[identifier]["name"] = self.songs[identifier]["name"][
53                0 : self.songs[identifier]["name"].find("|")
54            ]
55
56        self.songs[identifier]["page_name"] = self.songs[identifier]["name"]
57
58        try:
59            year = parse(entry_parts[2], fuzzy=True).year
60        except ParserError:
61            year = -1
62
63        self.songs[identifier]["year"] = year
64
65        self.songs[identifier]["composer_lyrics"] = entry_parts[3]
66        self.songs[identifier]["interpret"] = entry_parts[4]
67        self.songs[identifier]["disc_number"] = entry_parts[5]

```

Listing 3.2: Specific Parser for the Entry for Vamvakaris

```

1 {
2     "1": {
3         "name": "Εφουμέρναμε ένα βράδυ",
4         "page_name": "Εφουμέρναμε ένα βράδυ",
5         "year": 1932,
6         "composer_lyrics": "Μ. Βαμβακάρης",
7         "interpret": "Μ. Βαμβακάρης",
8         "disc_number": "Columbia/DG326/WG511",
9         "lyrics": "Εφουμέρναμ' ένα βράδυ\αργιλέν, σπαχάνη, μαύρη\Δίχων να' χουμε στην πόρτα \χωρίς να χουμε' στην
πόρτα χ2\n \τσιλιαδόρουσν όπως πρώτα.\nΚιν έρχουνται δύο πολιτμάνοι\καιν δε βρίσκουσε ντουμάνι\Ζούλαν όλοι οι
αργιλέδες χ3\ψυλαχτείτν' από τους τζέδες.\n\Στάσου, πολιτμάν, λεβέντη\Κιν ασ τον αργιλέ να καίει\Nan φουμάρει
το τουρκάκι χ3\πούν να' ψίνο δερβισάκι\n\Μαύρον ψέρνει από τη Σύμρνη\καιν καλάμι απ' τ' Αϊδίνι χ2\καιν χαρά
στον που την πίνει."
10     },

```

Listing 3.3: A Song by Vamvakaris in JSON Format (Before Data Correction)

index	name	year	composer	performance_metadata	composer_lyrics	singer	disc_number	location	dance	lyrics
1223	Εκατό δραχμές τη μέρα	1948	Κώστας Σκαρβέλης	(NULL)	(NULL)	Κώστας Μακρής	Balkan Αμερικής BAL-21	(NULL)	(NULL)	Ωχ, αμάν, αμάν, εκατό δραχμές τη
1224	Εκατό δραχμές τη μέρα	1948	Κώστας Σκαρβέλης	(NULL)	(NULL)	Κώστας Μακρής	Melody Ελλάδος ME-21	(NULL)	(NULL)	Ωχ, αμάν, αμάν, εκατό δραχμές τη
1225	Μες την Αθήνα τριγυρνά	1950	Κώστας Σκαρβέλης	(NULL)	(NULL)	Μάρκος Μέλκων	Melody Ελλάδος ME-192	(NULL)	(NULL)	Μες στην Αθήνα τριγυρνάς
1226	Καστανή μικρούλα	1937	Κώστας Σκαρβέλης	(NULL)	(NULL)	Κώστας Σκαρβέλης	Πύργος Κάδουρας	(NULL)	(NULL)	Χασάπικο Πατή μικρούλα καστανή πάντα με τ
1229	Ο φύλλο	1932	Σταύρος Παντελίδης	Λαϊκή ορχήστρα με δύο	Σταύρος Παντελίδης	Ρίτα Αμπατζή	Columbia DG 277	Αθήνα	Καμηλέρι	Να ήμουν φύλλο, μάτια μου, αμάν
1234	Γαργαλιστικός χορός	1933	Σταύρος Παντελίδης	(NULL)	(NULL)	Ρίτα Αμπατζή	Columbia DG 474	Αθήνα	Ζεϊμπέκι	Αμάν τι γαργαλιστικός που είναι τι
1237	Ζωή είν' αυτή ζωίτσα μ	1933	Αντώνης Διαμαντίδης	Ορχήστρα με κιθάρα, οι	(NULL)	Ρίτα Αμπατζή	Columbia DG 448	Αθήνα	Ζεϊμπέκι	Τη δύστυχη αυτή χρονιά,
1239	Θα αφήσω γένια	1933	Άγνωστος	Κλαρίνο (Κώστας Γιαού)	(NULL)	Ρίτα Αμπατζή	Odeon Ελλάδος GA 1635	(NULL)	(NULL)	Άιντε θ' αφήσω γέ-, βρε γένια και
1240	Καλλιόπκι	1939	Γρηγόρης Ασιτικής	Λαϊκή ορχήστρα	Γρηγόρης Ασιτικής	Ρίτα Αμπατζή	HMV AO-2090	Αθήνα	Συρτός	Καλλιόπκι με πευράδες,
1241	Καλλιόπκι	1933	Παναγιώτης Τούντας	Λαϊκή ορχήστρα	Παναγιώτης Τούντας	Ρίτα Αμπατζή	Columbia 7194-F	Αθήνα	Ζεϊμπέκι	Πάλι μεθυσμένος είμαι για να σέν
1242	Κατεργάρα	1933	Γιάννης Δραγάνσης (Γ	Λαϊκή ορχήστρα με βιο	Γιάννης Δραγάνσης (Γ	Ρίτα Αμπατζή	Columbia DG 362	Αθήνα	Κιοτσέκι	Κατεργάρα, μη γυρεύεις, να με β
1243	Κατινάκι μου για σένα	1933	Κώστας Καρίπης	(NULL)	(NULL)	Ρίτα Αμπατζή	Odeon Ελλάδος GA-1691	(NULL)	(NULL)	Κατινάκι μου για σένα, με σκοτώσ
1247	Μες του Ζαμπίκου τον τ	1933	Κώστας Τζόβενος	Ορχήστρα (κιθάρες), Δι	Κώστας Τζόβενος	Ρόζα Εσκενάζυ	Columbia Ελλάδος DG-27	Αθήνα	Ζεϊμπέκι	Μες του Τσαμπίκου τον τεκέ, αμέ
1248	Νέα μερακλού	1933	Κώστας Τζόβενος	(NULL)	(NULL)	Ρίτα Αμπατζή	Parlophone B-21731	(NULL)	(NULL)	Είμαι μερακλού τσαχπία
1249	Ο ασίκης	1933	Σταύρος Παντελίδης	Λαϊκή ορχήστρα	Σταύρος Παντελίδης	Ρίτα Αμπατζή	His Master's Voice AO-20	Αθήνα	Τσιφτετέ	Είσαι ασίκης και ντερβίσσης
1251	Παιχνιδάκι	1933	Κυριάκος Γρυπαιός	(NULL)	(NULL)	Ρίτα Αμπατζή	His Master's Voice AO-211	(NULL)	(NULL)	Μην ακούς τη γετονιά, τον κακούρ
1254	Τσερκές	1933	Σταύρος Παντελίδης	Βιολί (Δημήτρης Σέμσι)	Σταύρος Παντελίδης	Ρίτα Αμπατζή	His Master's Voice AO-20	Αθήνα	Ζεϊμπέκι	Αχ πού 'να 'κείνος ο καρός τα χρ
1255	Φερετζέ φορώ	1933	Ρόζα Εσκενάζυ	Βιολί (Δημήτρης Σέμσι)	(NULL)	Ρίτα Αμπατζή	His Master's Voice AO-20	(NULL)	(NULL)	Φερετζέ φορώ, γαβρί μου, να τον
1256	Χαρικλάκι	1933	Παναγιώτης Τούντας	(NULL)	Παναγιώτης Τούντας	Ρίτα Αμπατζή	Columbia DG-452	Αθήνα	Τσιφτετέ	Χτες το βράδυ Χαρικλάκι
1272	Χήρα και μάγκας	1933	Ιάκωβος Μοντανάρης	(NULL)	(NULL)	Ρίτα Αμπατζή, Κώσ	Parlophone B-21731	(NULL)	(NULL)	Δε μου λες καλέ μου χήρα
1260	Απελτισμένοι	1934	Γρηγόρης Ασιτικής	(NULL)	(NULL)	Ρίτα Αμπατζή	Columbia DG 472	(NULL)	(NULL)	Απελτισμένοι βρίσκονται, κλαίγυ
1262	Βρε μάγκες δυο στη φω	1934	Κώστας Τζόβενος	(NULL)	(NULL)	Ρίτα Αμπατζή	HMV AO 2201	Αθήνα	Ζεϊμπέκι	Δυό μάγκες μεσ' στη φυλακή
1263	Για την αγάπη σου	1934	Παναγιώτης Τούντας	(NULL)	(NULL)	Ρίτα Αμπατζή	HMV AO-2156	(NULL)	(NULL)	Μου 'πες να γίνω μεπιαλής, να γίν
1265	Δώδεκα χρονών κορίτσι	1934	Σωτήρης Γαβαλάς	(NULL)	(NULL)	Ρίτα Αμπατζή	ODEON Ελλάδος GA 1751	(NULL)	(NULL)	Άιντε δώδεκα χρονών κορίτσι,
1266	Ζούλα η μαριωρή	1934	Παναγιώτης Τούντας	Ορχήστρα με βιολί (Δημ	(NULL)	Ρίτα Αμπατζή	HMV AO 2182	(NULL)	(NULL)	Ζούλα η Μαριωρή, το βράδυ,
1267	Ζουρλοπαιεμένης γένν	1934	Βαγγέλης Παπαδόγλου	(NULL)	(NULL)	Ρίτα Αμπατζή	HMV AO-2191	(NULL)	(NULL)	Ζουρλοπαιεμένης γέννα, έλα στη
1270	Η Νίτσα	1934	Γρηγόρης Ασιτικής	Ορχήστρα με ούτι (Γρηγ	(NULL)	Ρίτα Αμπατζή	HMV AO 2140	(NULL)	(NULL)	Καρσλαμα
1271	Η σφρανή	1934	Γρηγόρης Ασιτικής	(NULL)	(NULL)	Ρίτα Αμπατζή	HMV AO 2124	(NULL)	(NULL)	Θα 'ρθω, βρε Νίτσα, για να σε βρι
1272	Η Συριανή	1934	Ν. Κλαπής	(NULL)	(NULL)	Ρίτα Αμπατζή	HMV AO 2167	(NULL)	(NULL)	Είμ' σφρανή και δυστυχής σαν έρτ
1274	Θέλω κόκκινα να βάλεις	1934	Κώστας Μακρής	(NULL)	(NULL)	Ρίτα Αμπατζή	HMV AO 2230	(NULL)	(NULL)	Στην αμοιούδι καθόμωνα κι έποτ
1275	Κακά μικρούλα	1934	Δημήτρης Σέμσις (Ει	(NULL)	(NULL)	Ρίτα Αμπατζή	HMV AO 2111	(NULL)	(NULL)	Χήρα μαυροφορεμένη
1276	Καλογριά θε να γινώ	1934	Νίκος Κατζητανιώου	(NULL)	(NULL)	Ρίτα Αμπατζή	HMV AO 2238	(NULL)	(NULL)	Πες μου, βρε κακιά μικρούλα,
1276	Καλογριά θε να γινώ	1934	Νίκος Κατζητανιώου	(NULL)	(NULL)	Ρίτα Αμπατζή	HMV AO 2238	(NULL)	(NULL)	Καλογριά θε να γινώ και μαύρα να

1 Δυό μάγκες μεσ' στη φυλακή
2 τα βάλαν με τον διαφυντή,
3 τον αέρα να του πάρουν
4 κι ότi θέλουν για να κάσουν.
5
6 Βαρά μαγκά το μπουζούκι
7 κι όσε το μαστουρούλακι,
8 θέλω η πενιά να κλαίει
9 και τα ντέρτια μου να λείει.
10
11 - Δπα, γεια σας μάγκες!
12
13 Κι όσ' τα αίματα σου βυά

SELECT *,rowid "NAVICAT_ROWID" FROM "main"."songs" LIMIT 1000,1000

Record 169 of 1000 in page 2

Figure 3.3: Rebetiko Corpus in Navicat

span. Moreover, although the year of release of the recording was verified using the sources mentioned earlier, I had to rely on secondary sources, as it is difficult to get access to the catalogues of the record companies as well as the record sleeves, which would contain this information.¹⁶ The verification of metadata such as the song name, disc record number, and names of the involved artists was easier due to the fact, that many images of disc records which contain this information on their labels are available on the web. For instance, the *Virtual Kounadis Archive* alone includes images of 2,820 78 rpm disc records. The labels of these records also often include information about instrumentation and the corresponding rhythm (dance). The name of a lyricist is difficult to determine, since, as mentioned in Section 2.3, until 1946 the work of lyricists was mostly not acknowledged, and thus their names

¹⁶ The records of rebetiko are located in different archives, which are often private, and the images of record sleeves cannot be found on the internet. The access to the digitized catalogues of the record companies via the internet is also very limited, and visiting several archives on site to verify the information would be very time-consuming.

do not appear on the record or anywhere else. This is why the column corresponding to the name of the lyricist is often empty, as one can see in Figure 3.3 (see column *composer_lyrics*). Information such as performance metadata including orchestration, music performers and, if available, also the conductor (see column *performance_metadata*, Figure 3.3), as well as location of recording and the dance (rhythm) accompanying the song are not included in the list of metadata available at *rebetiko.sealabs.net*. They were almost exclusively taken from the *Virtual Kounadis Archive* as well as from the information provided on the images of disc record labels found elsewhere on the internet. Although these additional metadata are not directly used in the scope of this thesis, they help in getting a better understanding of rebetiko in general as well as its evolution: for instance, the change of instrumentation over time.

Regarding the verification of the lyrics, there were three main cases to take care of: either the lyrics to a particular song are missing at *rebetiko.sealabs.net*, or a song is provided with incorrect text, or it includes some errors in the text such as spelling errors. In order to tackle these problems, the same sources were used as in the case with the metadata: namely, web-based sources mentioned in Section 3.1.

The problem of missing lyrics could only be solved to some extent, as not all song texts or recordings are available on the internet. Printed rebetiko anthologies mostly do not include the metadata that is necessary, such as the disc record number, a year, or the names of the involved artists, which is why they were not used as a source for lyrics, as it would have been guesswork to connect entries in the printed anthologies to specific recordings. Moreover, these anthologies do not have a suitable license for further publication of the data within the corpus. I did not perform the process of song lyrics transcription, if only a recording with the missing song lyrics was found, because the workload would be too high for the scope of a master's thesis.

The issue of incorrect song lyrics usually is caused by the fact, that most of rebetiko recordings exist in several variations: these may either be re-recordings or reinterpretations of older rebetika. Regarding reinterpretations, differences in text are in many cases not reflected in the data on *rebetiko.sealabs.net*. Thus, some earlier rebetika are provided with the text of older ones or vice-versa. For example, the song «*Το χαρέμι στο χαμάμ*» (“The Harem in the Hammam”) composed and sung by Anestis Delias (disc number: Columbia DG 6165)¹⁷ and Stratos Pagioumtzis’ reinterpretation of it named «*Μες της πόλης το χαμάμ*» (“In The City’s Hammam”) sung by Stratos Pagioumtzis himself and Anna Chrysafi (disc number: Phillips 7739)¹⁸ actually have slight, but important differences in the text, which are not present in the data on *rebetiko.sealabs.net*. The first recording by Delias appeared in 1935 and the second one by Pagioumtzis in 1961. While the lyrics of the first recording include some lines referring to the Turkish hashish, these lines are missing in the version recorded in 1961, a time when censorship did not allow songs with such content, so that the rebetiko exponents were forced to practice self-censorship. The possible differences between variants of the same song could not be verified in all cases, since one cannot always find the required records or song lyrics.

¹⁷ Available at https://rebetiko.sealabs.net/mediawiki/index.php/Χαρέμι_στο_χαμάμ (accessed January 22, 2022)

¹⁸ Available at https://rebetiko.sealabs.net/mediawiki/index.php/Μες_της_πόλης_το_χαμάμ (accessed January 22, 2022)

Finally, the raw data contained a range of different errors such as false or missing apostrophes, wrong quotation marks, missing spaces and word accents. These errors would make further lyrics analyses difficult: for instance, the missing space and apostrophe in the phrase $\sigma'αγαπώ$ instead of $\sigma' αγαπώ$ (English 'I love you') or spelling errors such as $πέρνω$ instead of the correct $παίρνω$ (English 'take'). While a human would probably understand the meaning of a couplet despite such errors, for a computer this is impossible. Thus, clean data is of vital importance for any computer-assisted analysis. For this reason, I had to manually go through the lyrics and the metadata of each song. On average, there were multiple errors in each song, which I have corrected. For this reason, it was not sufficient to randomly choose and correct a certain amount of songs, but it was necessary to check and correct all of them.¹⁹

In general, specialists in the field of data science and machine learning spend a great amount of time with data verification and annotation. This step is unavoidable, since otherwise one will not obtain reliable results.²⁰

3.2.3 Key Data of the Rebetiko Corpus

Before introducing key data of the *Rebetiko Corpus* that was created in the scope of this thesis, it is important to present the definition of a corpus. Lothar Lemnitzer and Heike Zinsmeister define a corpus as follows: "A corpus is a collection of written or spoken utterances. Corpus data are usually digitised, i.e. stored in computers and machine-readable. A corpus consists of the data themselves, and possibly metadata, which describe these data, and linguistic annotations, which are assigned to the these data." (Lemnitzer/Zinsmeister 2015, 13)

Corpora are used not only in linguistics, but also, for instance, in different fields of history, sociology, anthropology, ethnography and literary studies as well as in the field of natural language processing. In corpus linguistics, one makes a distinction between general and specialised corpora (Lemnitzer/Zinsmeister 2015, 145). Reference corpora are aimed to represent all possible variants of one or several languages with regard to a domain (e.g. art, leisure, social and natural sciences) and medium (e.g. books, spoken language, periodicals and social media): for instance, the *DWDS-Kernkorpus (1900–1999)*²¹ or the *Russian National Corpus*.²² Specialised corpora are usually created with a specific research goal in mind: for instance, the *DWDS Corona-Korpus*²³ and the *DIGITARIUM*.²⁴ Thus, according to the introduced terminology, the *Rebetiko Corpus* is a specialised corpus due to the fact that it includes

19 Random sampling was, for instance, used in a project on computer-assisted analysis of rebetiko song lyrics of Bob Dylan (Sipl/Fuchs/Burghardt 2018).

20 Some information about the use of the World Wide Web as a source for a corpus as well as problems that one has to tackle using such data can be found in (Lemnitzer/Zinsmeister 2015, 42-43).

21 *The DWDS-Kernkorpus (1900–1999)* is a reference corpus for the German language of the 20th century. The corpus is available at <https://www.dwds.de/d/korpora/kern> (accessed January 26, 2022).

22 Available at <https://ruscorporu.ru> (accessed January 26, 2022)

23 *The DWDS Corona-Korpus* is a collection of texts from "German websites that deal with the COVID-19 pandemic from different perspectives (medicine, politics, philosophy, society) and that have been published since 2018" (cited as in (Corona-Korpus 2020). Available at <https://www.dwds.de/d/korpora/corona> (accessed January 26, 2022)

24 *DIGITARIUM* is a corpus of digitised issues of the Viennese newspaper *Wien[n]erisches Diarium* (today *Wiener Zeitung*), which was founded at the beginning of the 18th century. It contains not only digitised images of the issues, but also a machine-readable and machine-searchable text of the issues, making it a valuable and user-friendly source in fields such as history and linguistics. It is available at <https://digitarium.acdh.oeaw.ac.at> (accessed January 26, 2022).

only song lyrics of a particular music genre. It is therefore created for a specific research goal, although it also may be used in different fields such as history, linguistics, sociology and natural language processing.

The *Rebetiko Corpus* will be described with the help of the so-called corpus typology, which is often used in corpus linguistics.²⁵ The *Rebetiko Corpus* is a specialised static monolingual corpus of written language in Greek without linguistic annotations. It currently consists of 5,165 songs, of which only 3,772 songs are provided with song lyrics (398,015 tokens). As mentioned previously, the corpus data are based on songs that were recorded, therefore most of the metadata, provided with the songs, refer to particular recordings. To be more precise, songs include metadata such as the song title, source page name, release year, instrumentation, names of the involved artists (composer, lyricist, instrument player(s) and singer(s)), disc number, dance (rhythm) and location of recording. The metadata are provided in Greek. The performance metadata, such as orchestration and instrument players as well information about songwriter, dance accompanying a song and place of the recording, are often missing.

When building a corpus, it is also important to take into consideration the question of balance and representativeness. As mentioned in the previous paragraph, the *Rebetiko Corpus* contains 5,165 songs, of which only 3,772 songs are provided with song lyrics. In this context one could raise the question of how good a corpus of this size reflects the music genre rebetiko as a whole, when referring to the song lyrics. This question is difficult to answer due to the fact that rebetiko, as thoroughly presented in Chapter 2, does not have strict chronological boundaries, and cross-genre boundaries in the context of rebetiko are unstable as well. One would never be able to give an accurate number, for instance, of existing jazz, blues or also rebetiko songs. The only question that may be answered with pretty high accuracy, is the question of how many songs specific genre exponents were involved in. Thus, there are different views on the size of the rebetiko heritage. Some researchers have a very narrow definition of rebetiko, counting only about 100 songs that were recorded between the 1920s and the 1950s as rebetiko (for which the music genre rebetiko is specified on the label of the records) (Savvopoulos 2006, 29). Other researchers count more than 10,000 songs that belong to the genre (Kounadis 2003a, 407-409).

Regarding the question about balance, some rebetiko exponents were involved in several hundreds of songs, and others in only a few. As Table 6.1 in the appendix shows, the corpus contains nine composers who have authored more than a hundred songs, including Vasilis Tsitsanis (582 songs), Panagiotis Toundas (429 songs), Giannis Papagioannou (391 songs), and Manolis Chiotis (331 songs). Remarkably, the second highest number in this table is the number of songs of unknown authorship, counting 537 songs. This number has to be taken with a grain of salt due to the fact that many rebetiko songs are rearrangements of earlier songs of oral tradition, such that some ascribe these songs to certain rebetiko exponents, while others attribute them to unknown authorship and speak of rearrangements (*διασκευή*) in this context. Thus, the number of rebetiko songs of unknown authorship, which are included in the corpus, could be either lower than 537 songs or higher, depending on the degree of change from the original version and the researcher's approach towards this issue. Table 6.1 presents the 33 composers included in the corpus that were involved in at least 20 songs, but the corpus includes compositions by over 300 rebetiko composers, with numerous composers that are associated with only a few songs, among them Giorgos Rovertakis

25 For information about corpus typology see (Lemnitzer/Zinsmeister 2015, 137)

(10 songs), Giovan Tsaous (4 songs) and Stella Haskil (4 songs). It is important to note that this does not necessarily mean that these composers were only involved in a small number of songs. The *Rebetiko Corpus* only reflects the data available at *rebetiko.sealabs.net*, as taking songs from additional sources would also mean, that it would possibly be necessary to work with several licenses. This would make the further publication of the corpus more difficult. Nevertheless, the corpus presents a large variety of rebetiko exponents, and thus is a representative corpus of rebetiko song lyrics. The inaccurate number of the approximate number of composers (about 300) reflects the fact that almost every rebetiko exponent had a pseudonym²⁶, under which he or she has published some works. Real names of some of the exponents still remain unclear, which is also due to the fact that record companies released some songs under names of exponents that have never even existed (Genitsaris 1992, 76-77). When it comes to pseudonyms that still remain unclear, it is supposed that behind the name N. Stamas either Nikos Gounaris (Kournazos), Nikos Statheros or Markos Vamvakaris is hiding.²⁷

The corpus includes only songs that appeared as recordings, and therefore it does not represent some songs of the early oral tradition or songs that never were recorded. This should be taken into consideration when drawing conclusions based on the corpus data.

Generally, when creating a corpus it is usually made an attempt to obtain a balanced number of data over time, so that one can get more reliable estimations of certain phenomena over time. Figure 3.4 shows the distribution of data in the *Rebetiko Corpus* over time from 1906 to 1992, with most of the songs being recorded between the mid-1920s and 1941 as well as the mid-1940s and circa 1970.²⁸

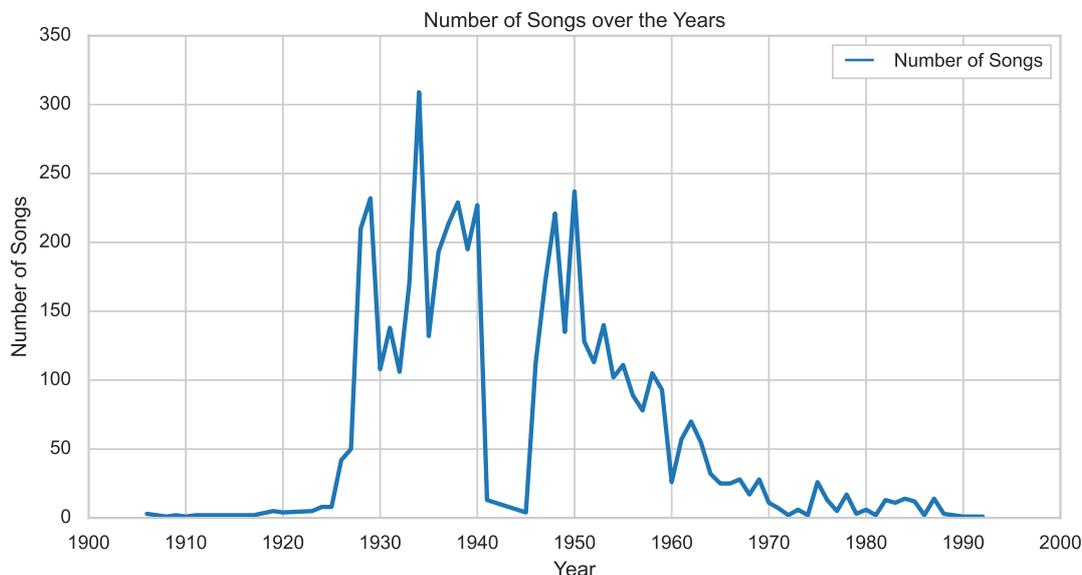


Figure 3.4: Distribution of Songs in the *Rebetiko Corpus* over Time

26 For instance, A. Rokos (Markos Vamvakaris), Marika Politissa (Frantzeskopoulou), K. Roumeliotis (Konstantinos Faltaits) and Dimitris Semsis (Salonikios)

27 See <https://rebetiko.sealabs.net/display.php?recid=6192> (accessed January 29, 2022)

28 The drop between the years of 1941 and 1946 was caused by the suspension of record production during the axis occupation (see Section 2.3).

3.2.4 *Corpus Publication on Zenodo*

The corpus has been published in JSON format on Zenodo and is freely available for non-commercial purposes under specification CC BY-NC-SA 4.0 (Arzt 2022). Zenodo is an open-access repository that is operated by CERN and is widely used for sharing datasets and other digital artefacts related to research in many areas. In this way the *Rebetiko Corpus* is visible to researchers from different fields and hopefully will spark new research interest in rebetiko.

The *Rebetiko Corpus*, like other corpora, may be used in many different fields such as linguistics, history, musicology and sociology. In this thesis, the corpus is mainly used for linguistic and historic analyses. Specifically, this chapter presents experiments regarding typical topics covered in rebetiko, looks at the development of the vocabulary over time and dives into linguistic properties of the rebetiko lyrics.

From a linguistic point of view, the corpus data is of particular value, as they are written in *Dimotiki*, the vernacular form of Modern Greek, and not in antiquitised *Katharevousa*, which de-facto was the official written language of Modern Greece since the formation of the state in the 1820s until 1976. The difference between these two variants of Modern Greek is so large that even not every Greek native speaker is able to read a text in *Katharevousa* without any problems. The debates regarding *Dimotiki* or *Katharevousa* diglossia, or the so-called *glossiko zitima* (the language question), which were actively lead in the nineteenth and twentieth century, actually go back to antiquity and the issue of *Koine*, or common Greek dialect.¹ Even though *Dimotiki*, or the demotic language, became an official language of Greece in 1976, the question of diglossia remains open until today, with many books being published in *Katharevousa* (Horrocks 2010, 466). *Rebetiko* songs also include numerous ‘underworld’ slang² words and expressions, which, with the growing popularity of rebetiko in the second part of the twentieth century, became and still remain part of the Greek colloquial language (Papadopoulou 2016). Moreover, due to the fact that many rebetiko exponents were originally from Asia Minor as well as the fact that Greece was part of the Ottoman Empire for almost four centuries, the language of rebetiko songs, unlike ‘purist’ *Katharevousa*, is rich in Turkish loanwords. *Rebetika* also contain many Italian and even some English loanwords. Due to the language of songs and the themes depicted, for a long time rebetiko exponents were considered by some to destroy the ‘pure’ Greek language and portray Greece in a bad light. The research of the vocabulary of rebetika may therefore enable a look from a different perspective at the history of the Modern Greek language, which is indissolubly connected with the strict state policy of Modern Greece directed against all loanwords, which remind of Greece’s connection to the East and therefore do not support the idea of continuity with the Ancient Greece.

What is more, one could claim that song texts fall into a separate category from both written and spoken language. Remarkably, song lyrics as a specific type of texts are not included in general language corpora such as the *British National Corpus*³ or the *DWDS-Kernkorpus (1900–1999)*⁴, which as reference corpora are aimed at representing as broad a spectrum of language forms as possible. Existing general corpora of Modern Greek like the *Hellenic*

1 On *Koine* Greek see (Horrocks 2010, 79-123)

2 In the context of rebetiko one often mentions the so-called *mangika*, or the sociolect of the mangas community, which influenced the language of rebetiko (Tragaki 2009, 32). At the same time it would be incorrect to equate the language of the mangas community with the language used by rebetiko exponents, at least due to the fact that many lyricists of rebetiko had only a loose connection to the mangas ‘underworld’.

3 Available at <https://www.english-corpora.org/bnc/> (accessed January 31, 2022)

4 Available at <https://www.dwds.de/d/korpora/kern> (accessed January 31, 2022)

*National Corpus (Εθνικός Θησαυρός της Ελληνικής Γλώσσας)*⁵ also do not contain song lyrics.⁶ When studying the meaning that a particular song conveys, it is impossible to separate the song lyrics from the other components of the song, such as music or dance (Turpin/Stebbins 2010, 2). Songs may follow ‘specific textual conventions’, that are different from both written and spoken language. It may be reflected in vocabulary, grammar, syntax, phonology and/or prosody. As Myfany Turpin and Tonya Stebbins point out, the songs may contain a higher number of vowels than in the spoken language, as they are ‘the main pitch-bearing unit of the voice’ (Turpin/Stebbins 2010, 4). Moreover, many songs contain non-lexical vocables such as *la la la* in English or *λάι λάι λάι (lai, lai, lai)* and *τραλαλά (tralala)* in Greek. Interjections that are used e.g. to express emotions are also very frequent in songs and in rebetika in particular: for instance, *γιάλα*, ‘giala (exclamation of misery and pain)’, *αμάν*, ‘oh, oh man (expression of surprise, disappointment or joy)’, *άιντε*, ‘let’s, go on’. All these peculiarities of the language of songs make a linguistic analysis of song lyrics an important area of research. Some characteristics of rebetiko song lyrics will be presented in Section 4.2.1, which deals with the challenges of using natural language processing techniques for the analysis of rebetiko song lyrics.

In the previous two paragraphs a brief overview of the importance of the rebetiko corpus data for a linguistic analysis was given, but one could also study the history of Greece through the lens of rebetiko. Such an approach may reveal narratives of social groups that were left out from the rest of society due to a range of historical events and are underrepresented in the official historiography. Moreover, although rebetiko has for a long time been connected with the Greek urban ‘underworld’, as time went on it increasingly touched on a broader social strata, as already mentioned in Chapter 2. In general, music is one of the components that forms a particular culture and society. Rebetiko is therefore an important primary source on studying the history and culture of Greece.

Regardless in which field rebetiko songs are intended to be studied, the *Rebetiko Corpus* may be used for distant reading in order to get a first impression about the data and explore them. First of all, the corpus is of course fully searchable, thus one can easily find all songs in which specific terms occur: for example, *αγάπη* (English ‘love’) or *Σμύρνα* which stands for the city of Smyrna. Furthermore, one might be interested in finding out, how the censorship possibly influenced the vocabulary of rebetika. Therefore, one could retrieve all songs that contain the word *χασίσι*, ‘hashish’ in the lyrics. This word appears in the corpus in 46 songs, almost all of which were recorded before 1936.⁷ Between 1936 and 1946 no records are included in the corpus that contain the word *hashish*, with the exception of one song that was recorded in the USA, namely «Μες του Μάνθου τον τεκέ» (English “‘Inside Manthos’ Tekes”), which was recorded in New York in 1938 (disc number: Orthophonic S-446-B). Even without consulting the corpus, it is a natural assumption that the recording of rebetika re-

5 The *Hellenic National Corpus (Εθνικός Θησαυρός της Ελληνικής Γλώσσας (ΕΘΕΓ))* is a corpus of written language, which contains circa 97,000,000 tokens, and includes data from media such as books, newspapers/ periodicals and the internet. It is available at <http://hnc.ilsp.gr> (accessed February 1, 2022).

6 For Modern Greek one could also mention the corpora of the newspapers *Τα Νέα (Ta Nea)* and *Μακεδονία (Macedonia)*. They are available at https://www.greek-language.gr/greekLang/modern_greek/tools/corpora/index.html, accessed February 1, 2022).

7 The corpus includes only two songs that contain the word *χασίσι* (English ‘hashish’) and were recorded in 1936, shortly before the establishment of the dictatorship and the following censorship: namely, the song «Παραπονούνται οι μάγκες μας» (English “Our Manges Are Complaining”) composed by Giovan Tsoulos and sung by Kalyvopoulos (Athens, disc number: HMV AO-2321) and «Σούρα και μαστούρα» (English “Drunkness and Kef”) composed by Anestis Delias (recorded in Athens in May 1936, disc number: HMV AO 2307).

ferring to drug consumption declined during Metaxas' regime established in 1936 and the following general suspension of record production in 1941 in Greece, which lasted until 1946. Nonetheless, it is of vital importance to compare one's assumptions and intuitions to the actual data. The word *hashish* only appeared again in song lyrics in Greece in 1946. These were namely two records of the same song that appeared in 1946 (both compositions by Vasilis Tsitsanis): «Το πρωί με τη δροσούλα» (English "Morning with the Dew") sung by Stratos Pagioumtzis and Vasilis Tsitsanis (Athens, disc number: Columbia 6598) and «Δροσούλα» (English "Dew") sung by Markos Vamvakaris, Apostolos Hatzichristos and Giannis Papagioannou (Athens, disc number: Parlophone B-74064), which were recorded within one week in June 1946. After 1946, when the censorship had been interrupted for a few months, the next and final recording in the corpus that includes the word *hashish* is the song «Το βαπόρι απ' την Περσία», English "The Ship from Persia" composed by Vasilis Tsitsanis and recorded in 1977 (disc number: Columbia 70529). It was therefore recorded already after the Junta fell in 1974. In this context it seems even more interesting to look at the song «Το χαρέμι στο χαμάμ», English "The Harem in the Hamman", composed by Anestis Delias and reinterpreted by Stratos Pagioumtzis. The first recording from 1935 (disc number: Columbia DG 6165) contains lines referring to the Turkish hashish, whereas the recording of 1961 named «Μες της πόλης το χαμάμ», English "In the City's Hamman" (disc number: Phillips 7739) does not, which was probably caused by the strict censorship in the 1960s. The fact, that after the fall of Junta in 1974 the corpus contains only one recording referring to drug consumption does not necessarily mean, that no other recordings referring to this topic have been made. Most of the corpus data, which are based on the data available at rebetiko.sealabs.net just date back to the period before the 1970s. The fact, that the word *χασίσι* (English 'hashish') occurs much more often until the mid-1930s strengthens the assumption made by Gauntlett and mentioned in Section 2.2, that dominant themes of rebetika until the establishment of the censorship of Metaxas were the so-called 'low life preoccupations' including hashish consumption, and that the censorship had an influence on the thematic spectrum of rebetika, so that from 1936 onwards love songs dominated the repertoire of rebetiko (Gauntlett 1985a, 100-101, 116).

This was just an example of how one could use the *Rebetiko Corpus* data for further research without even using sophisticated tools, but just using simple queries, in this case done via the database management tool *Navicat*, which was used for manual corpus data verification. In the next section, a few experiments with the *Rebetiko Corpus* data will be executed with the help of the much more powerful text analysis tool *Voyant Tools*.

4.1 DATA ANALYSIS WITH VOYANT TOOLS

There exist numerous text analysis tools or toolkits such as AntConc, Voyant Tools, Google Books Ngram Viewer or Sketch Engine. Such tools, for instance, allow to perform concordance analysis, build word clouds, analyse word frequencies over time and correlations between the words. In the present work, we will use the web-based application *Voyant Tools* for further analysis of the *Rebetiko Corpus*.

It is important to mention that the present work is not the first one that applies text analysis tools to study rebetiko song lyrics. For instance, in the scope of her dissertation Marianna Smaragdi compiled a corpus of 1,350 rebetiko song texts based on the lyrics available

in printed *rebetiko* anthologies and used the concordance program ConcappV4 to study metaphors of emotion in the context of *rebetiko* (Smaragdi 2012, 2, 5).⁸

Voyant Tools is an open-source web-based toolkit for text analysis and visualisation.⁹ In order to start working with a dataset or a single text, one can either insert the text into the corresponding field of the tool or upload the data. This application allows uploading the data in a range of formats such as PDF, XML, JSON or plain text. At the moment, Voyant tools consists of 28 different tools, including a word cloud generator, a concordance tool and a tool that shows a distribution of a word's occurrence across a text or a collection of texts. Currently, some of the tools are only provided for the English language: for instance, DreamScape which tries to detect locations in a text and then depicts them on a map, showing connections between them. The default view of Voyant Tools contains five tools, which are the so-called *Cirrus*, or a word cloud generator, the *Reader* which includes the entire original text or corpus, the *Trends* tool which builds a graph of a word's distribution across a corpus, a *Summary* tool that gives some general statistical information about the corpus such as number of tokens, vocabulary density, average number of words in a sentence and a list of the most frequent words in a corpus, and a *Contexts* tool with concordance lines which allows to analyse a keyword in context. A further benefit of Voyant Tools is that these tools interact with each other: for instance, if one clicks on a word in the *Reader* tool, one will get an update in the *Trends* and *Contexts* tools which will show the respective information about this word in the corpus. This allows to combine close reading in tools such as *Reader* and *Contexts* with distant reading in tools such as *Cirrus* or *Trends*. Next, some experiments with the *Rebetiko Corpus* using Voyant Tools will be presented.

4.1.1 Experiments with Voyant Tools: Word Clouds

Before beginning with the analysis, it is important to mention some general information about the form in which the data were used in the experiments. The basis for all following experiments in Voyant Tools is the *Rebetiko Corpus* containing 5,165 songs, of which only 3,772 songs are provided with song lyrics (398,015 tokens). More details can be found in Section 3.2.3.

Depending on the experiment, which one would like to perform, it is sometimes useful to first apply some text normalisation before the start of the experiments. Text normalisation is the process of transforming text into a form that preserves only the information relevant to the intended purpose of the investigation (Lemnitzer/Zinsmeister 2015, 197). Normalisation is performed in a preprocessing stage via natural language processing (NLP) algorithms, and may include such components as capital letter removal, tokenisation, lemmatisation, removal of punctuation or/and stop words, so the preprocessing steps depend on a particular task.¹⁰ Voyant only offers rudimentary text normalisation tools such as tokenisation and stop word removal, thus one often has to rely on external tools for data preprocessing.

⁸ The corpus created by Smaragdi contains songs composed between the 1880s and 1960s by around 80 song-writers and is not published (Smaragdi 2012, 2, 4).

⁹ Voyant Tools is available at <https://voyant-tools.org> (accessed February 5, 2022). This tool can also be installed locally. For more information see <https://github.com/sgsinclair/VoyantServer/releases/tag/2.4.0-M45> (accessed February 5, 2022).

¹⁰ For more information about text normalisation as well as NLP text preprocessing see (Duque 2020b) and (Duque 2020a)

For instance, one might like to perform an analysis of a word cloud of nouns for a particular collection of texts or a single text. Such an analysis may be helpful for finding main topics of a particular text or a corpus. In this case, one would need to obtain only nouns from the entire corpus, and it would also be advisable to perform a lemmatisation process, so that one can see all occurrences of a particular word across a corpus. When working with a Greek text, one has to tackle the problem of lemmatisation. Unfortunately, available lemmatisers¹¹ for Modern Greek do not solve this task as well as one would like to expect. For instance, the spaCy¹² lemmatiser for the Modern Greek language is rule-based¹³ and was not trained using modern machine learning methods.¹⁴ The accuracy of spaCy's rule-based lemmatiser for Greek is 56.46% to date, which is a quite poor performance.¹⁵ In general, the focus of many NLP algorithms is on high-resource languages (especially English), so that natural language processing for low-resource languages often still gives results of lower quality.¹⁶ While English, for which a large number of tools is available, is an analytic language with a low number of inflectional morphemes, Modern Greek is a highly inflected language, so that its grammar can be hardly described with a set of lemmatisation rules. The spaCy lemmatiser for Greek struggles the most with the lemmatisation of verbs, which in different tenses include such elements as an augment or a stem change. The spaCy lemmatiser faces some problems when performing lemmatisation of standard Modern Greek, and automatic lemmatisation of song lyrics is an even more challenging task. Nevertheless, the spaCy lemmatiser still is a great help for processing the Greek texts and can be adapted to some extent to solve certain problems. In order to perform the following experiments with the *Rebetiko Corpus* data, a short Python script was written that generates subcorpora, which will be used for specific analyses. With the help of the spaCy processing pipeline, which includes components such as a tokeniser, a lemmatiser and a part-of-speech tagger, subcorpora such as a subcorpus of lemmatised nouns and proper nouns included in the *Rebetiko Corpus* were created.¹⁷ In the following experiments it will be specified in which form the data were used.

To get a first impression of the corpus data, a word cloud of the entire corpus without any preprocessing of the data is presented in Figure 4.1. For this first analysis, primary data without text normalisation, e.g. removal of stop words or lemmatisation, were used. It is important to note that Voyant Tools itself automatically performs some basic text normalisation, namely tokenisation, i.e. splitting of the text into separate segments based on punctuation and whitespaces, as well as removal of punctuation marks.¹⁸ The analysis of primary data may show some general tendencies regarding rebetiko song texts, which will

11 A lemmatiser is an algorithm that automatically performs lemmatisation.

12 spaCy is a well-known open-source Python library for natural language processing. More information about spaCy can be found at <https://spacy.io> (accessed February 7, 2022).

13 More information at <https://github.com/eellak/gsoc2018-spacy> (accessed February 7, 2022)

14 A machine learning-based edit-tree spaCy lemmatiser is currently under development and is available in experimental mode for some languages (De Kok 2021).

15 This accuracy corresponds to the accuracy for the Greek model *el_core_news_lg-3.2.0*. The results can be found at https://github.com/explosion/spacy-models/releases/tag/el_core_news_lg-3.2.0.

16 As one could expect, the number of tools for the Ancient Greek is much higher than for the Modern Greek. See, for instance, Classical Language Toolkit (CLTK). Available at <http://cltk.org> (accessed February 7, 2022)

17 The source code as well as the generated subcorpora are available on GitHub at <https://github.com/kleinesgespenst/rebetiko-corpus-and-tools> (accessed February 20, 2022).

18 Due to this automatic tokenisation performed by Voyant Tools one can find single letters such as ς , ν μ in the word clouds, because the Voyant tokeniser treats apostrophes as distinct tokens and separates them from a word. For instance, ν' , a contraction of a particle $\nu\alpha$, is transformed to ν and $'$.

The next most frequent token μ' (3,270 times) is a contraction, which is polysemous and may refer either to a weak form of a first-person singular pronoun in accusative or genitive, or to a possessive pronoun corresponding to the English ,mine', or to a shortened version of a preposition $\mu\epsilon$,with'. This contraction corresponds, for instance, to a weak form of a first-person singular pronoun in accusative in the line «*Μ' έκανε να τρελαθώ με τα καμώματά της*' (English "She made me mad with her larks") from a song by Michalis Genitsaris «*Μικρούλα Πειραιώτισσα*» ("The Girl from Piraeus").¹⁹ In the line «*και τώρα πιά δεν μου μιλάς και μ' άλλη φεύγεις και περνάς*» (English "You don't talk to me anymore and go out with another one and spend your time with her") from the song «*Το γιασεμί*» ("The Jasmine")²⁰ by Dimitris Semsis one observes a contraction of the preposition $\mu\epsilon$,with'. In general, these examples with μ' were given in order to point out the general tendency of ambiguity of contractions in rebetiko song lyrics, which is a frequent phenomenon in song lyrics in general. As this example shows, one and the same contraction is often polysemous or homonymous. An identification of a word meaning or its grammatical function is an easy task for a human, but not for a machine. Solving the problem of word-sense disambiguation is a challenging task in natural language processing. This is one of the reasons why it is difficult to obtain a correct lemmatisation for the Greek language, and why, as mentioned previously, available tools for Modern Greek such as the spaCy lemmatiser do not perform the task properly.

The lexical item $\tau\eta$ (2,934 times) is homonymous and may serve either as a feminine definite article in accusative or a shortened feminine weak third-person personal pronoun in accusative (cf. $\tau\eta < \tau\eta\nu$). Using the interaction of Voyant tools, one easily obtains examples of this keyword in context in the tool *Contexts*. For instance, it occurs in the line «*Θα σαλτάρω, θα σαλτάρω, τη ρεζέρβα να τους πάρω*' (English "I'll leap, I'll leap, their spare tires I'll take") of the very well-known song by Michalis Genitsaris «*Θα σαλτάρω*», which was recorded several times.²¹

The last of the five most frequent tokens in the *Rebetiko Corpus* is $\kappa\alpha\rho\delta\acute{\iota}\alpha$,heart', which occurs 2,454 times. In this word cloud, one can also find words such as $\alpha\gamma\alpha\pi\acute{\omega}$,to love', $\alpha\gamma\acute{\alpha}\pi\eta$,love' (noun), $\mu\iota\kappa\rho\acute{\upsilon}\lambda\alpha$,(my) little girl', $\gamma\lambda\upsilon\kappa\acute{\iota}\alpha$,(my) sweet girl, honey', $\mu\acute{\alpha}\tau\iota\alpha$,eyes', $\gamma\upsilon\nu\alpha\acute{\iota}\kappa\alpha$,woman' and $\alpha\gamma\kappa\alpha\lambda\acute{\iota}\alpha$,hug'. Their presence in the word cloud confirms the observations of numerous rebetologists that a great number of rebetika is related to the topic of love.²² Moreover, in this word cloud one also finds tokens such as $\mu\acute{\alpha}\nu\alpha$,mom', $\mu\alpha\nu\acute{o}\upsilon\lambda\alpha$,(my) dear mom', thus, as also deeper analysis of context shows, a warm feeling of the protagonist is addressed not only to a beloved one, but also to a mother. This also attests the observation that many rebetiko songs address the figure of a mother, who is always kind and ready to come to help. Interestingly, as Smaragdi points out, the figure of a father and other relatives are mostly absent in rebetiko songs (Smaragdi 2012, 4). In the *Rebetiko Corpus* the figure of the father is mentioned only in 29 songs, while the figure of the mother appears in 559 songs out of 3,772 songs (almost 15% of the corpus data).

19 The song «*Μικρούλα Πειραιώτισσα*» ("The Girl from Piraeus") by Michalis Genitsaris ("was recorded in 1938 (disc number: HMV AO-2464).

20 The song «*Το γιασεμί*» ("The Jasmine (Giasemi)") by Dimitris Semsis (Salonikios) was first recorded in 1939 (disc number: HMV AO-2593).

21 See, for instance, the recording made in 1974 (disc number: Κυκλάδες 28923)

22 See, for instance, (Gauntlett 1985a, 164-165)

This word cloud also shows an abundance of interjections in rebetiko songs, such as *αμάν* ,oh, oh man (expression of surprise, disappointment or joy) ‘, *βρε* ,hey, dude, geez‘, *ρε* ,hey, dude, geez‘, *αχ* ,ah, oh, ouch‘ and *πω/ πω πω* ,oh, oh my, wow‘.²³

When looking at the word cloud in Figure 4.1 one also finds tokens such as *μένα* (653 times) and *σένα* (833 times), which correspond to the first- and second-person personal pronouns in accusative, respectively. Remarkably, these personal pronouns sometimes occur in the song lyrics with the neuter definite article in plural *τα*: for instance, in the line «*Με τα μένα να μπλεχτείς, μη γυρεύεις, βρε βλαμάκι*» (“If you mess with me, you won’t get back, you vlamis”) from the song «*Ρουλέτα*» (“Roulette”) composed by Minos Tsama (Matsas), who is primarily known as a rebetiko lyricist, and sung by Rita Abatzi (disc number: Odeon GA-1705, recorded in 1934). In this particular case *με τα μένα* means ‘with me’, which in standard Modern Greek corresponds to *με μένα* (,with me‘), i.e. without an article. The use of a prepositional definite article with a personal pronoun is attested in some dialects of Greek. It can be used as a nominalised pronoun, so that *τα μένα* (,mine‘) is synonymous with a nominalised possessive pronoun *τα δικά μου* (,mine‘). In the construction *για τα μένα* it may also mean ‘for myself’, being therefore synonymous with the construction with the reflexive pronoun *για τον εαυτό μου* (,for myself‘). Interestingly, a nominalised possessive pronoun with a prepositional definite article was also used in Ancient Greek (Webster 1864, 30). Constructions such as *τὰ ἐμά* (,mine‘) *τὰ σά* (,yours‘) are, for instance, commonly used in the New Testament. Such a use of a pronoun in Ancient Greek may be, for instance, found in the exclamation made by a priest during the Anaphora, in which bread and wine are lifted up: namely, «*Τὰ σὰ ἐκ τῶν σῶν Σοὶ προσφέρομεν κατὰ πάντα καὶ διὰ πάντα*» (“Thine own of Thine own (Thine own gifts) we offer unto Thee on behalf of all and for all.”). In songs though a prepositional definite article *τα* before a personal pronoun is mostly used not as a building element of a nominalised possessive pronoun, but as a filling element to match the metre. This is, for example, the case in the song mentioned previously called “Roulette”, which is written almost completely in decapentasyllabic verse. A prepositional definite article is also used in the song «*Σπάσ’ τα φως μου για τα μένα*» (“Smash it all, for Me, My Light”) composed by Dimitris Semsis and sung by Rita Abatzi (disc number: HMV AO-2320, recorded in 1936). The use of a personal pronoun with a definite article in some rebetiko songs was mentioned in the dissertation of Stathis Gauntlett (Gauntlett 1985a, 64). The neuter definite article in Plural with the pronoun *μένα* occurs in 12 songs presented in the *Rebetiko Corpus*, and with the pronoun *σένα* in 17 different songs. The construction *τα σένα* actually occurs 27 times in the corpus, but this includes some repetitions either within a refrain of a song or due to re-recordings of the same song.²⁴ In all the cases found in the *Rebetiko Corpus* such a construction is used to match the metre. Not only definite articles are used as filling components to match the metre in rebetiko songs, but also ‘extended’ pronoun forms such as

23 The token *πω* may also refer to a future simple form of the verb *λέω* ,say‘ (first person singular), and the existence of such homonyms should be taken into account during the analysis.

24 The number of occurrences of a word or a phrase in any corpus should be taken with a grain of salt: for instance, not all constructions which can be found in a corpus are grammatically correct. A corpus containing media such as newspapers will with high probability also include news tickers, and if a news item contains some mistake, it will appear in the corpus several times. For instance, even though the sentence “Ich habe fertig”, which according to the German Grammar is incorrect, was cited by a newspaper from a press conference of Giovanni Trapattoni, then coach of Bayern Munich, in 1998 and therefore appears in a newspaper corpus, this does not mean that this construction became grammatically correct. This phrase, for instance, appears in the DWDS corpus of the Berliner Zeitung (https://www.dwds.de/r/?q=ich+habe+fertig&corpus=bz&date-start=1994&date-end=2005&format=full&sort=date_desc&limit=10). Accessed February 3, 2022

μένανε (109 times, μένανε < μένα), σένανε (187 times, σένανε < σένα), αυτήνε (9 times, αυτήνε < αυτή(ν)), αυτόνε (19 times, αυτόνε < αυτόν) and even αυτόνανε (2 times, αυτόνανε < αυτόν). The form αυτόνανε, corresponding to a third-person masculine personal pronoun in accusative, appears in the line «Εγώ το μόρτη θ' αγαπώ, γιατί μ' αυτόνανε γλεντώ» (“I’ll love the rascal, cos I do party with him”) from the song «Ο μόρτης» (“The Rascal (Mortis)”) composed by Dimitris Barousis (Lorentzos) and sung by Kostas Roukounas (disc number: Odeon GA-1620, recorded in 1932). This extended pronoun is used to match the pattern of the octosyllabic verse, in which the song “The Rascal” is written. In the line «Το πλούτος για τα μένανε είναι τα δυο σου μάτια» (“The wealth for me is in your two eyes”) from the song «Ο κόσμος πλούτη λαχταρεί» (“People Long for Wealth”) composed by Kostas Skarvelis and sung by Stratos Pagioumtzis (disc number: Odeon GA-7231, recorded in 1939) one observes both the use of a prepositional definite article and an extended form of a personal pronoun in accusative. According to the corpus data, both the use of a prepositional definite article before a pronoun as well as an extended form of a personal pronoun in accusative serve as filling components in song lyrics, to fit the metre. As it was mentioned in the introduction to this chapter, song lyrics follow ‘specific textual conventions’, and rhythm is a main element which connects the musical and textual components of a song (Tait/Tabain/Sykes 2014, 189). The analysis of the use of pronouns was performed with the help of Voyant tools *Contexts* with the concordance lines as well as *Reader*, in which the entire corpus is available. Due to the fact that Voyant tools interact with each other, it is, for example, possible to make first observations using the tool *Cirrus* and then perform a more accurate analysis using tools such as *Contexts*, in which a query word’s surrounding context is shown. Such an approach makes it possible to combine distant reading with close reading.

The next word cloud, presented in Figure 4.2, demonstrates the most salient words in the *Rebetiko Corpus* after the removal of stop words. For the creation of the corpus without stop words a Python script was used. In general, Voyant Tools offers a function for the removal of stop words, but the stop words have to be entered in a relatively inconvenient way and are not persistent over multiple sessions. It is much easier to write a script, which generates all necessary corpora or subcorpora adapted for a specific task. The spaCy stop word list was extended and adapted to the rebetiko lyrics, so that the final stop word list includes 786 function words including prepositions, articles, pronouns, and auxiliary verbs such as *κάθε* (‘every’), *όταν* (‘when’), *έχω* (‘to have’) and *είμαι* (‘to be’). The removal of function words is certainly not necessary for every kind of text analysis, but in the context of the interpretation of a text or a corpus they carry less lexical meaning. After removal of stop words and punctuation, the rebetiko dataset contains 170,905 tokens. As one can see, the word cloud in Figure 4.2 now reflects the most important topics of rebetiko in a much clearer way. In this word cloud one can see even more interjections than in the previous word cloud, including *έλα* (‘come on’), *ώπα/όπα* (‘oops, golly, gee’), *γεια* (‘hi, hello, bye’, *άντε/άιντε* (‘let’s, go on’). The appearance of the token *γεια* (‘hi, hello, bye’ in the word cloud attests the frequent use of the so-called *yiassou* spoken interjections in rebetiko recordings, in which a rebetiko exponent is usually praised for his or her artistic merit (see Section 2.5, p. 44-46).²⁵

Additional lemmatisation of the texts in the *Rebetiko Corpus* results in a word cloud that is even easier to interpret (see Figure 4.3). In this word cloud, like in the word cloud in Figure

²⁵ For example, such *yiassou* interjection as «Γειά σου Μάρκο με το μπαγλαμά σου!» (“Health to you, Markos, with your baglamas!”) in the song «Όταν πλύνω τουμπεκάκι» (English “When I Clean the Hookah Tobacco”) by Markos Vamvakaris (disc number: His Master’s Voice AO-2065, recorded in 1933).

a rapid change of the audience's tastes, which the exponents of the older generation could only follow with difficulty, as well as the beginning of widespread use of the four-course bouzouki (see Section 2.3, p. 36). All these circumstances contributed to a pessimistic mood of some of the rebetiko songs of that period.

Looking at the word cloud in Figure 4.3 with lemmatised words, one also finds such forms as *γειτονιάς* (*γειτονιάς* < *γειτονιά* ,neighbourhood'), *γείας/γείας* (*γείας/γείας* < *γεία/γεία* ,hi, hello, bye') and *πεθάνω* (*πεθάνω* < *πεθαίνω* ,die'), which are incorrectly lemmatised. As mentioned earlier, these mistakes are caused by the fact that the spaCy lemmatiser for Modern Greek, which was used for the lemmatisation of the *Rebetiko Corpus* data, is rule-based, and the grammar of a highly inflected language such as Greek is next to impossible to be described by a set of rules, as well as the fact, that before performing lemmatisation, the spaCy pipeline includes a pre-trained part-of-speech tagger, so that the part-of-speech tagging varies depending on context. For instance, the spaCy pipeline performs a correct part of speech tagging and lemmatisation of the single line «Τα παιδιά της γειτονιάς σου με πειράζουνε» (“Your neighbourhood’s kids are teasing me”).²⁹ Though, when performing the part-of-speech tagging of the entire song «Τα παιδιά της γειτονιάς σου» (“Your neighbourhood’s kids”), it makes mistakes and therefore does not perform correct lemmatisation either, because the spaCy lemmatiser is dependent on the information it obtains from the POS-tagger.³⁰ Thus, for instance, instead of tagging the token *πειράζουνε* as a verb, it tags it as a proper noun, and the lemmatiser returns *πειράζουνε* (third-person plural Present Tense) as a lemma instead of the correct *πειράζω* (,tease'). This erroneous part-of-speech tagging happens for several reasons, including the type of data on which the components of the spaCy pipeline were trained, namely OSCAR Common Crawl and Wikipedia.³¹ Rebetiko lyrics frequently use non-standard word order, as it is often the case with song lyrics in general. Moreover, song lyrics use less punctuation, which makes it extremely difficult for a machine to ‘understand’ the sense of a long sequence of words with unusual word order, numerous contractions, frequent special vocabulary or even non-standard word spelling. When it comes to the lemmatisation rules³² that are currently used by the spaCy lemmatiser for the Greek language, they also show some limitations typical of a rule-based approach. For instance, according to the JSON file *el_lemma_rules.json*, if the lemmatiser gets the information that a particular token is a noun and it ends in the stressed *-ά*, then it should be inflected to *-άς*. The problem is that Greek nouns in nominative may end in both stressed *-ά* in feminine (e.g. *γειτονιά* ,neighbourhood') as well as in stressed *-άς* in masculine (e.g. *ντουνιάς* ,world'). This rule may be the reason, why, for instance, the word *γειτονιά* is lemmatised as *γειτονιάς*. Such a problem appears also with the verb form *πεθάνω*. The form *πεθάνω* corresponds to a first-person singular present tense (perfective aspect) of the verb *πεθαίνω* ,die'. One would suggest to write the rule, which transforms *-άνω* to *-άινω*, but there are also Greek verbs which end in their base form in *-άνω* including the verb *κάνω* ,do, make'. This problem may be solved with the help of a better morphologiser, so that a correct stem for each

29 The line is from the song «Τα παιδιά της γειτονιάς σου» (“Your neighbourhood’s kids”) of unknown authorship which was recorded several times including the recording made in 1929 (disc number: Columbia 8390, sung by Andonis Dalgas).

30 See the spaCy pipeline description at <https://spacy.io/models> (accessed February 10, 2022)

31 See <https://spacy.io/models/el> (accessed February 10, 2022)

32 The lemmatisation rules can be found in the JSON file in the explosion GitHub repository at https://github.com/explosion/spacy-lookups-data/blob/master/spacy_lookups_data/data/el_lemma_rules.json (accessed February 10, 2022).

an excerpt from the song «Ελενάκι» (“Little Eleni”)³⁴ is presented, in which the word *φαρμάκι* ‘poison’ occurs.

Ελενάκι, Ελενάκι	Little Eleni, little Eleni,
’συ με πότισες φαρμάκι	you’ve given me poison to drink
...	...
Στραβά το ’βαλες Ελένη	You stuck the knife badly
το μαχαίρι και δεν βγαίνει	And it doesn’t go out

In these lines we observe a metaphorical use of the word *φαρμάκι* ‘poison’. In the song “Little Eleni” the protagonist is suffering because of a woman named Eleni, who tortures him and ‘stuck the knife’ in his body, which is another metaphor in this song. In this song therefore the word *poison* refers to emotions, caused by the heartless Eleni. As Marianna Smaragdi in her dissertation on metaphors of emotion in rebetiko shows on numerous examples, a metaphorical use of the word *φαρμάκι* ‘poison’ is frequent when referring to the emotions with which a protagonist has to struggle because of a woman (Smaragdi 2012, 139-141).

In the word cloud in Figure 4.4 one also finds the word *χήρα* ‘widow’, which occurs 179 times in the nouns’ subcorpus. Below one can find a quatrain from the song «Χασικλήδες» (“Hash Smokers”) of unknown authorship.³⁵

Βρε τη νταμίρα τη νταμίρα	Oh hash, oh hash,
μου την έμαθε μια χήρα	I learned it from a widow
μ’ έκανε και αλανιάρη	She made me to a bum,
χασικλή και κουρελιάρη.	a hash smoker dressed in rags.

In these lines the protagonist blames some widow that he became a hashish smoker and a shabby man of the street supposedly due to the fact that he is in a relation with this widow. There are also numerous rebetiko songs, in which the protagonist declares his love to a widow and depicts her beauty: for instance in the song «Χήρα μ’ έκαψες» (“Widow, You’ve Ruined Me”) composed by Kostas Skarvelis and sung by Rita Abatzi (disc number: Columbia DG-361, recorded in 1933). In general, in the 1920s and the 1930s a great number of rebetiko songs was recorded, which refer to the figure of a widow. This was caused by tragic events of that time including the Balkan Wars, World War I and the Asia Minor Disaster, so that many women became widows (Politis 2014, 42). The *Rebetiko Corpus* includes 95 songs referring to a widow, of which 87 songs were recorded before 1940. It is of interest, that the corpus only includes very few songs recorded after World War II that refer to a widow. Remarkably, there are also rebetiko songs, in which a protagonist praises the beauty of a divorced woman (*ζωντοχήρα*), as it is, for instance, the case in the song «Το χασαπάκι» (“The Sweet Butcher”) composed by Kostas Skarvelis and sung by Zacharias Kasimatis (disc number: Pathé X-80160, recorded in 1929).

34 The song «Ελενάκι» (“Little Eleni”) is of unknown authorship and was recorded several times including the recording made in 1926 (disc number: Odeon GA-1153, sung by Giorgos Vidalis).

35 The song «Χασικλήδες» (“Hash Smokers”) of unknown authorship was recorded under different titles several times including the recording made in 1927 (disc number: Columbia 8017, sung by Kostas Karipis).

Είμαι χασάπης σεβνταλής, πού μ' έχει μπλέ- ξει πάλι	I'm a butcher, a lover who got in mess again
μια ζωντοχήρα σε μπελά, σε ντέρτια θα με βάλει	A divorcee will put me in trouble, in sor- rows

Interestingly, in rebetiko songs not only a man is striving for a carefree life refusing traditional values such as marriage and family (see Section 2.1, p. 7). A woman is also often depicted as the one, who does not accept the established values. The word cloud in Figure 4.4 includes not only words such as *μάγκας*, 'vagabond, wide boy' (703 times), but also such as *αλανιάρρα*, 'vagrant, bum (feminine)' (73 times).

The word cloud in Figure 4.4 also shows a relatively high occurrence of words such as *φυλακή*, 'prison' (165 times) *λουλάς*, 'hookah, waterpipe' (107 times), *αργιλές*, 'hookah, waterpipe, arghileh' (77 times), but in comparison to topics such as love, songs depicting prison life and hashish consumption occur less frequently, at least in the *Rebetiko Corpus* data. It is at the same time important to note that the fact that words such as *χασίσι*, 'hashish' are not present in a word cloud with the most salient nouns for the *Rebetiko Corpus*, does not automatically mean that this topic was not important in rebetiko songs. On the one hand, there are numerous synonyms for words such as *χασίσι*, 'hashish' and *φυλακή*, 'prison', which are used in rebetiko texts (e.g. *μαύρο*, 'black (hash)' for *χασίσι*, 'hashish' and *σίδερα*, 'iron (Pl.), bars'³⁶ for *φυλακή*, 'prison'). On the other hand, the *Rebetiko Corpus* only contains song texts of a part of the recordings that survived until our days, and numerous rebetiko songs, especially those related to the 'underworld', have never been recorded or preserved for the future in other forms (e.g. as sheet music). Moreover, the word *χασίσι*, 'hashish' occurs 75 times in the corpus, but it got assigned a false part-of-speech tag 21 times and therefore occurs only 54 times in the subcorpus of nouns. All these factors should be kept in mind when analysing the data with the help of a text analysis tool such as a word cloud. A deeper analysis of the topic of hashish consumption in the *Rebetiko Corpus* will be presented in the next section.

A closer look at the word cloud in Figure 4.4 also reveals that it contains words such as *μπουζούκι*, 'bouzouki' (172 times), *μπουζουκάκι*, 'little/sweet bouzouki' (74 times), *μπαγλαμάς*, 'baglamas' (117 times) and *βιολί*, 'violin' (139 times). These are some of the most important music instruments used by rebetiko exponents (see Section 2.3). While a violin along with a santour was used by the representatives of the Smyrna style of rebetiko, the bouzouki and the baglamas dominated the Piraeus style of rebetiko. Interestingly, these music instruments appear in the rebetiko recordings not only in the context of spoken interjections, in which an addressee was usually praised for his musical mastery (see Section 2.5), but also directly in the lyrics of the songs. These music instruments appear often in the context of *glendi* (party, revelry), as one of the components of merrymaking. Below, an excerpt from the song «Ο γλεντζές» ('The Merrymaker') composed by Stavros Pandelidis is presented.

Παίχτε τα βιολιά, να χορέψω έμορφα κορίτσια για να βγάλω το μεράκι πο' 'χω στην καρ-	Play the violins so that I dance with beautiful girls and show the passion of my heart
--	--

³⁶ See, for instance, expressions such as «βάζω στα σίδερα» ('to put smb. behind bars')

διά.

This song was recorded several times, including two recordings made in 1934 with two different singers, namely Kostas Roukounas (disc number: Columbia DG-6107) and Rita Abatzi (disc number: HMV AO-2181). Both the composer of the song Stavros Pandelidis and Rita Abatzi were originally from Smyrna, one of the centres of the Smyrna style of rebetiko, and Kostas Roukounas was originally from the island of Samos. Both these recordings, as one would also expect from the text, include the violin among other instruments. In the word cloud in Figure 4.4 one also finds rebetiko venues, in which this merrymaking accompanied by music took place, namely *ταβέρνα* ,tavern‘ (136 times) and *τεκές* ,tekke, hashish den‘ (95 times). When it comes to other rebetiko venues, *καφέ αμάν* ,café aman‘ occurs two times in the *Rebetiko Corpus*, in two recordings of the same song called «Στο καφέ-αμάν» (“In the Café Aman”) composed by Kostas Skarvelis.³⁷ In the corpus data one also finds another rebetiko venue called *μπύρες* ,beer dens‘ (50 times). Remarkably, in the song «Κατεργάρα μου Ειρήνη» (“My Mischievous Irini”)³⁸, composed by Andonis Dalgas, not only *μπύρες* ,beer dens‘ are mentioned, but also *καφέ σαντάν* ,café chantant‘. Cafés chantants came to the Ottoman Empire and Greece from France and offered a westernised music repertoire (see Chapter 2). This is the only time *café chantant* occurs in the corpus data.

Next, we will briefly look at the word clouds depicting the most salient words in the subcorpora of four different rebetiko exponents: namely, the subcorpora with the songs composed by Giannis Dragatsis (Ogdondakis), Markos Vamvakaris, Ioanna Georgrakopoulou and Vasilis Tsitsanis.³⁹ The word clouds can be found in Figure 4.5 (Dragatsis, subcorpus size 3,687 tokens), Figure 4.6 (Vamvakaris, subcorpus size 7,485 tokens), Figure 4.7 (Georgrakopoulou, subcorpus size 1,914 tokens), and Figure 4.8 (Tsitsanis, subcorpus size 24,571 tokens). The fact that these four rebetiko exponents are listed as the composers of the songs does not automatically mean that they were also the lyricists. As it was mentioned, for instance, in Section 2.3, lyricists of many songs recorded before 1946 were not specified and therefore remain unclear.

First, we will take a look at the subcorpus of Giannis Dragatsis (Ogdondakis). He is a prominent representative of the Smyrna style of rebetiko, born himself in Smyrna to a family of professional musicians in 1886. He came to Greece in 1923 due to the mandatory exchange of population between Greece and Turkey. Giannis Dragatsis worked simultaneously as a composer, lyricist and a music instrument player, being one of the most prominent violinists of rebetiko. Dragatsis died in Athens in 1958. The subcorpus with the songs composed by Giannis Dragatsis contains 125 songs (3,687 tokens after removal of stop words).

The word cloud of the Dragatsis subcorpus (see Figure 4.5) contains numerous interjections such as *αμάν* ,oh, oh man‘ (157 times), *αχ* ,ah, oh, ouch‘ (93 times), *έλα* ,come on‘ (20 times), *άντε* ,let’s, go on‘ (20 times), *βρε* ,hey, dude, geez‘ (37 times), *μπράβο* ,bravo, well

37 It was recorded in 1933 and sung by Roza Eskenazi (disc number: Columbia DG-335) and Marika Politissa (Frantzeskopoulou) (disc number: Parlophone B-21691).

38 The song “My Mischievous Irini” was recorded in 1931 and sung by Andonis Dalgas himself (disc number: HMV AO-2013).

39 The *Rebetiko Corpus* contains 125 songs of Giannis Dragatsis, 158 songs of Markos Vamvakaris, 46 songs of Ioanna Georgrakopoulou and 582 songs of Vasilis Tsitsanis. For more detail see Appendix 6.1

In Figure 4.5, we can also find words such as *φουμάρω*, ‘to smoke’ (23 times), *λουλάς*, ‘hookah, waterpipe’ (28 times). Interestingly, in the *Rebetiko Corpus*, instead of the standard verb *καπνίζω*, which stands for ‘smoke’, the Italian borrowing *φουμάρω* is used much more often. A more detailed analysis of the use of verbs *καπνίζω* and *φουμάρω* will be presented in the next section (see Section 4.2.2). In this word cloud, we also find a word connected to rebetiko songs about prison life, namely *φυλακή*, ‘prison’ (20 times). Dragatsis’ subcorpus also shows a high frequency of lexical items such as *μάγκας*, ‘vagabond, wide boy’ (18 times) and *σερέτης*, ‘obstinate, wilful’ (10 times), which are used to define a protagonist of rebetiko songs. Moreover, in the songs composed by Giannis Dragatsis, the protagonist is also often defined as a *φίνος μάγκας*, ‘fine/right mangas, vagabond, wide boy’. The appearance of the word *Κόνιαλης*⁴³ is caused by two songs of Dragatsis, which were recorded several times, namely «Κόνιαλης» (“Man from Konya”)⁴⁴ and «Νέος Κόνιαλης» (“Young Man from Konya”).⁴⁵ In general, in rebetiko songs many geographical places as well as the inhabitants of particular places are mentioned, especially women (e.g. *Πειραιώτισσα*, ‘woman from Piraeus’ and *Σμυρνιά*, ‘woman from Smyrna’).

To summarise, looking at the word cloud for Dragatsis’ subcorpus one can expect songs related to topics such as love (often unrequited one and resulting sufferings of a protagonist), vagabond (mangas) life, drug consumption and prison life.

Next, we will analyse the subcorpus of songs by Markos Vamvakaris, who is one of the most prominent representatives of the Piraeus style of rebetiko, being even called a ‘patriarch of rebetiko’ for his influence on the genre. He was born in 1905 in Syros and came to Piraeus in 1917. He was simultaneously a composer, a lyricist and a bouzouki player. Vamvakaris died in Nikaia in 1972.⁴⁶ He belongs to the so-called first generation of bouzouki players in the context of rebetiko. The subcorpus with the songs composed by Markos Vamvakaris contains 158 songs (7,485 tokens after removal of stop words).

The word cloud in Figure 4.6 reflects the most salient words for Vamvakaris’ subcorpus. Based on the word cloud, one can make the observation that the song texts also contain a large number of interjections, but they are less diverse than in the Dragatsis’ subcorpus. In the Vamvakaris subcorpus one can find interjections such as *αμάν*, ‘oh, oh man’ (84 times), *αχ*, ‘ah, oh, ouch’ (46 times), *έλα*, ‘come on’ (86 times), *άντε*, ‘let’s, go on’ (20 times), *βρε*, ‘hey, dude, geez’ (88 times) and *γεια*, ‘hi, hello, bye’ (52 times). Remarkably, this word cloud also contains the given name of Vamvakaris, namely Markos, which though was not correctly lemmatised and remained in vocative (see *Μάρκο*). This reflects the frequent use of the so-called *yiassou* interjections, which praise the artistic merit of a rebetiko exponent, even in the studio recordings.

Like in Dragatsis’ subcorpus, the most frequent noun of the Vamvakaris subcorpus is *καρδιά*, ‘heart’ (124 times). According to the word cloud, in the songs by Vamvakaris the topic of love also prevails over all other ones. Additionally, one can find words such as *αγαπώ*, ‘to love’ (76 times), *μάτι*, ‘eye’ (70 times), *αγκαλιά*, ‘hug’ (22 times), *γυναίκα*, ‘woman’ (19 times), *καρδούλα*, ‘sweetheart’ (15 times) and *φιλί*, ‘kiss’ (24 times). Interestingly, the figure of the

43 The word *Κόνιαλης* was incorrectly lemmatised most probably due to the fact, that it is used in song texts in vocative with a double stress *Κόνιαλή*, and the automatic spaCy lemmatiser is sensitive towards a word stress.

44 The song “Man from Konya” was recorded two times in 1933 (disc numbers: Columbia DG-449, Parlophone B-21707)

45 The song “Young Man from Konya” was recorded in 1934 (disc number: HMV AO-2093).

46 More about Vamvakaris can be found in the Section 2.1, p. 10

In most other cases it is used to refer to the black eyes (*μαύρα μάτια*) or the black hair (*μαύρα μαλλιά*) of a woman.

The word cloud of the Vamvakaris subcorpus also includes lexical items such as *μπουζούκι* ,bouzouki‘ (15 times) and *πενιά* ,stroke, strum‘ (17 times). The appearance of the bouzouki in the texts of the songs composed by Markos Vamvakaris could be expected due to the fact that he is one of the most prominent representatives of the Piraeus style of rebetiko, in which the bouzouki as an accompanying music instrument played a dominant role, along with the baglamas. In contrast, in the subcorpus of songs composed by Giannis Dragatsis, who is a representative of the Smyrna style of rebetiko, the bouzouki does not appear in song texts at all.⁴⁸

Based on the word cloud of the subcorpus of Vamvakaris, one can define topics similar to those listed for the Dragatsis subcorpus, namely love (often unrequited with following sufferings of a protagonist), vagabond (mangas) life, drug consumption and prison life. In the Vamvakaris subcorpus, the figures of a mother (58 times) and a child (27 times) appear more often compared to the subcorpus of Giannis Dragatsis (17 and 11 times respectively).⁴⁹ Also, based on the word cloud data and further deeper analysis of the text, the number of Turkish loanwords is lower in comparison to the subcorpus with songs composed by Dragatsis. The reasons for this might again be the origins of both rebetiko exponents, as Vamvakaris was from Syros, while Dragatsis from Smyrna.

Next, we will take a look at the word cloud depicting the most salient words in the dataset consisting of the songs composed by Ioanna Georgakopoulou (1920-2007), who is originally from the town of Pyrgos in the northwestern Peloponnese. She is primarily known as a singer of rebetiko, but numerous records were published under her name as her own compositions. Some rebetologists claim that these songs were not her own compositions, but were given as a present to Georgakopoulou. Independently of how one looks at this issue, we will analyse the song texts as those that appeared under her name without touching the question of authorship, which always is a difficult one in rebetiko. In any case, it is of interest to analyse the songs which appeared under the name of a female rebetiko exponent. The subcorpus with the songs composed by Ioanna Georgakopoulou contains 46 songs (1,914 tokens after removal of stop words).

The word cloud created from this subcorpus is shown in Figure 4.7. The first striking difference compared to the word clouds depicting the Dragatsis and Vamvakaris subcorpora (see Figures 4.5 and 4.6 respectively), is that the interjections, which play a prominent role in the lyrics of Dragatsis and Vamvakaris, are missing almost completely in the word cloud depicting the subcorpus of songs composed by Ioanna Georgakopoulou. The only interjections included in the word cloud are *ψι-ψι* ,puss, puss (used usually when calling a cat)⁵⁰ (7 times) and *χάιντε* ,let’s, come on‘ (8 times).

Like in the previous two subcorpora, the most frequent noun again is *καρδιά* ,heart‘ (38 times), which is also the most frequent word in the subcorpus in general.

The words that are reflected in the word cloud in Figure 4.7 refer almost exclusively to the topic of love, relationship and breakup. These are lexemes such as *αγαπώ* ,to love‘ (23 times),

48 Regarding music instruments, in the song texts of the subcorpus of Giannis Dragatsis the words *βιολί* ,violin‘ and *βιολιτζής* ,violinist‘ occur each two times.

49 The two corpora are of comparable size with Dragatsis’ corpus containing 125 songs and Vamvakaris’ corpus containing 158 songs.

50 The interjection *ψι-ψι* ,puss, puss‘ is included in the only one song composed by Georgakopoulou in the corpus, namely «Οι ζαργάνες» (‘‘The Beauties’’) recorded in 1947 (disc number: Columbia DG-6638).

Tsitsanis subcorpus, the Voyant tool Contexts was used on the dataset of Tsitsanis songs without any normalisation to look at the context of all corpus entries with the stem *μαύρ-*. There are 179 entries for the query *μαύρ-* and none of these entries refer to the use of an adjective *μαύρη* or *μαύρο* as a synonym for *χασίσι* ‘hashish’, like it is the case in some other rebetiko songs. There are two central meanings, in which the Greek adjective corresponding to the English ‘black’ was used: either referring to the appearance of a woman (e.g. *μαύρα μάτια* ‘black eyes’) or to a feeling of doom, fatalism and general pessimism (e.g. *μαύρη μοίρα* ‘black fate’, *μαύρη ώρα* ‘black hour’, *μαύρες σκέψεις* ‘black thoughts’ and *μαύρη αγκαλιά* ‘black hug’). Especially in Tsitsanis’ songs with pessimistic lyrics this adjective is frequent: for instance, in the song «Μαύρα τα βλέπω κι άραχνα» (“I See It All Dreary and Miserable”) composed by Vasilis Tsitsanis and released in 1953 (disc number: Columbia 7041). Overall, the adjective *μαύρος* (in all three grammatical genders) occurs in 74 songs of Tsitsanis, i.e. in over 10% of the songs of the subcorpus. Such a high frequency of this adjective, which is used in most cases to depict the gloomy atmosphere of a song, may be influenced by the fact that 65 of these 74 songs were recorded after World War II and either during or after the Greek Civil War. Below one can find an excerpt from the song «Παίξε Χρήστο το μπουζούκι» (“Play Bouzouki, Christos”) composed by Vasilis Tsitsanis and released in 1949 (disc number: HMV 2890, sung by Prodromos Tsaousakis and Marika Ninou).

Παίξε Χρήστο, άλλο ένα,
μαύρες σκέψεις μου περνούν,
κάποια μέρα, μες τη στράτα,
ξαπλωμένο θα με βρουν.

Play, Christos, one more (song),
My black thoughts go away,
One day, on the street,
I’ll be found lying down.

Drugs and alcohol only play a peripheral role in the subcorpus of the songs by Tsitsanis. For instance, the verb *πίνω* ‘drink’ occurs only 49 times, *κρασί/κρασάκι* ‘wine’ 30 times, *αργιλές* ‘arghileh, hookah, waterpipe’ 8 times, *λουλάς* ‘hookah, waterpipe’ only once, *χασίσι* ‘hashish’ 3 times and *φουμάρω/φουμέρω* ‘smoke’ 6 times.

As one can also see from the word cloud, the figure of a mother is also present in the lyrics of songs composed by Vasilis Tsitsanis, as there are 178 entries for the lexemes *μάνα* ‘mom’ and *μανούλα* ‘(my) sweet mom’.

The word cloud also includes the lexical item *ντουμ* ‘dum’, which occurs 126 times in the Tsitsanis dataset. In the songs it is used as a syllable when humming a tune (like also in some songs in English) and does not refer to dum-dum bullets, as they occur, for instance, in the song «Γιουσουρούμ» (“Giousouroum”) of the Greek rock musician Nikolas Asimos, in which *σφαίρες ντούμ ντούμ* ‘dum-dum bullets’ are explicitly mentioned.

To summarise, based on the word cloud in Figure 4.8, songs referring to love, relationship and mother as well as songs expressing a feeling of pessimism and fatalism of a main character seem to dominate the dataset with the songs composed by Vasilis Tsitsanis.

The word clouds for the subcorpora of the four different rebetiko composers, who represent rebetiko of different periods, show several tendencies such as a change of vocabulary to a more complex one, decrease in use of Turkish loanwords, shift of underworld themes to the periphery and a general diversification of the topics, so that these songs addressed a broader audience. Giannis Dragatsis’ songs are an example of rebetika of the Smyrna style, which dominated the genre until the mid-1930s. Markos Vamvakaris was a prominent expo-

ment of the Piraeus style of rebetiko, which started to dominate rebetiko afterwards. Ioanna Georgrakopoulou and Vasilis Tsitsanis belong to the next period of rebetiko, the so-called *laiko* period of rebetiko with the songs resembling serenades, as Tsitsanis called his compositions himself.

Overall, as it was shown via the examples presented above, word clouds facilitate a quick analysis of the topics of the *Rebetiko Corpus* data and give a first impression of its contents. Based on the performed analyses one can highlight themes related to beloved ones, unrequited love, glendi (party), death, poverty, prison life, music making, hashish consumption, mother, misery life and sufferings. Moreover, when defining topics of rebetiko songs it is important to keep in mind, that most of the songs refer to more than one topic simultaneously: for instance, one and the same song may contain references to spending time in joyful music making and wine drinking in a company of beautiful women (e.g. in the song «Με ζουρνάδες με νταούλια» (“With Zurnas and Davuls”)).⁵⁴

In conclusion, word clouds are a helpful tool for exploring rebetiko lyrics, but at the same time when using them one should keep in mind their limitations. A word, even if it often occurs in a dataset, does not give enough information without context (Temple 2019). As John Rupert Firth wrote, “You shall know a word by the company it keeps” (Firth 1957, 11). For this reason, the analysis of word clouds of the *Rebetiko Corpus* was combined with the close reading of texts. Voyant Tools enables such an approach due to the fact that its tools interact with each other. This makes it possible to not only scratch the surface of the *Rebetiko Corpus*, but also dive deeper into the data.

4.1.2 Experiments with Voyant Tools: Trends

As mentioned previously, the Voyant tool *Trends* allows to build a graph of a word’s distribution across different parts of a corpus. For this analysis the entire *Rebetiko Corpus* with lyrics of 3,772 songs (398,015 tokens) was used without any text normalisation except for that performed by Voyant Tools automatically, namely tokenisation. As the goal of this analysis is to observe the use of particular words over time, the *Rebetiko Corpus* data were divided into quinquenniums.⁵⁵ The data for the experiments in the *Trends* tool were not divided in a more fine-grained way, e.g. per year, as this tool cannot handle the data correctly and build a proper graph, if the data for some of the years within a time span are missing, which is the case with rebetiko data.⁵⁶

With the help of this tool we can observe particular trends and verify, if the assumptions we have intuitively, can also be confirmed based on a relatively large collection of rebetiko song texts. Due to the fact that rebetiko is still often connected almost exclusively with topics such as drug consumption, it is interesting to look at the frequency of use of terms connected to drug consumption over time. The graph in Figure 4.9 shows the use of four different words that are directly or indirectly related to hashish. These are *χασίσι* ‘hashish’ (82 entries), *αργιλές* ‘arghileh, hookah, waterpipe’ (157 times), *λουλάς* ‘hookah, waterpipe

54 The song «Με ζουρνάδες με νταούλια» (English “With Zurnas and Davuls”) composed by Panagiotis Toundas was recorded several times including the recording released in 1934 (disc number: Columbia DG-498, sung by Stellakis Perpiniadis).

55 The Jupyter notebook with the Python script which enables such data splitting is available on GitHub at <https://github.com/kleines-gespenst/rebetiko-corpus-and-tools> (accessed February 20, 2022).

56 For instance, there are no lyrics for the period between 1941 and 1946 due to the suspension of a record production in Greece.

(bowl)‘ (124 entries) and *μαύρος/μαύρη* ,black (masculine/feminine)‘ (397 times).⁵⁷ In the parentheses the frequency of the occurrence of these words in the entire corpus is given.

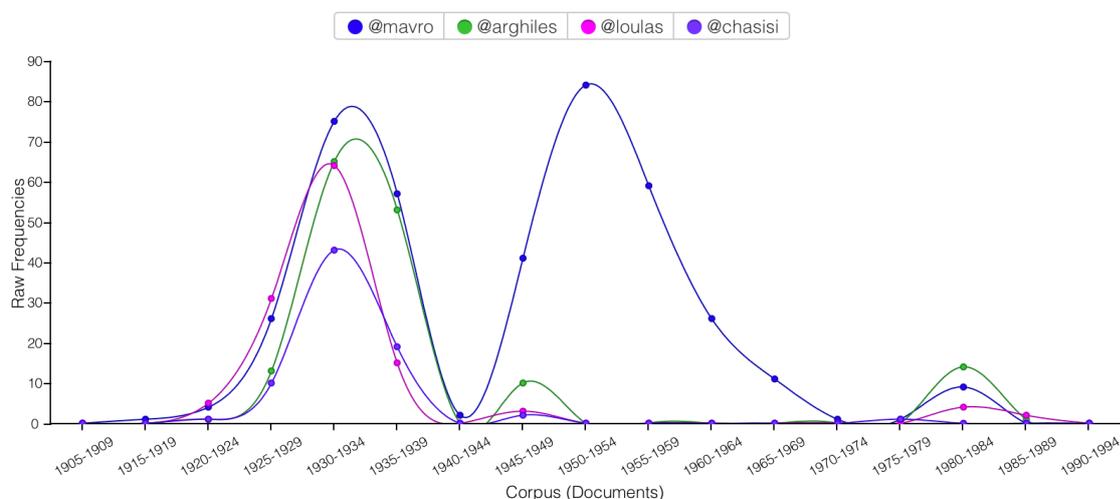


Figure 4.9: Graph Showing the Use of the Word *hashish* and its Synonyms in the *Rebetiko Corpus*. Graph Generator Source: Stéfán Sinclair and Geoffrey Rockwell, “Trends”, Voyant Tools

The tool *Trends* enables the creation of so-called categories, which are lists of words that describe one particular concept or category: for instance, all words referring to the topic of drug consumption.⁵⁸ This function makes it possible to create a line graph, which includes all possible forms of each word. The *Rebetiko Corpus* with lemmatised words was not used for this task firstly because of frequent wrong lemmatisation performed by spaCy and secondly because it is useful to be able to access the context of a keyword by clicking a particular node on a graph. The categories were named using latin letter, as this is a requirement of Voyant Tools, and English titles would be confusing. The graph in Figure 4.9 shows these categories: namely, @mavro which stands for *μαύρος/μαύρη* ,black (masculine/feminine)‘, @arghiles for *αργιλές* ,arghileh, hookah, waterpipe‘⁵⁹, @loulas for *λουλάς* ,hookah, waterpipe (bowl)‘ and @chasisi for *χασίσι* ,hashish‘.

The graph in Figure 4.9 shows a tendency, which a rebetologist would probably expect. There is a very sharp rise in the use of all four words referring to hashish around 1930 and a sharp decline afterwards. The most frequently occurring word in this period of these four is *μαύρος/μαύρη* ,black (masculine/feminine)‘, but due to the fact that it is also used in its literal meaning, the real statistics for its use in reference to hashish is undoubtedly different. Thus, for instance, for the time span between 1930 and 1934 the graph shows 75 entries for @mavro, but analysing the context we find out that only 57 entries of those are in fact referring to hashish, which is still a relatively high number. The rise of the use of words referring to hashish consumption at the end of the 1920s and in the beginning of the 1930s happens for two reasons: namely, that active record production in Greece began only at the end of the 1920s and at that time rebetiko songs referring to hashish smoking were popular (see, for instance, Section 2.4). The sharp decline around the mid-1930s is caused

⁵⁷ The form *μαυράκι* ,black (diminutive)‘ is also included in the query.

⁵⁸ Voyant Tools also offers a search for several words/subwords as a single term (e.g. by entering *χασίσι*|χασισάκι** in the Voyant search mask), but if the list of different forms is long, it is easier to create a category.

⁵⁹ The category @arghiles includes, for instance, the whole paradigm of the word *αργιλές* ,arghileh, hookah, waterpipe‘ as well as forms like *αργιλεδάκι* ,sweet/little arghileh‘.

by the establishment of the dictatorship of Metaxas and the following censorship, so that the entire underworld vocabulary almost completely disappeared from the rebetiko songs, and songs referring to hashish could not be recorded anymore. Therefore, such songs could only continue to exist in rebetiko venues. The complete drop of all four lines between 1940 and the mid-1940s is caused by the suspension of record production in Greece in general. Noticeably, around the mid-1940s we observe a jump of the lines corresponding to the occurrence of the words *χασίσι* ,hashish‘, *αργιλές* ,arghileh, hookah, waterpipe‘ and *λουλάς* ,hookah, waterpipe (bowl)‘. The reason for this increase is that in 1946, when the record production in Greece was restarted, there was no censorship for a few months. The @mavro line corresponding to the number of entries for *μαύρος/μαύρη* ,black (masculine/feminine)‘ after the restoration of record production grows again very quickly and reaches its peak around the mid-1950s. Similar to the analysis of the Tsitsanis subcorpus (see Section 4.1.1), a deeper analysis of the keywords and their contexts shows that this lexeme was used in its literal meaning (especially depicting a pessimistic mood of the song texts in phrases such as *μαύρη γη* ,black ground‘) and not metaphorically, referring to hashish. For instance, the highest peak of the @mavro line around 1955 shows 84 entries, of which 8 entries refer to the appearance of a woman, while the remaining 76 refer to the topic of miseries of life and fatalism.

The complete decline in the use of even the lexeme *μαύρος/μαύρη* ,black (masculine/feminine)‘ with its neutral meaning, i.e. not defining hashish, in the period between 1970 and 1974 may be caused by the small number of songs included for this period in the *Rebetiko Corpus*, namely only 20, so that the probability of the appearance of the word *μαύρος/μαύρη* ,black (masculine/feminine)‘ in this small subset of songs is very low. After 1975, all four words again appear in the song lyrics, which is caused by the end of the strict censorship due to the fall of the Junta in 1974. Still, the increases shown in the graph are relatively small due to the fact that also for the remaining time span until 1995 the *Rebetiko Corpus* contains a relatively small number of songs: 50 songs for the period between 1975 and 1979, 38 songs between 1980 and 1984, 29 songs between 1985 and 1989, and only two songs between 1990 and 1994. The small number of songs in the corpus does not mean that in this period a small number of rebetiko songs was recorded. The number is in fact on the contrary very high, but the corpus data are based only on data available at *rebetiko.sealabs.net*. In the *Rebetiko Corpus*, the first song with a reference to hashish that was recorded after the fall of the military junta in Greece is «Το βαπόρι απ’ την Περσία» (English “The Ship from Persia”), composed by Vasilis Tsitsanis and released in 1977 (disc number: Columbia 70529). This song is not concerned with the process of hashish smoking, but narrates a real story that took place in 1977 near the Isthmus of Corinth, where a ship coming from Beirut was caught with 11 tons of cannabis on board.⁶⁰

The words presented in the graph in Figure 4.9 are not all terms that occur in the corpus in reference to hashish or the waterpipe. Other possible synonyms of hashish or words that are related to *χασίσι* ,hashish‘, which occur in the song texts of the *Rebetiko Corpus*, are *νταμίρα* ,hashish‘ (55 entries), *σύρμα* ,wire, prison, hash‘ (19 entries), *καλάμι* ,reed, argheleh‘ (54 entries) and *τσικά* ,piece of hashish (cf. Spanish *chica*)‘ (14 entries). The songs included in

60 More information about this incident can be found in *Οι λμηνικές αρχές εντοπίζουν και κατάσχουν 11 τόνους χασίς επί του υπό Κυπριακή σημαία M/S «Γκλόρια» στον Ισθμό της Κορίνθου* [The Port Authorities find and confiscate 11 tonnes of hashish on board of a vessel MS “Gloria” under the Cyprus flag in the Isthmus of Corinth.] *Cyprus Times*, January 7, 2022. Available at <https://cyprustimes.com/san-simera/7-ianoyarioy-1977-to-vaporir-ap-tin-persia-pianetai-stin-korinthia-11-tonoi-chasis-sto-ypo-kypriaki-simeia-m-s-gkloria/> (accessed February 14, 2022)

the corpus also contain references to other substances such as *κοκαΐνη* ,cocaine‘ (11 entries) and *ηρωΐνη* ,heroin‘ (5 entries). The rebetiko songs included in the *Rebetiko Corpus* also contain words defining alcohol such as *κρασί* ,wine‘ (257 entries), *ούζο* ,ouzo‘ (90 entries), *μπύρα* ,beer‘ (19 entries)⁶¹ and *τεκίλα* ,tequila‘ (6 entries).

With the help of the *Trends* tool, we can also look at the use of terms defining a ‘care-free man/woman of the street’⁶² such as *μάγκας* ,mangas, vagabond, wide boy‘ and words related to it, which are used in the context of rebetiko as synonyms (e.g. *αλάνης/αλανιάρης* ,vagrant, bum‘). The graph in Figure 4.10 shows the frequency of the occurrences of five words referring to the concept of a ‘carefree man/woman of the street’ across the *Rebetiko Corpus*. These are the following five nouns: *μάγκας/μάγκισσα* ,mangas, vagabond (masculine/feminine)‘ (659 entries), *αλάνης/αλανισσα* ,vagrant, bum (masculine/feminine)‘ (307 entries), *ρεμπέτης/ρεμπέτισσα* ,outlaw, vagabond (masculine/feminine)‘ (25 entries), *μόρτης/μόρτισσα* ,scoundrel, bum (masculine/feminine)‘ (154 entries) and *ντερβίσης* ,dervish, hashish smoker, fine man (cf. *λεβέντης*)‘ (144 entries).⁶³ Both masculine and feminine nouns defining male and female characters in rebetiko song lyrics were taken for the analysis. As for the previous analysis on the use of words related to hashish, five categories were created, so that the entire paradigm for each word is covered, and each category includes both masculine and feminine forms.⁶⁴ The graph shows five categories, namely @mangas⁶⁵, @alanis, @mortis, @dervisis and @rembetis.

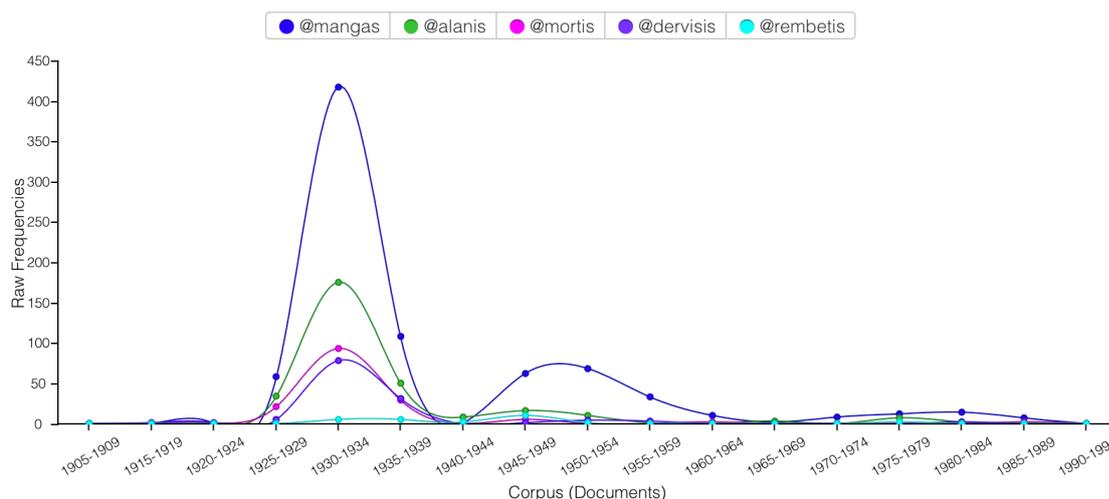


Figure 4.10: Graph Showing the Use of the Word *mangas* and its Synonyms in the *Rebetiko Corpus*. Graph Generator Source: Stéfán Sinclair and Geoffrey Rockwell, “Trends”, Voyant Tools

- 61 The *Rebetiko Corpus* includes in fact 69 entries for the word *μπύρα*, but 50 entries refer to a kind of beer bar called *byra* and not to an alcoholic drink itself.
- 62 The translation of concepts such as *mangas* into English is difficult, because depending on the context these concepts may have different meanings.
- 63 In parentheses the frequency of the occurrence of these words in the entire corpus is given.
- 64 The noun *ντερβίσης/δερβίσης* ,dervish, hashish smoker, fine man‘ does not have a feminine form, which might be caused by the main meaning of this word. Not only the word *dervish* entered rebetiko, but also a *tekke*, which in rebetiko songs defines a hashish den and not a dervish lodge. So the word *dervish*, which in a neutral context usually defines a member of the Sufi fraternity, was introduced to rebetiko due to a particular mystic atmosphere of a hashish den, where its frequenters used to gather not only to smoke hashish, but also to make music.
- 65 The category @mangas includes, for instance, forms such as *μάγκας*, *μάγκα*, *μάγκες*, *μαγκάκι*, *μαγκάκια*, *μαγκίτης*, *μαγκίτη*, *μαγκίτες*, *μαγκίτισσα*, *μαγκίτισσας*, *μαγκίτισσες*, *μάγκισσα*, *μάγκισσας* and *μάγκισσες*.

The nouns presented in the graph are the most frequent ones that denote a ‘carefree man/woman of the street’ in the *Rebetiko Corpus*, with the exception of *ρεμπέτης/ρεμπέτισσα*, ‘outlaw, vagabond (masculine/feminine)’, which was included in the analysis, because it is directly connected to the name of the music genre *rebetiko*. As the graph shows, the most frequent term in the corpus out of these five is *μάγκας/μάγκισσα*, ‘mangas, vagabond (masculine/feminine)’. Its use reaches its peak in the period between 1930 and 1934 with 417 occurrences. It disappeared around 1940 due to the suspension in record production in Greece between 1941 and 1946. After the resumption of record production its use goes up again, but reaches only 68 entries in the period between 1950 and 1954. This is a substantial relative decrease of the use of the word *μάγκας/μάγκισσα*, ‘mangas, vagabond (masculine/feminine)’, as the number of songs recorded before the war and after the war, which are included in the corpus, is similar: the *Rebetiko Corpus* contains 1,962 songs recorded before the record suspension in 1941 and 1,810 songs recorded after the war, so the reason is not a generally smaller number of recorded songs in the period after the war. At the same time, one should always keep in mind the fact that a corpus cannot include all texts ever created, so that the results should always be critically viewed.

The second most frequent word of these five is *αλάνης/αλάνισσα*, ‘vagrant, bum (masculine/feminine)’, and there are 175 entries for this lexeme in the mid-1930s. After World War II, it almost disappears with only 16 entries for the period between 1946 and 1949, and only 10 between 1950 and 1954. After the mid-1950s it occurs only in 5 songs. The third most frequent word out of the five is *μόρτης/μόρτισσα*, ‘scoundrel, bum (masculine/feminine)’. It occurs 93 times in the period between 1930 and 1934, but after the resumption of record production in 1946 it almost completely disappears. Remarkably, the word *μόρτης* comes most probably from Italian *morti* < *morto*, ‘dead’ and defined originally someone who was immune to the plague and worked as a gravedigger, burying the bodies of those who died of this disease. Therefore, being originally synonymous to *απόλοιμος*, it experienced a semantic shift and became synonymous with *μάγκας* meaning a vagabond or a tough guy (Sarandakos 2015).

The word *ντερβίσης*, ‘dervish, hashish smoker, fine man’ occurs relatively frequent in the mid-1930s (78 entries between 1930 and 1934), but like *αλάνης* and *μόρτης* it occurs only a few times after 1946. Due to its reference to hashish, this word in fact almost disappears from the records already in 1936, when the censorship was established. Between 1937 and 1941 only four songs include the word *ντερβίσης*, and in these few records the word appears only in the spoken interjections and not in the song text itself.⁶⁶ The lexeme *ρεμπέτης/ρεμπέτισσα*, ‘outlaw, vagabond (masculine/feminine)’ occurs only 25 times in the entire corpus.

To summarise, the words *μάγκας/μάγκισσα*, *αλάνης/αλάνισσα*, *ντερβίσης* and *μόρτης/μόρτισσα* are frequently used from the mid-1920s until the mid-1930s, but after 1946, with the exception of *μάγκας/μάγκισσα*, they all occur only a few more times. The word *ρεμπέτης/ρεμπέτισσα* is in general very rare in the entire corpus.

In the previous graph in Figure 4.10 feminine and masculine forms of nouns were treated as one term. It is also of interest to look at the correlation between feminine and masculine nouns defining ‘carefree man/woman of the street’ in the *Rebetiko Corpus*. As already mentioned, not only a male figure in rebetiko songs can be depicted as a freedom-loving person that refuses traditional values, but also a female figure (see Section 4.1.1). The graph in Fig-

⁶⁶ For instance, in the song «Μικρός αρραβωνιάστηκε» (English “I Got Engaged Young”) composed by Markos Vamvakaris (disc number: Odeon GA-7041, recorded in 1937)

ure 4.11 shows the separate use of some masculine and feminine nouns denoting a ‘careless person of the street’ in the *Rebetiko Corpus*. For this analysis, six different nouns were chosen, which exist both in feminine and masculine, so that the distribution of masculine and feminine forms is comparable: *μάγκας* (645 entries) / *μάγκισσα* (14 entries) ,mangas, vagabond (masculine/feminine)’, *μόρτης* (62 entries) / *μόρτισσα* (92 entries) ,scoundrel, bum (masculine/feminine)’, *ρεμπέτης* (23 entries) / *ρεμπέτισσα* (2 entries) ,outlaw, vagabond (masculine/feminine)’, *μποέμ* (4 entries) / *μποέμισσα* (20 entries) ,bohemian (masculine/feminine)’, *αλάνης* (213 entries) / *αλάνισσα* (94 entries) ,vagrant, bum (masculine/feminine)’ and *βλάμης* (57 entries) / *βλάμισσα* (17 entries) ,blood brother, fraterniser (masculine) / blood sister, girlfriend (feminine)’.

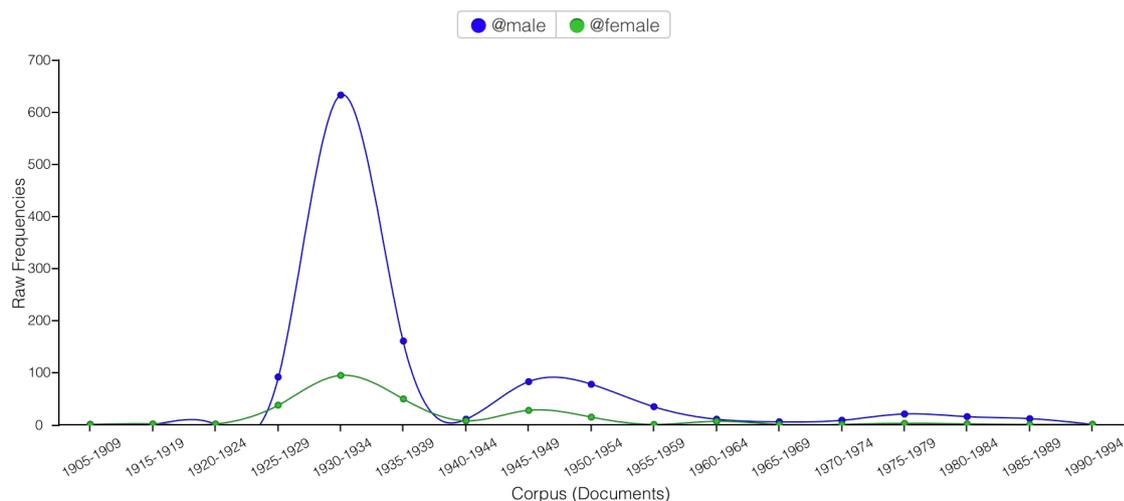


Figure 4.11: Graph Showing the Use of Masculine and Feminine Nouns denoting a ‘Careless Person of the Street’ in the *Rebetiko Corpus*. Graph Generator Source: Stéfan Sinclair and Geoffrey Rockwell, “Trends”, Voyant Tools

As we can see in the graph, the use of masculine forms of nouns related to the concept of *mangas* is higher than the use of feminine forms. Between 1930 and 1934 there are 632 entries for the masculine forms of the above-mentioned nouns and 94 entries for feminine forms. In the *Rebetiko Corpus* the most frequent noun denoting a male ‘careless person of the street’ is *μάγκας* ,mangas, vagabond (masculine)’ (645 entries), and the most frequent noun denoting a female ‘person of the street’ is *αλάνισσα*⁶⁷ ,vagrant, bum (feminine)’ (94 entries).

This was an example for a possible analysis of the synonyms of the lexeme *μάγκας* ,mangas, vagabond, wide boy’ and *ρεμπέτης* ,outlaw, vagabond’, but many extensions are possible.⁶⁸ A deeper analysis of the different terms that are used in rebetiko song lyrics to define male and female characters is outside of the scope of this thesis but definitely deserves attention in future works on rebetiko. The frequency of the word *μάγκας* testifies once more the observations presented in Section 2.1 that not *ρεμπέτης* but *μάγκας* is the main figure of rebetiko songs.

⁶⁷ The search in the corpus for the term *αλάνισσα* also included search for forms such as *αλάννα* and *αλανιάρρα* .

⁶⁸ For instance, one could also include in this analysis the word *αλήτης* (73 entries) / *αλήτισσα* (2 entries) ,loafer, bum (masculine/feminine)’.

In conclusion, Voyant Tools provides an easy to use interface for text analyses, which combine both close and distant reading due to the fact that the tools interact with each other. The experiments with the *Rebetiko Corpus* data have shown that even from such a generalised representation as a word cloud, which has its limitations, one can learn a lot about the data, especially with the help of additional Voyant tools such as *Contexts*, which shows a keyword in context. The *Trends* tool enables us to observe particular trends over time, such as the presented visualisations regarding words denoting hashish and the main character of rebetiko song lyrics. One could think of many other experiments, but this is beyond the scope of this work.

4.2 DATA ANALYSIS WITH CUSTOM TOOLS WRITTEN IN PYTHON

In general, as it was mentioned in the previous section, there are multiple tools for text analysis that are readily available, easy to use, and enable quite sophisticated analyses of texts. At the same time, one is bound to the functionality of a particular tool, and tools which are suitable to some extent for some experiments, can still show some limitations. For instance, in Section 4.1.2 the Voyant tool *Trends* was used to build graphs of frequencies of occurrence of particular terms over time. The visualisation of the trends is based on a smoothed line plot, which is helpful for analysing general tendencies, but makes it difficult to detect sudden changes such as the disappearance of a particular term at a particular point in time. The implementation of custom tools enables not only to write a custom-made algorithm that suits a particular task such as the visualisation of the frequency of words occurring over time, but also approach much more complex problems of text analysis by using, for instance, machine learning transformer models.

4.2.1 Challenges for NLP Tasks with the *Rebetiko Corpus*

When working with the *Rebetiko Corpus*, numerous challenges for the processing of the song lyrics were encountered, which make it difficult to apply even the limited range of language processing tools for standard Modern Greek. Some of these challenges were already mentioned in Section 4.1. For instance, available lemmatisers as well as other elements of a standard NLP pipeline such as a part-of-speech tagger for Greek do not exhibit good performance when processing the *Rebetiko Corpus* data. This is caused mostly by the models used in the processing pipeline, which were trained or optimised on texts written in standard Greek, so that they are not able to always correctly process song texts which show numerous peculiarities including specific differences in syntax and grammar. Some of these tools do not include a pre-trained model for some of the steps of the processing pipeline, but are based on rules (like the default lemmatiser for Greek in spaCy). It is difficult to describe highly inflected languages such as Greek with a set of rules, and if one deals with texts with, for instance, non-standard spelling and prosody⁶⁹, the performance of such a rule-based lemmatiser will be even poorer (see Section 4.1.1, p. 69).

⁶⁹ In Greek, word accents are indicated in writing, and in song texts one often deals with non-standard accent marking, which is often caused by the necessity to match the metre (e.g. *ἀμᾶν* instead of standard *ἀμάν*).

While working with the *Rebetiko Corpus*, four different lemmatisers were tested. In addition to the spaCy lemmatiser, I also experimented with tools such as UDPipe⁷⁰, Sketch Engine⁷¹, as well as the Stanza lemmatiser.⁷² Unfortunately, none of the four lemmatisers performs lemmatisation well for the rebetiko song lyrics. As already mentioned, there are multiple reasons for the sub-par performance of the NLP pipeline, e.g. untypical word order, missing punctuation marks, and a very large number of contractions. Some problems of the lemmatisation for Greek with spaCy were already presented in Section 4.1.1, thus below an exemplary lemmatisation of rebetiko lyrics with UDPipe will be presented.

Id	Form	Lemma	UPosTag	XPosTag	Feats	Head	DepRel	DepS	Misc
# generator = UDPipe 2, https://lindat.mff.cuni.cz/services/udpipe									
# udpipe_model = greek-gdt-ud-2.6-200830									
# udpipe_model_licence = CC BY-NC-SA									
# newdoc									
# newpar									
# sent_id = 1									
# text = Αντί σχολιό μου πάγαινα μες του Καραϊσκάκη έπινα διάφορα πιοτά να μάθω μπουζουκάκι									
1	Αντί	αντί	ADP	ADP	_	2	case	_	TokenRange=0:4
2	σκολιό	σκολιό	NOUN	NOUN	Case=Acc Gender=Neut Number=Sing	8	obl	_	TokenRange=5:11
3	μου	μου	PRON	PRON	Case=Gen Gender=Masc Number=Sing Person=1 Poss=Yes PronType=Prs	2	nmod	_	TokenRange=12:15
4	πάγαινα	πάγαινα	NOUN	NOUN	Case=Nom Gender=Neut Number=Plur	2	nmod	_	TokenRange=16:23
5	μες	μες	ADP	ADP	_	7	case	_	TokenRange=24:27
6	του	ο	DET	DET	Case=Gen Definite=Def Gender=Masc Number=Sing PronType=Art	7	det	_	TokenRange=28:31
7	Καραϊσκάκη	Καραϊσκάκη	PROPN	PROPN	Case=Gen Gender=Masc Number=Sing	4	nmod	_	SpacesAfter="\s\^\n" TokenRange=32:42
8	έπινα	έπινο	VERB	VERB	Case=Acc Gender=Neut Number=Plur	0	root	_	TokenRange=45:50
9	διάφορα	διάφορος	ADJ	ADJ	Case=Acc Gender=Neut Number=Plur	10	amod	_	TokenRange=51:58
10	πιοτά	πιοτός	NOUN	NOUN	Case=Acc Gender=Neut Number=Plur	8	obj	_	TokenRange=59:64
11	να	να	AUX	AUX	_	12	aux	_	TokenRange=65:67
12	μάθω	μάθω	VERB	VERB	Aspect=Perf Mood=Ind Number=Sing Person=1 VerbForm=Fin Voice=Act	10	acl	_	TokenRange=68:72
13	μπουζουκάκι	μπουζουκάκι	NOUN	NOUN	Case=Acc Gender=Neut Number=Sing	12	obj	_	SpaceAfter=No TokenRange=73:84

Figure 4.12: Example of a Lemmatisation Performed by UDPipe. Source: <http://lindat.mff.cuni.cz/services/udpipe/> (accessed February 18, 2022)

Figure 4.12 shows a text passage processed by UDPipe of the two lines *αντί σχολιό μου πάγαινα μες του Καραϊσκάκη έπινα διάφορα πιοτά να μάθω μπουζουκάκι* (“Instead of going to school I went to Karaiskaki (coffeeshop at Karaiskaki square), I had there some drinks to learn bouzouki”)⁷³ from the song by Michalis Genitsaris «Εγώ μάγκας φαινόμευνα» (“I looked like a mangas”).⁷⁴ While some of the words obtain correct part-of-speech tags and gram-

70 More information on UDPipe can be found at <http://lindat.mff.cuni.cz/services/udpipe/> (accessed February 18, 2022).

71 More information on the Sketch Engine corpus manager and text analysis software as well as its components can be found at <https://www.sketchengine.eu> (accessed February 18, 2022).

72 More information on the Stanza NLP library can be found at <https://stanfordnlp.github.io/stanza/> (accessed February 18, 2022).

73 In this song, Genitsaris refers to the coffeeshop of Giorgos Batis *Zorz Bate* (*Ζώρζ Μπατέ*) located near Karaiskaki square in Piraeus. As a child, Genitsaris used to go to the coffeeshop of Batis instead of school and play bouzouki and baglamas there.

74 The song «Εγώ μάγκας φαινόμευνα» (“I looked like a mangas”) by Michalis Genitsaris was recorded in 1938 (disc number: Columbia DG-6312).

matical features as well as correct lemmata (e.g. *διάφορα* < *διάφορος* ,different‘), four words out of 13 obtain false labels and lemmata. Remarkably, all three verbs used in this sentence were lemmatised incorrectly, and one of them was even tagged as a noun. The form *έπινα* is a standard Greek first-person singular imperfect form of the verb *πίνω* ,drink (verb)‘, but due to the fact that a comma between the two lines before *έπινα* as well as a full stop at the end of a ‘sentence’ are missing, the processing pipeline treats the verb as a noun/adjective and gives it corresponding grammatical features, even though the part-of-speech tag is correct. This occurs because most tools such as the lemmatiser, are context sensible, which is of crucial importance for accurate text analysis, and thus punctuation marks also play an important role. If a comma before *έπινα* and a full stop at the end of these two lines are added, UDPipe treats *έπινα* as a verb, assigning it the correct grammatical features (see Section 6.3 in appendix), but still does not manage to perform correct lemmatisation. Therefore, *έπινα*, which is treated by the UDPipe lemmatiser as a plural neuter noun/adjective in accusative, is lemmatised to **έπινος*. The form *πάγαινα* is a non-standard Greek form of the verb *πηγαίνω* ,go‘ in imperfect (first-person singular). Due to the fact that UDPipe does not know such a lexical item, it is treated as a noun, due to its similarity to a Greek feminine noun, which is understandable but of course wrong. The next incorrectly lemmatised word is *πιotá*, which is also a non-standard variant of the plural neuter accusative form *ποτά* < *ποτό* ,drink (noun)‘, so that its incorrect lemmatisation is also understandable. The last incorrectly lemmatised word is *μάθω*, which is a standard Greek first-person singular form in present tense (perfective aspect) of the verb *μαθαίνω* ,learn‘. To summarise, the UDPipe pipeline struggles the most with the correct identification of verbs and their lemmatisation. It is also with the lemmatisation of verbs that spaCy struggles the most, which is probably caused by the fact that the Greek language has a very complex verb inflectional system. This example also shows the important role that the punctuation marks play in the interpretation of text by a machine. It is a difficult question, how to handle this problem, as adding punctuation marks manually would be very time consuming and also change the form of the song lyrics. The *Rebetiko Corpus* lyrics are based on the data available at *rebetiko.sealabs.net*, but the missing punctuation in this data source does not mean that the songwriter’s version was also without punctuation.⁷⁵ The Stanza and the Sketch Engine lemmatisers exhibit similar problems with rebetiko song lyrics. Due to the fact that the spaCy pipeline generally showed a better performance for the *Rebetiko Corpus* and spaCy is a user-friendly Python NLP library with wide functionality, it was chosen as the main NLP package used in the scope of this thesis. Next, some further characteristics of rebetiko song lyrics will be presented, which make lyrics preprocessing more difficult.

Elision, which is a common phenomenon in colloquial Modern Greek, is also frequent in rebetiko lyrics (see Section 4.1). For instance, we observe vowel deletion at word boundaries when the first word ends with a vowel and the second word starts with a consonant, and the first one loses its final vowel (apocope) (e.g. *δώσε μου* > *δωσ’ μου* ,give me‘). A special case of vowel deletion is called hiatus resolution⁷⁶, which refers to vowel deletion caused by the occurrence of two vowels at word boundaries, i.e. the first word ends on a vowel and the second one begins with a vowel, and one of the two vowels is deleted. For example, in the

75 Generally, the question of punctuations is a difficult one, as in most cases we do not have access to the song lyrics as written by the songwriter. Moreover, adding punctuation may influence the rhythm of a song. This issue becomes even more involved due to the fact that some rebetiko songs, and songs in general, use ‘ungrammatical’ constructions.

76 On hiatus resolution in Modern Greek see (Baltazani 2006).

case of *έκθλιψη*, the first word loses its final vowel, e.g. *για αυτό > γι' αυτό*, 'that's why'. On the other hand, in the case of an aphaeresis the second word loses its initial vowel, e.g. *το έφερα > το 'φερα*, 'I brought it'. In rebetiko songs, we also observe a more complex vowel reduction such as crasis, which is, for instance, present in the phrase *σούδωσα διαζύγιο* instead of *σου έδωσα διαζύγιο*, 'I divorced you'. These types of vowel deletion often occur in rebetiko song lyrics due to the requirements of metre, as it is typical for poetry and song lyrics in general (Holton et al. 2019, 80). While it is desirable that an NLP algorithm treats forms such as *'φερα* and *έφερα* in the same way, the preprocessing task of the substitution of deleted vowels is not trivial. Moreover, such elisions are often combined with the shift of a word stress, which in Greek is indicated in writing, so that algorithms, which are usually sensitive to an accent mark, will with a high probability not be able to identify a word correctly (e.g. the second-person personal pronoun in genitive *σου* and *σου*). While for English there are tools that can deal with the task of expanding contractions (Lukei 2019), there are no such tools for Modern Greek.

Another characteristic of rebetiko song lyrics refers to an augment in past tenses, which is used much more often in rebetiko song lyrics than in standard Modern Greek (e.g. the first-person plural imperfect form *εκάναμε* instead of *κάναμε < κάνω*, 'do' and the second-person singular aorist form *εφαρμάκωσες* instead of *φαρμάκωσες < φαρμακώνω*, 'poison (verb)'). At the same time it does not seem to be used as an additional grammatical marker of a past tense, but similar to pronouns like *ταμένα* and *μένανε* (see Section 4.1, p. 72), it is used to match the metre. Such a deviation from verbal forms of standard Greek also influences the automatic text analysis due to the fact that the existing models were trained on standard Greek, which also includes the standard pattern of conjugation.

Dealing with rebetiko lyrics, one also finds many repetitions of hemistichs with subwords like in the line «*σ' αυτόν τον ψε-, τον ψεύτικο τον κόσμο*» ('in this fa-, in this fake world') from the song «*Απόψε να μην κοιμηθείς*» ('Don't Sleep Tonight') by Vasilis Tsitsanis.⁷⁷ For an NLP algorithm that was trained on a different kind of data, it is impossible to 'understand' that the syllable *ψε-* refers to a masculine adjective in accusative *ψεύτικο < ψεύτικος*, 'fake', and due to the fact that most NLP tools take into account also the context, it may lead to misinterpretations of the entire hemistich.

Due to the fact that lexical stress tends to align with the musical beats, some words change their typical stress in songs: for instance, in the previously presented line «*αντί σκολιό μου πάγαινα μες του Καραϊσκάκη*» ('Instead of going to school I went to Karaiskaki (coffeeshop at Karaiskaki square)') with a fifteen-syllable metre from the well-known autobiographic song by Michalis Genitsaris «*Εγώ μάγκας φαινόμευνα*» ('I looked like a mangas'), one observes a shift of lexical stress from the diphthong *-ει* to the final *-ο*. Due to the fact that lexical stress is expressed explicitly in Modern Greek, its shift influences the way an algorithm observes and interprets such an information. The strong relationship between linguistic prosody and musical rhythm is a known phenomenon which was studied, for instance, on American hip hop (see (Tait/Tabain/Sykes 2014)). Moreover, in this line the word *σκολιό* in comparison to standard Greek *σχολείο* shows a phonological change from *-χ* to *-κ* and a fusion of diphthong *-ει* to *-ι*.

⁷⁷ The song «*Απόψε να μην κοιμηθείς*» ('Don't Sleep Tonight') composed by Vasilis Tsitsanis was recorded in 1939 and sung by Rita Abatzi and Stellakis Perpiniadis (disc number: HMV 2538).

Moreover, in this line we also see the construction *μες του Καραϊσκάκη*⁷⁸ (“to Karaiskaki”), a phrase including a possessive construction, in which an attribute denoting a possessee is missing. Most probably in this line Genitsaris refers to the Karaiskaki square in Piraeus close to which back then the coffeeshop of Giorgos Batis was located, which he used to visit as a child. Thus, we could reconstruct the phrase *«μες του Καραϊσκάκη»* to *«μεσ’ στου Καραϊσκάκη την πλατεία»* (“to Karaiskaki square”), although Genitsaris meant not the square itself, but the venue located there. We find similar constructions also in many other rebetiko songs: for instance, in the song titles *«Μες του Συγγρού τη φυλακή»* (“In Syngrou Prison”) composed by Andonis Dalgas⁷⁹ and *«Μες του Μάνθου τον τεκέ»* (“Inside Manthos’ Tekes”) composed by Kostas Tzovenos.⁸⁰ The more neutral construction with the same meaning would be *«μεσ’ στην πλατεία του Καραϊσκάκη»* (“to/at Karaiskaki square”), so that in possessive constructions in Greek, an argument in genitive usually follows a head-noun in nominative (NG). Such an emphatic word order (GN) inside a possessive construction like in the song “I Looked Like a Mangas” by Genitsaris occurs quite often in verses in general, but is not typical for other types of texts. Interestingly, both orders (GN and NG) inside a possessive phrase were also used in Ancient Greek.⁸¹ Such a construction with untypical word order and the missing preposition *σε* ,in‘ before the definite article in genitive *του*, like in *μες (σ)του Καραϊσκάκη*, also has an influence on automatic text analysis. For instance, if the goal is to extract all named entities (e.g. persons and locations), the algorithm will struggle with the named entity recognition (NER). When the spaCy NER component⁸² is applied to the lyrics of the song *«Εγώ μάγκας φαινόμουνα»* (“I Looked Like a Mangas”) by Genitsaris, it is unable to identify *Καραϊσκάκη* (Karaiskaki) as a person, nor as a location or as a geopolitical entity. In contrast, if the sentence is entered as *«Αντί σκολιό μου πάγαινα στην πλατεία Καραϊσκάκη, έπινα διάφορα πιστά να μάθω μπουζουκάκι»*, the algorithm identifies *πλατεία Καραϊσκάκη* as a geopolitical entity. Due to the fact that the spaCy NER pipeline component was trained on standard ‘neutral’ Greek texts, it does not know how to treat constructions like in this song, although they appear quite often in songs in general (cf. also frequent restaurant names such as *«Στου Θανάση»* (“By Thanasis”)).

These were some of the challenges that one faces when applying NLP techniques to rebetiko lyrics. Rebetiko song lyrics show deviations from the standard Modern Greek phonology, prosody, grammar and syntax including, for instance, elements such as untypical word order and word accent marking. Moreover, rebetiko lyrics are rich in metaphors, which also makes automatic text analysis a more challenging task. In conclusion, the non-standard language of rebetiko songs as well as a very limited number of tools available for Greek in comparison to such high-resource languages as English influence the process of automatic lyrics analysis.

78 In most rebetiko song recordings including the construction *μεσ’ στου*, the preposition *σε* ,in‘ is missing. It can also be seen in the labels of records, where titles are included.

79 The song *«Μες του Συγγρού τη φυλακή»* (“In the Syngrou Prison”) composed by Andonis Dalgas was recorded several times including the recording made in 1932 (disc number: Parlophone B-21642).

80 The song *«Μες του Μάνθου τον τεκέ»* (“Inside Manthos’ Tekes”) was recorded several times including the recording made in 1933 (disc number: HMV AO-2074).

81 On word order in possessive constructions in Ancient Greek see (Viti 2008).

82 More about the spaCy NER component can be found at <https://spacy.io/usage/linguistic-features> (accessed February 19, 2022).

4.2.2 Custom Frequency Graph Over Time

As mentioned in the introduction to Section 4.2, although there is a range of helpful available text analysis tools, their functionality is often limited. For instance, in Section 4.1.2 the Voyant tool *Trends* was used to build graphs of the frequency of occurrence of a particular term over time. The visualisations in the *Trends* tool are heavily smoothed, which is helpful for analysing general tendencies, but hides local phenomena. In some cases, this might be desired, as sometimes these phenomena are just artefacts of the data, but in other cases the smoothing hinders the analysis of these local phenomena. Moreover, as mentioned previously, the Voyant tool *Trends* cannot handle the data correctly and build a proper graph, if the data for some of the years within a time span are missing, which is the case with *rebetiko* data, therefore the *Rebetiko Corpus* data were divided into quinquenniums. This makes the *Trends* graphs even more inaccurate for some experiments. For instance, in the graph in Figure 4.9 (Section 4.1.2), which shows the occurrence of words referring to hashish across the corpus, we observe a smooth decline in the time span from 1930-1934 to 1940-1944. The smoothing hides the fact that words referring to hashish consumption actually disappeared in 1936 with the establishment of the dictatorship of Metaxas, and the fact that record production was suspended completely in Greece between 1941 and 1946.

To solve this problem, I have written a Python script⁸³, which splits the data on a year-wise basis and builds a graph with term distributions over time without smoothing. This script allows to define a list of words that define a term. For instance, to find all occurrences of the term *λουλάς* ‘hookah, waterpipe’, the following words were included in the search: *λουλάς*, *λουλά*, *λουλάδες*, *λουλαδιές*, *λουλάδων*, *λουλα*, *λουλαδάκι* (other forms of the word do not occur in the corpus).

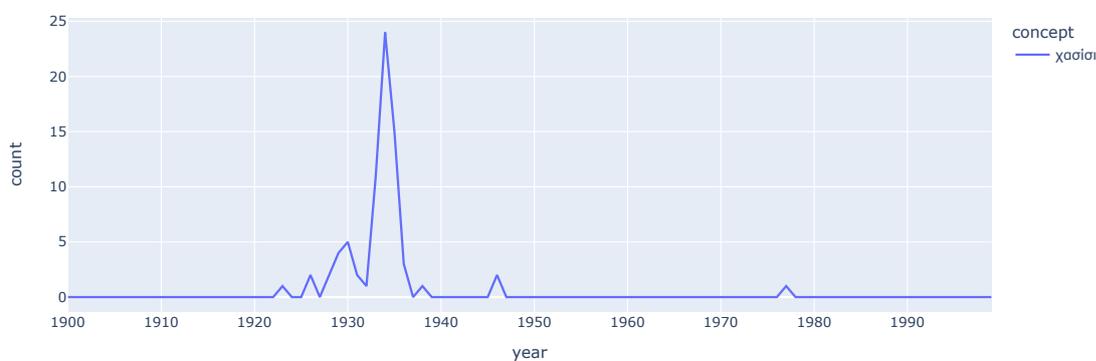


Figure 4.13: Custom Graph: Use of the Word *hashish* in the *Rebetiko Corpus*

The graphs in Figures 4.13 and 4.14 demonstrate the use of the words *χασίσι* ‘hashish’ and *λουλάς* ‘hookah, waterpipe’ over time. The terms were not presented in one graph due to the fact that they have very similar curves with all spikes being almost identical. In comparison to the graph in Figure 4.9 (Section 4.1.2) one can clearly see in these graphs both the introduction of censorship in 1936 and the record suspension in 1941. The graphs, which

⁸³ The Jupyter notebook with this Python script is available on GitHub at <https://github.com/kleines-gespenst/rebetiko-corpus-and-tools> (accessed February 20, 2022).

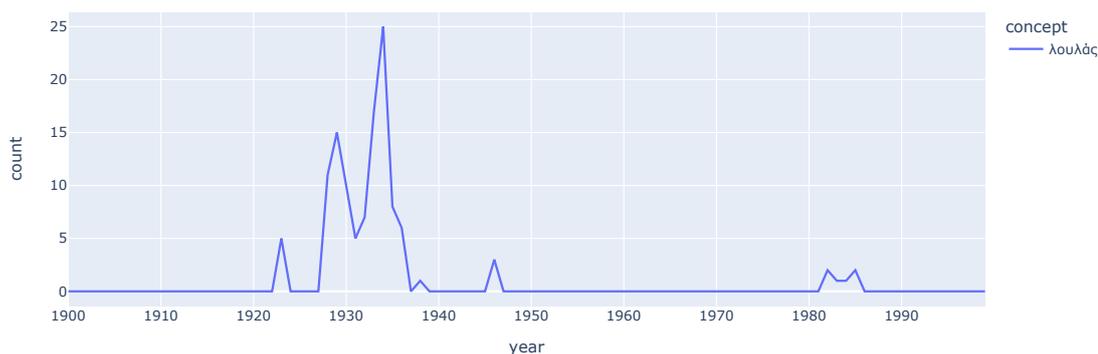


Figure 4.14: Custom Graph: Use of the Word *hookah* in the *Rebetiko Corpus*

are built with the help of the Python script, are interactive, so that one can see exactly, how many entries correspond to each year and therefore can also better interpret peaks and low points. Thus, for instance, the drops in both graphs in the 1930s start in 1937, which one would expect due to the introduction of strict censorship. The slight rise and then again a complete decline in 1938, i.e. at times of strict censorship, is caused by several recordings made in the United States: among them the recording of the song «Μεσ του Μάνθου τον τεκέ» (“Inside Manthos’ Tekes”), which was rerecorded by Giorgos Katsaros in New York in 1938 (disc number: Orthophonic S-446-B). This song contains four different words, which stand for hashish or are related to it: namely, *χασίσι* ‘hashish’, *λουλάς* ‘hookah, waterpipe’, *μαύρο* ‘black (hash)’ and *αργιλές* ‘hookah, waterpipe, argihleh’.

The graph in Figure 4.15 shows the frequency of use of the verbs *καπνίζω* and *φουμάρω* ‘smoke’ across the corpus. As it is mentioned earlier (see Section 4.1.1, p. 82), the standard Greek verb *καπνίζω*, which stands for ‘smoke’ is hardly ever used in the rebetiko lyrics, whereas the Italian borrowing *φουμάρω* is used much more often. Its use reaches its absolute maximum in the *Rebetiko Corpus* in 1934 with 58 occurrences. The decline in its use is caused by the introduction of censorship and the following shift in topics depicted in rebetiko.

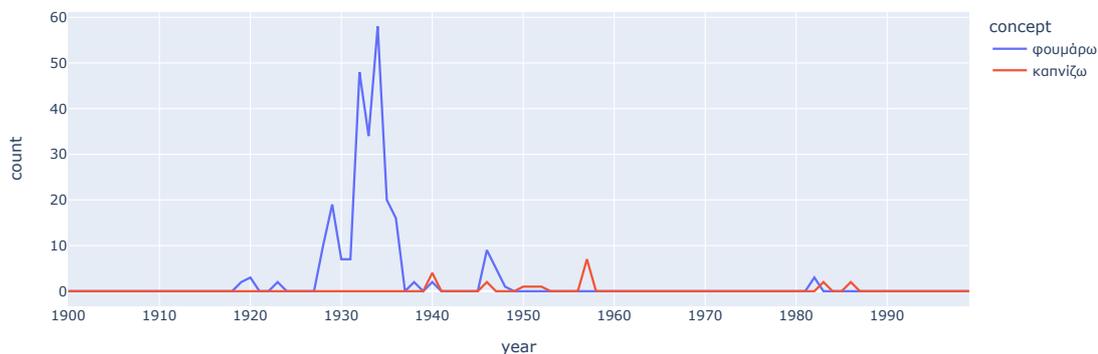


Figure 4.15: Custom Graph: Use of two Verbs Standing for *smoke* in the *Rebetiko Corpus*

Interestingly, a slight rise in the graph in 1940 for the verb *φουμάρω* ,smoke‘ is caused by two entries of this word in two recordings of the same song, which is «Σε γελάσανε» (“They Laughed at You”) by Kostas Skarvelis rerecorded in Athens (disc numbers: Parlophone B-74054, Parlophone B-74371), and in these songs the word *φουμάρω* is part of the phrase in the line «κατάλαβα τη γνώμη σου και τι καπνό φουμάρεις» (“I understood your opinion and what sort of person you are”), which is used metaphorically. So, the phrase *τι καπνό φουμάρεις* literally means ‘what kind of smoke you smoke’, but in its metaphorical use does not refer to smoke itself.

Such visualisations support a quick analysis of various concepts in a relatively large collection of rebetiko song lyrics. Another important advantage of such a custom tool is that one is very flexible regarding preprocessing or filtering the corpus data. Additionally, the definition of the concepts via lists of words is very easy compared to e.g. Voyant Tools, in which one has to repeat this process every time one opens the browser application via a rather awkward user interface.

4.2.3 *Topic Modeling*

Topic modeling is a popular statistical method for extracting salient topics from a collection of texts. It is in its classical form an unsupervised machine learning task, which means, that one does not need annotated data to train a model. In general, a computer when performing any kind of text analysis, often does not operate with the words themselves, but with numerical data. Statistical and machine learning methods in natural language processing mainly work with vectors and matrices. There are multiple ways of representing a text in a numerical form. The easiest encoding techniques are based on parameters such as an occurrence of a word in a document: for instance, one-hot encoding, in which one obtains a vector filled with only two possible values, which are either 1 (if a particular word occurs in a document) or 0 (if a word does not occur). The length of a vector in this case is defined by the number of unique words in the entire corpus. For instance, if we have as an entry a single sentence «Εγώ μάγκας φαινόμευνα να γίνω από μικράκι» (“I seemed to become mangas from an early age”), the sentence in Greek includes 7 tokens, and in a one-hot-encoded representation the first word *εγώ* will be represented as [1, 0, 0, 0, 0, 0, 0]. Although such an approach for text vectorisation is not sophisticated and does not enable to reflect in this way, for instance, word similarity as each word obtains its own one-hot vector (e.g. *soccer* and *football*), it is helpful in multiple tasks of NLP including text classification. Another way of transforming a text in a vector form is to train a model on a large dataset. One of the most popular text representation techniques are the so-called word embeddings⁸⁴, such that each word is mapped to a multi-dimensional vector.⁸⁵ For training a word embedding model, word vectors are usually randomly initialised⁸⁶, and then adapted step by step to get vector representations with small distances between them for words that occur in the same or similar context, and vector representations with large distances between them for words that do not appear in the same context. A common size for word embeddings is between 100 and

84 Another popular technique of text representation in NLP are sentence embeddings, which means that the entire sentence is mapped to a vector.

85 More information about word embeddings as well as one of the most popular ways to train word embeddings, namely Word2vec, can be found in (Karani 2018).

86 For the random initialisation of word vectors, often the normal distribution is used.

1000 dimensions. For calculating the similarity between two word embeddings, one usually uses the cosine similarity. Thus, the closer the words are to each other, the closer the value of the cosine similarity is to 1, i.e. the closer they are located in the representation space. The trained vectors do not carry the meaning in a way humans refer to a meaning of a word, but at the same time with numbers they are able to reflect not only semantic relations, but also morphological properties of lexical items (Kocmi/Bojar 2017, 56). Independently of how the numerical representations of a text, like word vectors, were obtained, they then may be used for building a topic model. Finally, interpretable visualisations of topics are created with the help of dimensionality reduction techniques such as t-SNE, which transform high-dimensional vectors into a low-dimensional space, i.e. a 2D space.

Topic Modeling with the LDA Model

In this section, an attempt will be made to automatically cluster the songs in the *Rebetiko Corpus* with regard to the main topics they are covering. This will be done with the help of one of the most common topic modeling algorithms, namely Latent Dirichlet allocation (LDA) (Blei/Ng/Jordan 2003). Latent Dirichlet allocation is a probabilistic model that builds clusters using two probability values: namely, $P(\text{word } w \mid \text{topic } t)$ and $P(\text{topic } t \mid \text{document } d)$. $P(\text{topic } t \mid \text{document } d)$ reflects the proportion of the words in a document d which are assigned to a topic t , and $P(\text{word } w \mid \text{topic } t)$ reflects the proportion of assignments of a word w (over all documents) to a topic t (Bansal 2020; Chawla 2017). At first, both sets of values are randomly initialised, and then probabilities are repeatedly updated over multiple iterations, until the algorithm converges and the topics are assigned. To make it more clear, the Latent Dirichlet allocation algorithm is based on the “assumption that every document in a literature corpus is a mixture of latent topics and each of these topics themselves is a probability distribution over words” (cited as in (Denzler 2021)). Due to the fact that most rebetiko songs refer simultaneously to more than one topic, such a technique is well suited for the task of rebetiko topic modeling. In general, this combination of topics in one single song text makes the task of topic modeling more complicated in comparison, for instance, to the topic modeling of a newspapers’ corpus, as newspapers usually include typical sections.

To perform topic modeling via LDA, I have written a Python script⁸⁷, which includes several important steps. For this experiment, like in Section 4.1, the *Rebetiko Corpus* with lyrics of 3,772 songs was used. The preprocessing stage includes removal of stop words and punctuation as well as tokenisation. As in Section 4.1.1, the spaCy NLP library was used. The lemmatisation was not performed, because as has been established in the previous experiments, the spaCy lemmatiser for Greek shows quite poor performance, so that lemmatisation would result in worse LDA topic models.

Next, I had to transform the normalised lyrics data into a numerical representation, which is necessary for using the LDA topic model. This representation was computed with the help of the so-called *Bag of Words* technique (BOW), which is a representation for documents (in this case lyrics of individual songs).⁸⁸ This approach is similar to the one-hot encoding mentioned in the introduction to this section, but the BOW representation reflects not only the fact of occurrence of a word, but also the frequency of its occurrence in a document. In essence, these vectors are built as follows. First, a *dictionary* with all the unique words in a

87 The Jupyter Notebook with this experiment can be found on GitHub at <https://github.com/kleines-gespenst/rebetiko-corpus-and-tools> (accessed February 26, 2022).

88 More information about *Bag of Words* can be found in (Gharibi 2019).

corpus is produced, so that each word also obtains an index. In case of the *Rebetiko Corpus*, after the preprocessing stage the *dictionary* contains 17,832 unique words. For this experiment, all words that occur less than three times in the entire corpus as well as those that occur in more than 60% of the song texts were removed, as these words are less meaningful for the topic extraction. Also other numbers were tested, and these were chosen as they lead to a better topic extraction. After this step, the vocabulary is reduced to 6,251 unique tokens. Then, for each document, i.e. for each rebetiko song, a vector of dimension 6,251 is produced, which means that the dimension of a vector is defined by the size of the reduced vocabulary. This vector is then used for counting the occurrences of each word in a specific song text. This means that a mapping of word indices to the frequency of their occurrence in a particular song is obtained. Due to the fact that the size of the corpus vocabulary is quite large and the length of a song is relative short, most entries in these vectors will be zeros, so it is a sparse vector.⁸⁹ Overall, after processing the data 3,772 songs were extracted. Thus, the *Rebetiko Corpus* after the reduction of the vocabulary can be represented in form of a matrix of dimension 3,772x6,251.

After I obtained the *bag of words* representation for the song lyrics, an LDA topic model was trained. For training the LDA model, the Gensim library for unsupervised topic modeling was used.⁹⁰ As the models before training are initialised randomly, each run results in different topic definitions. The number of topics one wants to obtain is set before training a model. In the case of the *Rebetiko Corpus* data, it was set to five topics, as it showed the best results. The trained topics were visualised using the pyLDAvis topic visualisation library⁹¹ with the help of the t-SNE method of projecting a high-dimensional space into a low-dimensional, to be more precise, in the scope of this experiment into a two-dimensional space. Below one can see this projection.

Figure 4.16 shows how close each topic is to the others in a 2D projection of the high-dimensional space. On the right, one can see the 30 most salient tokens in the reduced *Rebetiko Corpus*. As we can see in this visualisation (like with the help of word clouds in Section 4.1.1), some of the most frequent words in the corpus are *καρδιά* ,heart‘, *αμάν* ,oh, oh man‘, *αχ* ,ah, oh, ouch‘, *βρε* ,hey, dude, geez‘, *μάτια* ,eyes‘, *αγάπη* ,love‘ and *μάννα* ,mom‘.

Next, we will look at single topics. The parameter λ weights the probability of a term within a particular topic, so it allows to adjust the terms which are shown for each topic: the closer the parameter is to zero, the more topic specific terms will be presented, but at the same time the more rare these words are in general. In this particular visualisation this parameter is set to a rather low value, namely 0.14.

Figure 4.17 shows the most salient words for the first topic, defined by the LDA model. It includes tokens related to a beloved one such as *γλυκιά* ,(my) sweet girl, honey‘, *μικρή* ,little, (my) little girl‘, *ναζιάρα* ,mincing girl‘ and *μικράκι* ,little, (my) little girl‘. In this visualisation we also find words such as *καρδιά* ,heart‘, *φιλιά* < *φιλί* ,kiss (noun)‘, *φιλί* ,kiss (noun)‘, *χάδια* < *χάδι* ,caress‘, which all refer to the topic of love. Remarkably, one also finds among the most salient tokens for this topic terms such as *φως* ,light‘, *φλόγα* ,flame, passion‘, *έκαιψες* < *καίω* ,burn‘ and *άναιψες* < *ανάβω* ,turn on/ switch on (light)‘, which are in the context of rebetiko often used metaphorically referring to the emotions of a protagonist caused by

89 In Python these sparse vectors are stored as dictionary objects, which need much less memory than dense vectors.

90 More about Gensim can be found at <https://radimrehurek.com/gensim/> (accessed February 20, 2022).

91 More about pyLDAvis Python library can be found at <https://pyldavis.readthedocs.io/en/latest/readme.html> (accessed February 20, 2022).

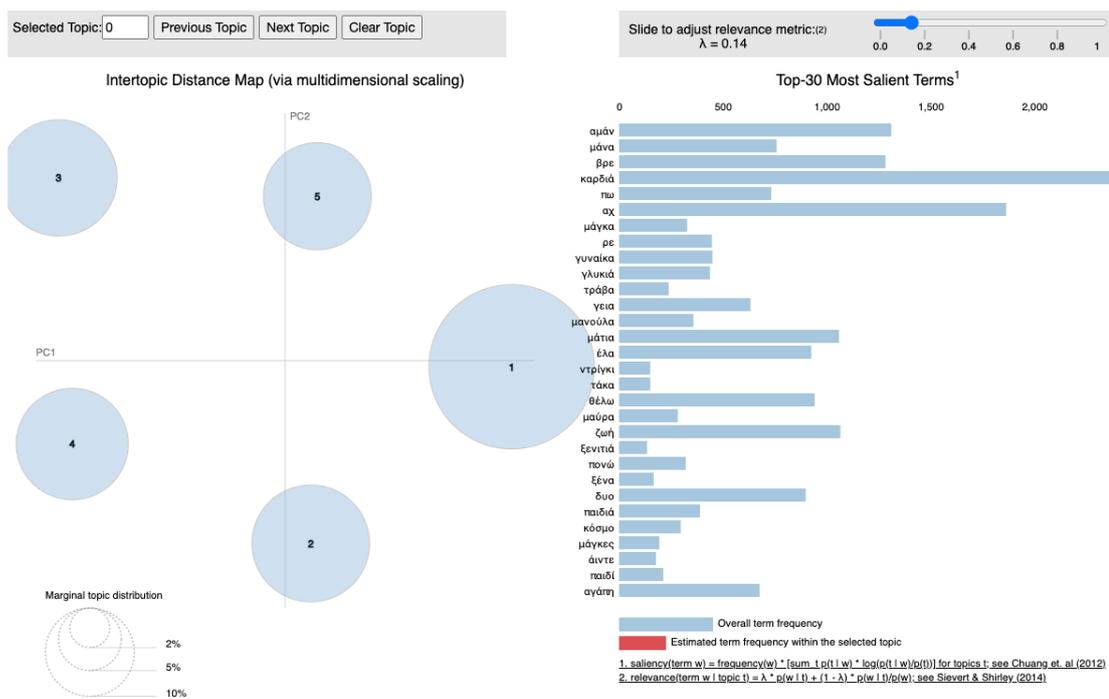


Figure 4.16: LDA Topic Modeling: General View of Topics with overall Term Frequency

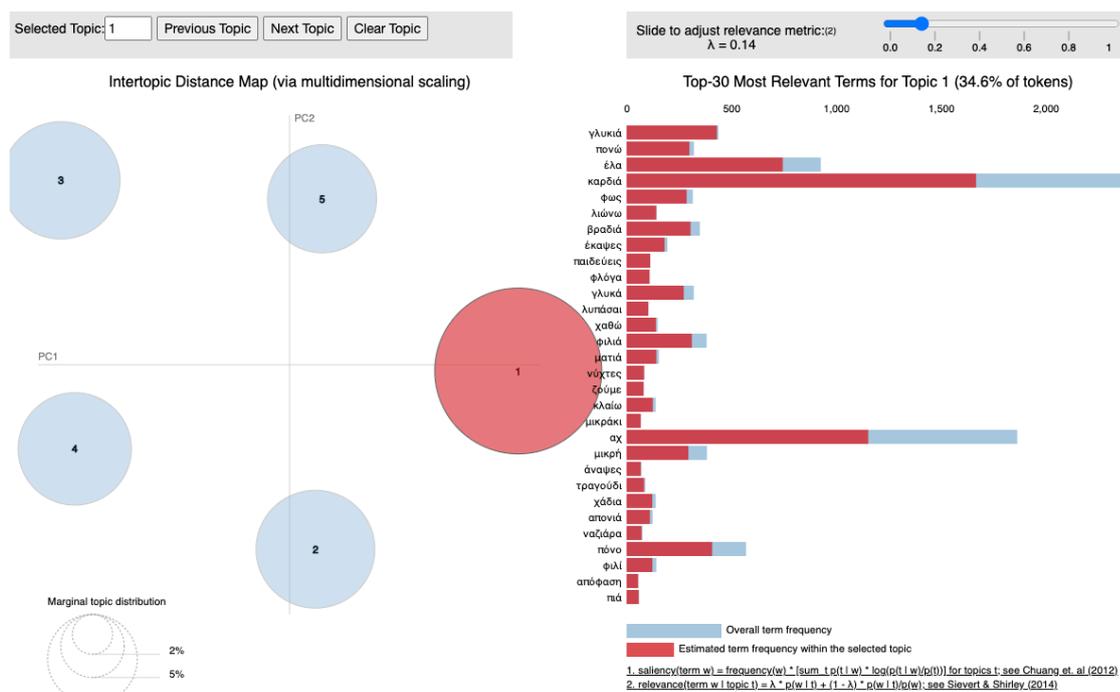
his love. Among the words presented in this visualisation there are also those referring to the sufferings of a protagonist of rebetiko, which are often caused by an unrequited love towards a heartless beloved one: *απονιά* ,heartlessness‘, *λυπάσαι* < *λυπάμαι* ,regret‘, *κλαίω* ,cry‘ and *πόνο* < *πόνος* ,pain‘. So, one could make an assumption that these tokens may define a topic referring to love as well as sufferings caused by love to a heartless woman. The size of the circle defining this topic also shows that it prevails over the four other ones.

Figure 4.18 shows the most salient words for the second topic. The first connecting characteristic of these words, which may strike our attention, is the number of tokens, which refer to the topic of hashish consumption. These include a number of tokens that refer directly to hashish, namely *αργιλέ* < *αργιλές* ,hookah, waterpipe, arghileh‘, *νταμίρα* ,hashish (in argot)‘ and *χασίσι* ,hashish‘. Moreover, the topic also includes the term *τεκέ* < *τεκές* ,tekke, hashish den‘, one of the rebetiko venues in which its frequenters used to smoke hashish and make music. The appearance of a word such as *πλένεις* < *πλένω* ,wash‘ is caused by the fact that the verb *πλένω* ,wash‘ also occurs in the context of hashish smoking with words such as hashish and waterpipe. The lines below are from the song «*Η χασικλού*» (“The Hash Smoker (female)”) composed by Panagiotis Toundas.⁹²

θα σου γεμίζω τον λουλά, θα πλένω το κα-
λάμι,
μαζί να τη φουμάrouμε, βρε μάγκα μου κι α-
λάνι.

I will fill your hookah and wash the reed⁹³,
so that we smoke it together, my vagrant
and bum.

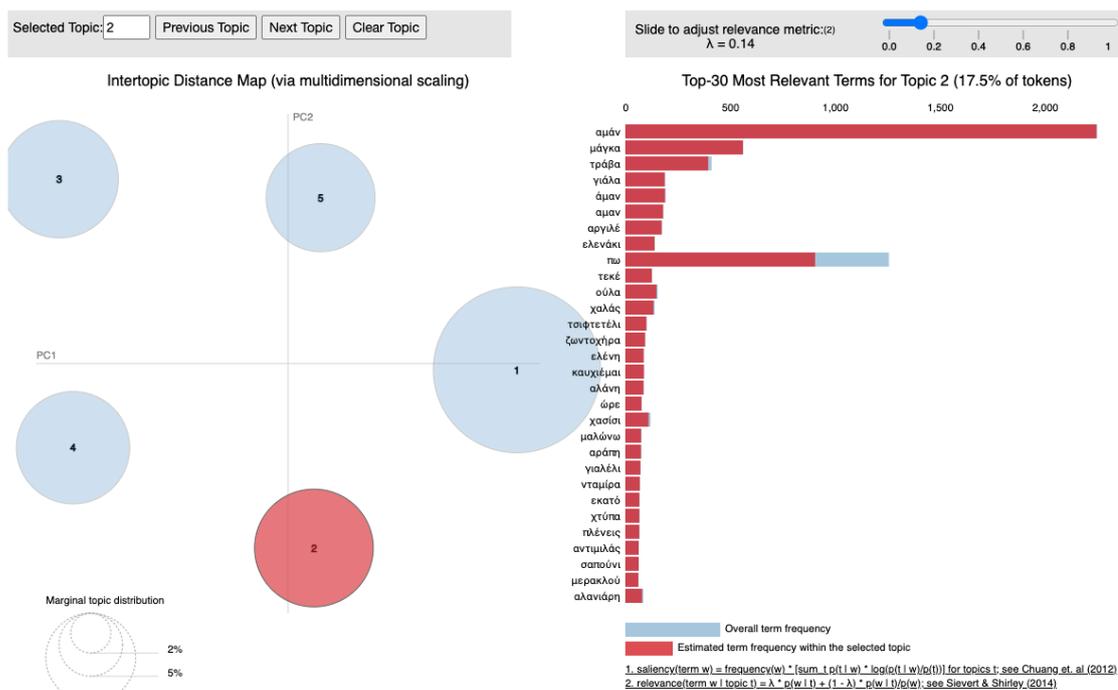
⁹² The song «*Η χασικλού*» (“The Hash Smoker (female)”) composed by Panagiotis Toundas was recorded several times including the recording made in 1930 (disc number: Pathé X-80188, sung by Andonis Dalgas).

Figure 4.17: LDA Topic Modeling: Topic 1 *Love*

Probably because the verb *πλένω*, 'wash' occurs in the song lyrics referring to hashish, the LDA algorithm also considers that the word *σαπούνι*, 'soap' refers to this topic, so that it also appears in this visualisation. In Figure 4.18 we also find a range of interjections such as *αμάν*, 'oh, oh man', *γιάλα*, 'giala (exclamation of misery and pain)' and *πω*, 'oh, oh my, wow'. This is caused by the fact that most rebetiko songs referring to hashish were recorded before the mid 1930s, and the number of interjections in the songs of that period is also much higher than in later periods, as it was shown earlier, for instance, in the example of the Dragatsis subcorpus (see Section 4.1.1, p. 80). Tokens such as *μάγκα* < *μάγκας*, 'vagabond, wide boy' *αλανιάρη* < *αλανιάρης*, 'vagrant, bum' also often occur in the context of songs referring to hashish consumption.

Figure 4.19 shows the most salient words for the topic 4. Interestingly, among them are several terms, which refer to family members including *μάννα/μητέρα*, 'mom/mother', *πατέρα* < *πατέρας*, 'father' and *γιο* < *γιος*, 'son'. Among the words referring to family, it also includes words such as *πατρίδα*, 'homeland', *ξενιτιά*, 'foreign lands, abroad, overseas', *ξένα* < *ξένος*, 'foreign', *φύγω* < *φεύγω*, 'go away' and *βάρκα*, 'boat'. So, this topic refers not only to family, but also to a dichotomy *foreign lands* and *homeland*. As it was discussed in Chapter 2, many representatives of rebetiko were forced to leave their homes either due to the mandatory exchange of population between Greece and Turkey in 1923 or due to a difficult economic situation. Most migrants, who went to the United States, left their entire family either in Greece or the Ottoman Empire/Turkey. The lines below from the song «*Μη με στέλνεις, μάννα,*

93 A part of a hookah, which is known as a hookah hose, was back then often made of reed.

Figure 4.18: LDA Topic Modeling: Topic 2 *Hashish Consumption*

στην Αμερική» (“Mom, Don’t Send Me to America”)⁹⁴ by Dimitris Semsis narrate about the unwillingness of a female protagonist to go to America.

Μη με στέλνεις μάννα στη Αμερική
θε να μαραζώσω, να πεθάνω εκεί.

Mom, don’t send me to America,
you want me to wither away, to die there

Δολάρια δεν θέλω πώς να σου το πω
καλιά, ψωμί, κρεμμύδι, κι αυτόνε π’ αγαπώ.

I don’t want dollars, how can I explain it
to you
Only bread, onions, and the one I love.

- Όπα, κακούργα ξενιτριά!

- Oh evil foreign land!

This song continues by the female protagonist telling her mother that she loves a man from a village. It is an interesting case of a song, which depicts a woman that is forced to go overseas, as 90% of Greek migrants to the USA between 1890 and early 1920s were male (see Section 2.3, p. 28).

The last two topics (i.e. topics 3 and 5), which were identified by the LDA model, are much harder to interpret. Their visualisations can be found in Section 6.4 in the appendix. It is possible that these two remaining topics try to summarise the remaining documents, but they do not present coherent topics. The outputs of topic models are in general often difficult to interpret. In case of topic extraction for the *Rebetiko Corpus* it may be caused by several factors. One of them lies in the missing lemmatisation of the corpus, which was not

⁹⁴ The song «Μη με στέλνεις, μάννα, στην Αμερική» (“Mom, Don’t Send Me to America”) by Dimitris Semsis was recorded in 1934 and sung by Rita Abatzi (disc number: HMV AO 2196).

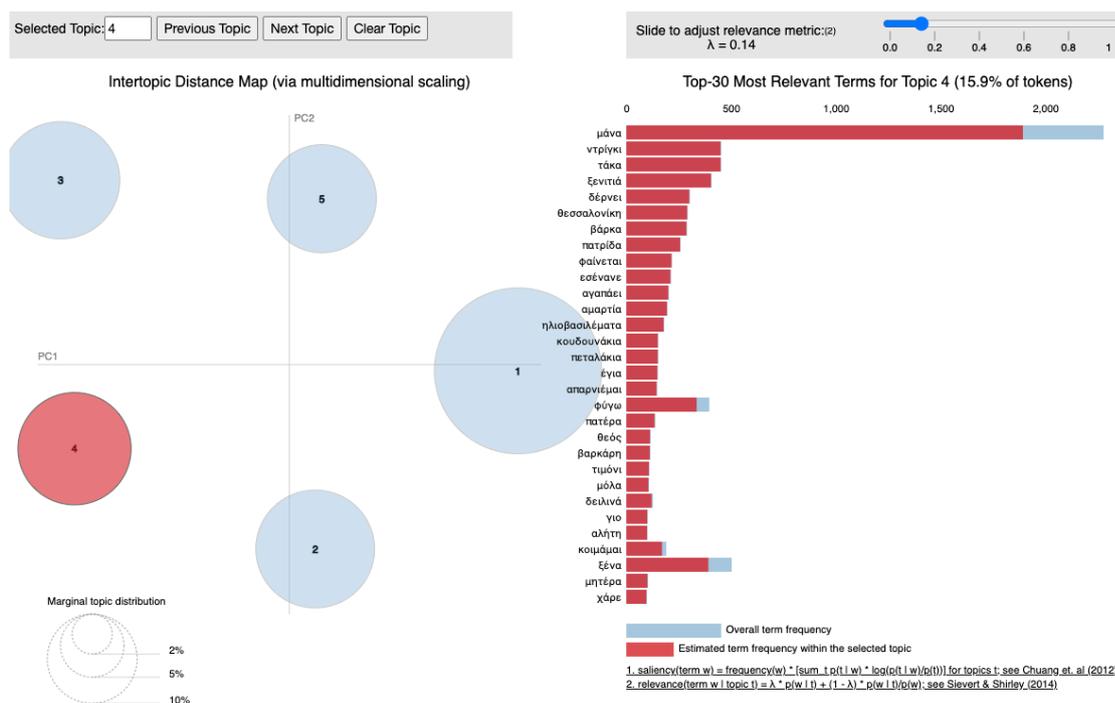


Figure 4.19: LDA Topic Modeling: Topic 4 *Family, Homeland and Foreign Land*

performed due to the unreliability of spaCy lemmatiser. It would also be of advantage to test other data representations than *Bag of Words*. In this context it may be of interest to test the so-called LDA2vec model, which integrates word2vec embeddings into its architecture (Hulstaert 2017). When working on topic modeling of the *Rebetiko Corpus* I also tested an Embedded Topic Model, or ETM⁹⁵, in combination with fastText⁹⁶ embeddings for Greek, but so far the best results for the *Rebetiko Corpus* were achieved with the LDA model which uses *Bag of Words*.

As the presented visualisations have shown, even with the current performance the LDA model allows to look at the rebetiko lyrics from another perspective, which may be a starting point for further and deeper analysis of the song lyrics.

95 More information about ETM can be found at <https://github.com/lffloyd/embedded-topic-model> (accessed February 20, 2022).

96 More information about fastText can be found at <https://fasttext.cc> (accessed February 20, 2022).

CONCLUSION AND FUTURE WORK

Rebetiko is a fascinating and multi-faceted music genre, with a rich history. Its lyrics reflect many aspects of the history of Greece in the 20th century and are thus a valuable source and a worthwhile object of research. This thesis has investigated the viability of a corpus-based analysis of the lyrics of rebetiko.

To this end, the background of rebetiko was discussed extensively, including the etymology and the history of the name of the genre, its main characters such as *mangas* and *rebetis*, the issue of marginality of rebetiko, the question of authorship, the importance of the oral tradition, and the relationship to other Greek music genres such as *laika* or *elafra*, as well as neighbouring music traditions. The history of rebetiko was presented in detail, focusing on the impact of events such as World War II, the Greek Civil War, the dictatorship of General Ioannis Metaxas, and the military Junta in Greece on the evolution of rebetiko, its lyrics, instrumentation, performances, and recordings. Typical topics of rebetiko, which include love, sufferings of a protagonist, hashish consumption and the dichotomy of homeland versus foreign land, were presented and discussed in their social and historical contexts throughout this treatise.

To facilitate a quantitative analysis of the rebetiko lyrics, the *Rebetiko Corpus* was created. This corpus is based on data available at rebetiko.sealabs.net and contains 5,165 songs, of which 3,772 songs are provided with song lyrics. Furthermore, the corpus provides important metadata like song title, release year, instrumentation, names of the involved artists (composer, lyricist, instrument player(s) and singer(s)), disc number, dance (rhythm) and location of recording. Both the metadata and lyrics are in Modern Greek. The corpus was published on Zenodo and is freely available for further research under the license CC BY-NC-SA 4.0.

Based on the *Rebetiko Corpus*, various kinds of quantitative analyses were performed, with the goal of verifying some common assumptions about rebetiko, as well as to uncover novel insights into rebetiko. To this end, the well-known text analysis tool Voyant Tools, as well as custom scripts implemented in the programming language Python were used. For instance, word clouds created with Voyant Tools as well as the topic model based on latent Dirichlet allocation have shown that songs referring to love play an important role in rebetiko and that the topic of love prevails over all other topics, at least based on the data of the *Rebetiko Corpus*. This also means that the number of songs referring to hashish consumption or prison life, which are commonly assumed to be very important topics in rebetiko, is much lower than those that refer to love, at least in the created corpus. The word cloud visualisations in Voyant Tools, the topic model and further deeper analysis of texts based on these data have also demonstrated that many rebetiko songs include the figures of a mother and of a widow. Much more less common are the figures of a father, a daughter and a son. Due to the fact that the corpus data are provided with all necessary metadata, one can study trends over time, as it was shown, for instance, on the example of the terms with the meaning of *hashish*. In this case, the analysis clearly shows the impact of the censorship under Metaxas on the lyrics and the topics of rebetiko.

This thesis can only scratch on the surface, but from the presented experiments it is clear, that quantitative data analysis is a viable tool for researching rebetiko lyrics, and many more possible kinds of analyses come to mind. For instance, one could try to extract named entities such as toponyms from the rebetiko songs and visualise geographic relationships within specific songs or within the corpus. One could also follow more experimental paths and try to train machine learning models that generate new rebetiko songs or try to identify the song writer by analysing the lyrics. As the *Rebetiko Corpus* has been made publicly available, these and many more avenues for future research are now open for any interested researcher.

The thesis has also uncovered some limitations of natural language processing tools available for Modern Greek. In general, natural language processing for low-resource languages like Modern Greek is a difficult task and an important area of research. In this thesis the deficiencies of certain parts of the processing pipeline, especially the lemmatiser and the part-of-speech tagger, impeded a number of experiments. Improving these algorithms is vital for better text analysis results. A further difficulty lies in the specific properties of the lyrics themselves, which are typically not reflected in natural language processing algorithms. For instance, this type of texts exhibits specific syntax, grammar, prosody, phonology as well as a wide use of metaphors. The punctuation, which influences drastically the way how text analysis tools interpret the text, is almost completely missing in rebetiko song lyrics, as it is typical for song lyrics in general. All of the above makes the use of the available natural language processing tools even more difficult.

The *Rebetiko Corpus* lends itself as a relevant data source for research in various different fields, such as musicology, linguistics, history, anthropology and sociology. For instance, the rebetiko song lyrics contain colloquial language and use a specific vocabulary, of which many elements have entered Modern Greek and became popular and commonly used. The study of rebetiko lyrics is thus of vital importance for understanding some elements of Modern Greek as it is spoken today. Rebetiko song lyrics include many topics that are underrepresented in official historiography. Even if the songs do not directly refer to particular events, they are important documents of oral history, and songs referring to drug consumption, illnesses such as tuberculosis, gambling, life as a refugee or a migrant, give insights into the struggles of every day life and the impact of historical events, like World War I, World War II or the exchange of population between Turkey and Greece 1923, on the general population.

On a personal note, this thesis has motivated me to listen to thousands of rebetika of different chronological periods and thus introduced a new music genre to my everyday life. It also allowed me to take a look at the history of Greece through the lens of rebetiko, which emphasises aspects of the history which might be underrepresented in the general historiography of Greece. This thesis has enabled me to combine two fields of which I am passionate about, namely linguistics and data science. I am looking forward to applying the new tools and techniques, which I have learned while working on this thesis, in my future research.

APPENDIX

6.1 COMPOSERS INCLUDED IN THE *REBETIKO CORPUS*

Table 6.1 lists all composers in the corpus with at least 20 songs.

Composer	Number of Songs
Vasilis Tsitsanis – Βασίλης Τσιτσάνης	582
Unknown – Άγνωστος	537
Panagiotis Toundas – Παναγιώτης Τούντας	429
Giannis Papagioannou – Γιάννης Παπαϊωάννου	391
Manolis Chiotis – Μανώλης Χιώτης	331
Spyros Peristeris – Σπύρος Περιστέρης	255
Kostas Skarvelis – Κώστας Σκαρβέλης	207
Markos Vamvakaris – Μάρκος Βαμβακάρης	158
Giannis Dragatsis (Ogdondakis) – Γιάννης Δραγάτσης (Ογδοντάκης)	125
Apostolos Kaldaros – Απόστολος Καλδάρας	123
Stavros Pandelidis – Σταύρος Παντελίδης	92
Kostas Roukounas – Κώστας Ρούκουνας	92
Michalis Genitsaris – Μιχάλης Γενίτσαρης	91
Apostolos Chatzichristos – Απόστολος Χατζηχρήστος	90
Andonis Diamantidis (Dalgas) – Αντώνης Διαμαντίδης (Νταλγκάς)	84
Stratos Pagioumtzis – Στράτος Παγιουμτζής	78
Giorgos Mitsakis – Γιώργος Μητσάκης	77
Dimitris Semsis (Salonikios) – Δημήτρης Σέμισης (Σαλονικιός)	63
Kostas Karipis – Κώστας Καρίπης	62
Manolis Chrysafakis – Μανώλης Χρυσάφης	61
Stavros Tzouanacos – Σταύρος Τζουανάκος	59
Dimitris Gongos (Bagiaderas) – Δημήτρης Γκόγκος (Μπαγιαντέρας)	57
Ioanna Georgakopoulou – Ιωάννα Γεωργακοπούλου	46
Stelios Chrysinis – Στέλιος Χρυσίνης	45
Agapios Tomboulis – Αγάπιος Τομπούλης	42
Giorgos Katsaros (Theologitis) – Γιώργος Κατσαρός (Θεολογίτης)	38
Babis Bakalis – Μπάμπης Μπακάλης	34
Georgia Mittaki – Γεωργία Μηττάκη	32
Iakovos Montanaris – Ιάκωβος Μοντανάρης	31
Vangelis Papazoglou – Βαγγέλης Παπάζογλου	31
Stellakis Perpiniadis – Στελλάκης Περπινιάδης	30
Grigoris Asikis – Γρηγόρης Ασίκης	27
Marinos Gavriil (Marinakis) – Μαρίνος Γαβριήλ (Μαρινάκης)	21
Panagiotis Petsas – Παναγιώτης Πετσάς	20

Table 6.1: List of Composers with more than 20 Pieces in the *Rebetiko Corpus*

6.2 USE OF PRONOUNS IN THE *REBETIKO CORPUS*

The tables below demonstrate the use of personal pronouns in the *Rebetiko Corpus*. They present both strong and weak forms of pronouns.¹ Due to the fact that some of these forms are polysemous or homonymous, i.e. may have different grammatical roles (for instant, direct or indirect object) or even may be used as different parts of speech, some of the numbers in the tables are estimations based on query results, without further time-consuming manual analysis. Determining the exact number of occurrences in the corpus is difficult, as some forms may be used as different parts of speech (cf. a preposition *με* 'with' and a weak form of first-person singular personal pronoun in accusative *με*), or play a role of different types of pronouns (cf. masculine singular demonstrative pronoun *αυτός* 'this' in nominative and a strong form of a masculine singular personal pronoun *αυτός* in nominative), or share the same form for different grammatical cases (cf. strong form of a first-person singular personal pronoun *εμένα* in genitive and accusative). Thus, the sign < defines an upper bound of occurrences for these forms. Respectively, the sign > hints to a minimum number of occurrences. For instance, it is difficult to determine an exact number of occurrences in the corpus, if for a form additional contractions may be used. These contractions in turn may be interpreted differently and may be a different part of speech or different grammatical case depending on the context: for instance, *μ'* may be a contraction of a preposition *με* 'with', first-person singular personal pronoun in accusative *με* or first-person singular personal pronoun in genitive *μου*. Due to the fact that going through all the query results would be very time consuming, after the sign > only minimal number of occurrences was given.

To help the reader in the interpretation of these tables, a further example will be given. The token *του* occurs in the *Rebetiko Corpus* 1,448 times, but it is polysemous and may either serve as a weak masculine/ neuter third-person pronoun in genitive or as a masculine/ neuter definite article in genitive, and the use in the last role is much more frequent than as a personal pronoun. For this reason the sign < was used, as its use in the role of a particular form of personal pronoun is definitely lower than the query result. Even if we take the highest possible number for all the pronouns except of first- and second-person singular, the first- and second-person personal pronouns too will be the most frequent.

Additionally, it is important to mention, that this analysis included queries for all possible variants of a particular form of a pronoun: for instance, the query for *εμένα* also included a query for *εμέ*, *εμένανε*, *μένανε*, *μένα* as well as these variants with missing word stress, so that the number in the table presents the sum of the query results of all possible variants of particular forms of a pronoun. The sign ? was used highlight cases for which the query results almost exclusively referred to a definite article and not to weak forms of personal pronouns (and thus the use of the sign < would be misleading).

1 More information about personal pronouns can be found in (Triantafyllidis 2018, 286-292).

	Singular	Plural
Nominative	εγώ (1,590)	εμείς (111)
Genitive	εμένα (<524), μου (>15,982)	εμάς (0), μας (<1,632)
Accusative	εμένα (<524), με (<11,847)	εμάς (4), μας (<1,632)

Table 6.2: Use of First-person Pronouns in the *Rebetiko Corpus*

	Singular	Plural
Nominative	εσύ (1,306)	εσείς (49)
Genitive	εσένα (< 1,462), σου (> 8,676)	εσάς (0), σας (< 238)
Accusative	εσένα (< 1,462), σε (< 6,164)	εσάς (2), σας (< 238)

Table 6.3: Use of Second-person Pronouns in the *Rebetiko Corpus*

	Singular		
	Masculine	Feminine	Neuter
Nominative	αυτός (< 110) τος (1)	αυτή (< 418) τη (?)	αυτό (< 620) το (?)
Genitive	αυτού (< 7) του (< 1448)	αυτής (0) τής (<1831)	αυτού (< 7) του (<1448)
Accusative	αυτόν (130) τον (< 2,906)	αυτήν (< 99) τήν (< 4,340)	αυτό (< 620) το (?)

Table 6.4: Use of Third-person Singular Pronouns in the *Rebetiko Corpus*

	Plural		
	Masculine	Feminine	Neuter
Nominative	αυτοί (< 62) τοι (0)	αυτές (< 52) τες (2)	αυτά (< 337) τα (?)
Genitive	αυτών (0) τους (< 846)	αυτών (0) τους (< 846)	αυτών (0) τους (< 846)
Accusative	αυτούς (< 30) τους (< 846)	αυτές (< 52) τις/τες (< 830)	αυτά (< 337) τα (?)

Table 6.5: Use of Third-person Plural Pronouns in the *Rebetiko Corpus*

6.3 EXAMPLE OF LEMMATISATION PERFORMED BY UDPIPE WITH PUNCTUATION

Id	Form	Lemma	UPosTag	XPosTag	Feats	Head	DepRel	Deps	Misc
# generator = UDPIPE 2, https://lindat.mff.cuni.cz/services/udpipe									
# udpipe_model = greek-gdt-ud-2.6-200830									
# udpipe_model_licence = CC BY-NC-SA									
# newdoc									
# newpar									
# sent_id = 1									
# text = Αντί σκολιό μου πάγαινα μες του Καραϊσκάκη, έπινα διάφορα πιστά να μάθω μπουζουκάκι.									
1	Αντί	αντί	ADP	ADP	_	2	case	_	TokenRange=0:4
2	σκολιό	σκολιό	NOUN	NOUN	Case=Acc Gender=Neut Number=Sing	9	obl	_	TokenRange=5:11
3	μου	μου	PRON	PRON	Case=Gen Gender=Masc Number=Sing Person=1 Poss=Yes PronType=Prs	2	nmod	_	TokenRange=12:15
4	πάγαινα	πάγαινα	VERB	VERB	_	2	acl:relcl	_	TokenRange=16:23
5	μες	μες	ADP	ADP	_	7	case	_	TokenRange=24:27
6	του	ο	DET	DET	Case=Gen Definite=Def Gender=Masc Number=Sing PronType=Art	7	det	_	TokenRange=28:31
7	Καραϊσκάκη	Καραϊσκάκη	PROPN	PROPN	Case=Gen Gender=Masc Number=Sing	4	obl	_	SpaceAfter=No TokenRange=32:42
8	,	,	PUNCT	PUNCT	_	9	punct	_	SpacesAfter='\n' TokenRange=42:43
9	έπινα	έπινος	VERB	VERB	Aspect=Perf Mood=Ind Number=Sing Person=1 Tense=Past VerbForm=Fin Voice=Act	0	root	_	TokenRange=45:50
10	διάφορα	διάφορος	ADJ	ADJ	Case=Acc Gender=Neut Number=Plur	11	amod	_	TokenRange=51:58
11	πιστά	πιστά	NOUN	NOUN	Case=Acc Gender=Neut Number=Plur	9	obj	_	TokenRange=59:64
12	να	να	AUX	AUX	_	13	aux	_	TokenRange=65:67
13	μάθω	μάθω	VERB	VERB	Aspect=Perf Mood=Ind Number=Sing Person=1 VerbForm=Fin Voice=Act	11	acl	_	TokenRange=68:72
14	μπουζουκάκι	μπουζουκάκι	NOUN	NOUN	Case=Acc Gender=Neut Number=Sing	13	obj	_	SpaceAfter=No TokenRange=73:84
15	.	.	PUNCT	PUNCT	_	9	punct	_	SpaceAfter=No TokenRange=84:85

Figure 6.1: Example of Lemmatisation Performed by UDPipe with Punctuation. Source: <http://lindat.mff.cuni.cz/services/udpipe/> (accessed February 18, 2022)

6.4 LDA TOPIC MODELING: TOPICS 3 AND 5

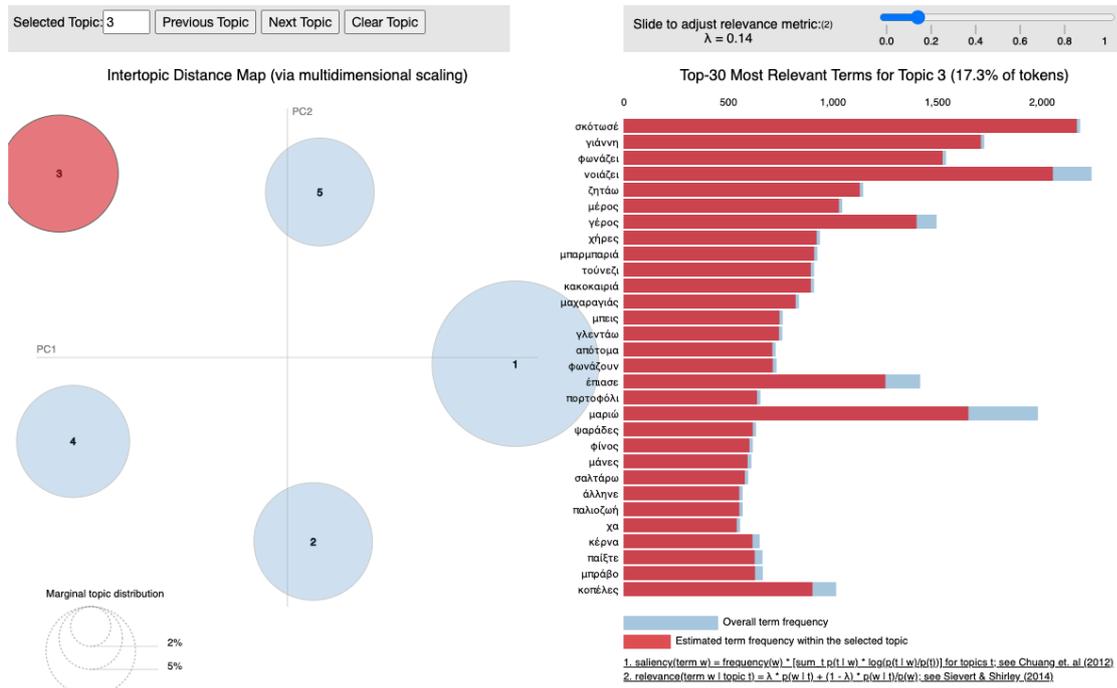


Figure 6.2: LDA Topic Modeling: Topic 3

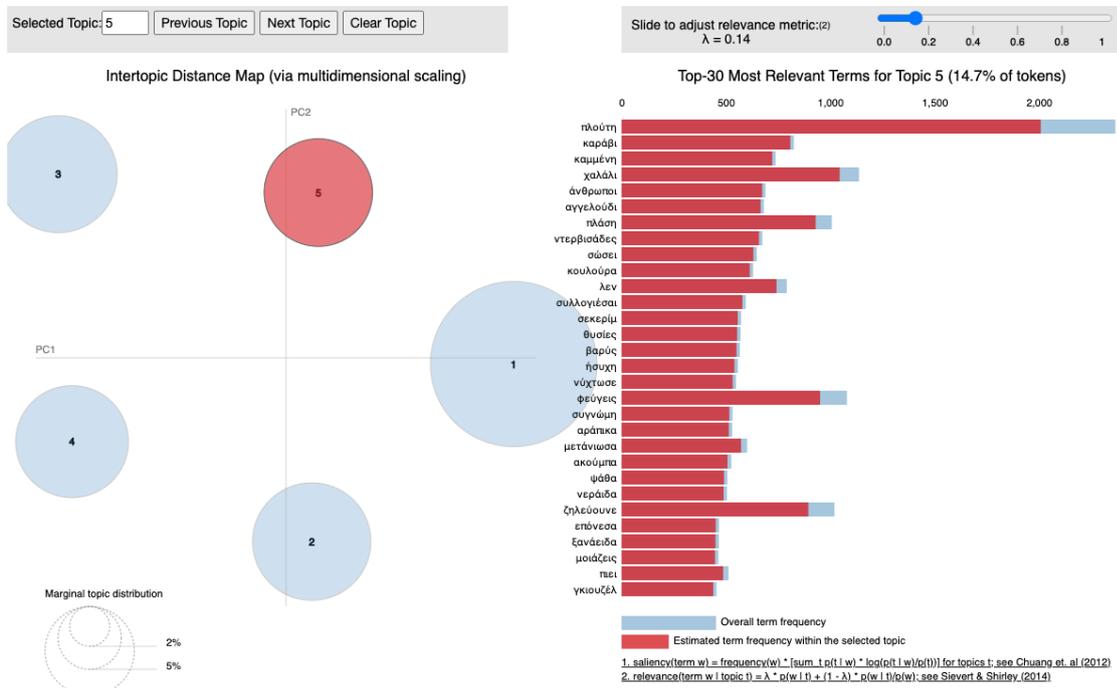


Figure 6.3: LDA Topic Modeling: Topic 5

ABSTRACTS

7.1 ABSTRACT

Rebetiko is a Greek popular music genre that is commonly seen as Greek blues, due to the similarities in the genres' formations. It is often associated with the urban underworld and marginal groups and had and still has an immense influence on the Greek culture and language.

This thesis combines the music genre rebetiko with the idea of computer-assisted analysis, i.e. it tackles the question of the creation of a rebetiko corpus, with a focus on the song lyrics, and investigates the viability of corpus-based text analysis via programmes like Voyant Tools and via custom programmes as research tools. The corpus created in the scope of this thesis contains 5,165 songs, of which 3,772 songs are provided with song lyrics, and is made publicly available for non-commercial purposes as a source for research in diverse areas such as history, sociology, linguistics, and musicology.

Based on the created corpus, quantitative analyses were conducted, with the goal of verifying common assumptions about rebetiko and uncovering novel insights into this music genre. These quantitative methods were successfully used for re-discovering and thus confirming suppositions like the importance of the topic of love and the prevalence of a first-person narrative. Furthermore, these methods were successful in highlighting issues that formerly might have been neglected, like the multi-faceted roles of women in rebetiko, as well as particular linguistic properties of rebetiko. These discoveries and the rebetiko corpus in general provide multiple directions for further research.

7.2 ZUSAMMENFASSUNG

Rebetiko ist eine griechische Musikgattung, die aufgrund der Ähnlichkeiten in den Entstehungsgeschichten der Genres oft als griechischer Blues bezeichnet wird. Das Genre wird oft mit der städtischen Unterwelt und mit Randgruppen der Bevölkerung in Verbindung gebracht und übt auch heute noch einen immensen Einfluss auf die griechische Kultur und Sprache aus.

Diese Arbeit kombiniert das Musikgenre Rebetiko mit der Idee der computergestützten Datenanalyse. Genauer gesagt befasst sie sich mit der Frage der Erstellung eines Rebetiko-Korpus mit Fokus auf die Liedtexte und untersucht die Eignung von korpusbasierter Textanalyse mit Programmen wie Voyant Tools und selbst programmierten Skripts als Forschungswerkzeug. Das im Rahmen dieser Masterarbeit erstellte Korpus umfasst 5.165 Lieder, wovon 3.772 mit Liedtexten versehen sind, und es wurde für nicht-kommerzielle Zwecke als Quelle für die Forschung in vielfältigen Bereichen wie Geschichte, Soziologie, Sprachwissenschaft und Musikwissenschaft öffentlich zugänglich gemacht.

Basierend auf dem erstellten Korpus wurden quantitative Analysen durchgeführt, mit dem Ziel, bekannte Annahmen über das Rebetiko zu überprüfen und neue Einblicke in dieses Musikgenre zu gewinnen. Mithilfe dieser Analysen wurden beispielsweise Annahmen wie die Wichtigkeit des Themas Liebe und die überwiegende Verwendung der Ich-Erzählung wiederentdeckt und damit quantitativ bestätigt. Darüber hinaus wurden diese Methoden auch erfolgreich eingesetzt, um Themen zu entdecken, denen früher möglicherweise zu wenig Beachtung geschenkt wurde, wie die facettenreiche Rolle der Frau im Rebetiko, sowie besondere sprachliche Eigenschaften dieses Genres. Diese Themen und das Rebetiko-Korpus im Allgemeinen bieten mehrere spannende Ansatzpunkte für die weitere Forschung.

BIBLIOGRAPHY

- Abbott, Lynn / Seroff, Doug. 1996. ““They cert’ly Sound Good to Me”: Sheet Music, Southern Vaudeville, and the Commercial Ascendancy of the Blues.” *American Music*, 402–454.
- Alexatos, Gerassimos. 2018. *Die Griechen von Görlitz 1916–1919*. Berlin: Frank & Timme GmbH. Verlag für wissenschaftliche Literatur.
- Anagnostou, Panagiota. 2018. “Did You Say Rebetiko? Musical Categories, Their Transformation, and Their Meanings.” *Journal of Social History* 52 (2): 283–303.
- Anoyanakis, Fivos. 1961. “Για το Ρεμπέτικο Τραγούδι [About the Rebetiko Song].” *Επιθεώρηση Τέχνης (Revue d’Art)* (Athens) 79:11–20.
- Arzt, Varvara. 2022. “Rebetiko Corpus (1.0) [Dataset].” *Zenodo*, accessed February 22, 2022. <https://doi.org/10.5281/zenodo.6211555>.
- Ayverdi, İlhan. 2005. *Misalli Büyük Türkçe Sözlük [Misalli Grand Turkish Dictionary]*. Kubbealtı. <http://lugatim.com>.
- Baltazani, Mary. 2006. “Focusing, Prosodic Phrasing, and Hiatus Resolution in Greek.” *Laboratory phonology* 8:473–494.
- Bansal, Harsh. 2020. “Latent Dirichlet Allocation.” *Medium*, accessed February 18, 2022. <https://medium.com/analytics-vidhya/latent-dirichlet-allocation-1ec8729589d4>.
- Blei, David M / Ng, Andrew Y / Jordan, Michael I. 2003. “Latent Dirichlet Allocation.” *Journal of machine Learning research* 3 (Jan): 993–1022.
- Bucuvalas, Tina, ed. 2018. *Greek Music in America*. Jackson: University Press of Mississippi, Kindle edition.
- Chafe, Wallace / Tannen, Deborah. 1987. “The Relation between Written and Spoken Language.” *Annual review of anthropology* 16 (1): 383–407.
- Chamoudopoulos, Minas. 1871. *Οι Μυστηριώδεις Νυκτοκλέπται [Mysterious Night Burglars]*. Smyrna: Typois Smyrnis.
- Chatzigrigoriou, Pavlos et al. 2019. “Ο Νεότερος Ελληνικός Αστικός Πολιτισμός μέσα από το Ρεμπέτικο. Η Περίπτωση του Εικονικού Μουσείου «Αρχείου Κουνάδη»: από τη Σύλληψη στην Υλοποίηση [Modern Greek Urban Culture through Rebetiko. The Case of the Virtual Museum “Kounadis Archive”: from the Concept to Realisation].” *Διεθνές Συνέδριο Φερεκίδης: Ο Δυτικός πολιτισμός και οι Ελληνικές αναφορές του [International Conference “Ferekidis”: Western Culture and its Greek references.]*
- Chawla, Ravish. 2017. “Topic Modeling with LDA and NMF on the ABC News Headlines dataset.” *Medium*, accessed February 18, 2022. <https://medium.com/ml2vec/topic-modeling-is-an-unsupervised-learning-approach-to-clustering-documents-to-discover-topics-fdfbf30e27df>.

- Corona-Korpus, DWDS. 2020. “Digitales Wörterbuch der deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart.” Last accessed January 26, 2022. <https://www.dwds.de/d/korpora/corona>.
- De Kok, Daniël. 2021. “Neural Edit-tree Lemmatization for spaCy.” *Explosion*, accessed February 7, 2022. <https://explosion.ai/blog/edit-tree-lemmatizer>.
- Denzler, Tim. 2021. “What’s in a Song? LDA Topic Modeling of over 120,000 Lyrics.” *Medium*, accessed February 18, 2022. <https://tim-denzler.medium.com/whats-in-a-song-using-lda-to-find-topics-in-over-120-000-songs-53785767b692>.
- Dragoumis, Markos. 2003. *Η Παραδοσιακή μας Μουσική [Our Traditional Music]*. Vol. 1. Athens: Κέντρο Μικρασιατικών Σπουδών.
- Droulia, Loukia. 1999. “Οι Ενδυματολογικές Μεταλλάγες στα Χρόνια της Εθνικής Διαμόρφωσης του Νέου Ελληνισμού [Transformation of Costume in the Times of Nation Building of the New Hellenism].” *Πρακτικά Επιστημονικού Συμποσίου: Ο Ρομαντισμός στην Ελλάδα* (Athens), 127–145.
- Duque, Tiago. 2020a. “NLP Preprocessing Pipeline — What, When, Why?” *Medium*, accessed February 7, 2022. <https://medium.com/analytics-vidhya/nlp-preprocessing-pipeline-what-when-why-2fc808899d1f>.
- . 2020b. “Text Normalization — Why, What, and How.” *Medium*, accessed February 7, 2022. <https://towardsdatascience.com/text-normalization-7ecc8e084e31>.
- Firth, John Rupert. 1957. “Applications of General Linguistics.” *Transactions of the Philological Society* 56 (1): 1–14.
- Gauntlett, Stathis. 1983. “Rebetiko Tragoudi as a Generic Term.” *Byzantine and Modern Greek Studies* 8:77–102.
- . 1985a. *Rebetika: Carmina Graeciae Recentioris*. Vol. 1. Athens: Denise Harvey & Company.
- . 1985b. *Rebetika: Carmina Graeciae Recentioris*. Vol. 2. Athens: Denise Harvey & Company.
- . Forthcoming. ““Γράμματα, Χασάπη!” Bringing Rebetes and Rebetologists to Book.” *Popular Music of the Greek World. British School at Athens*.
- Gauntlett, Stathis / Payvanas, D. / Chatzinikolaou, Anna. 1994. “Corpus Rebetorum: a Preview.” *Modern Greek Studies (Australia and New Zealand)* 2.
- Genitsaris, Michalis. 1992. *Μάγκας από Μικράκι. Αυτοβιογραφία [Mangas from the Early Age. Autobiography]*. Edited by Stathis Gauntlett. Athens: Dodoni.
- Gharibi, Mohamed. 2019. “About Text Vectorization. The Magic of Converting Text to Numbers.” *Towards Data Science*, accessed February 18, 2022. <https://towardsdatascience.com/the-magic-behind-embedding-models-part-1-974d539f21fd>.
- Gotsinas, Kostis. 2021. *Επί της Ουσίας: Ιστορία των Ναρκωτικών στην Ελλάδα (1875-1950) [On the Substance: The History of Drugs in Greece (1875-1950)]*. Ηράκλειο: Πανεπιστημιακές Εκδόσεις Κρήτης - Ecole française d’Athènes.

- Holton, David / Horrocks, Geoffrey / Janssen, Marjolijne / Lendari, Tina / Manolesou, Io / Toufexis, Notis. 2019. *The Cambridge Grammar of Medieval and Early Modern Greek*. Cambridge University Press.
- Horrocks, Geoffrey. 2010. *Greek: A History of the Language and its Speakers*. John Wiley & Sons.
- Hulstaert, Lars. 2017. "LDA2vec: Word Embeddings in Topic Models." *Towards Data Science*, accessed February 20, 2022. <https://towardsdatascience.com/lda2vec-word-embeddings-in-topic-models-4ee3fc4b2843>.
- Karani, Dhruvil. 2018. "Introduction to Word Embedding and Word2Vec." *Towards Data Science*, accessed February 16, 2022. <https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa>.
- Kocmi, Tom / Bojar, Ondřej. 2017. "An Exploration of Word Embedding Initialization in Deep-Learning Tasks." *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)* (Kolkata), 56–64.
- Kostis, Kostas. 2018. *History's Spoiled Children: The Formation of the Modern Greek State*. London: Hurst & Company.
- Kotzamanis, Byron. 1997. "Αθήνα, 1848-1995. Η Δημογραφική Ανάλυση μιας Μητρόπολης [Athens, 1848-1995. A Demographic Analysis of a Capital]." *Επιθεώρηση Κοινωνικών Ερευνών* 92:3–30.
- Kounadis, Panagiotis. 2003a. *Εις Ανάμνησιν Στιγμών Ελκυστικών [In Remembrance of Attractive Moments]*. Vol. 1. Athens: Katarti.
- . 2003b. *Εις Ανάμνησιν Στιγμών Ελκυστικών [In Remembrance of Attractive Moments]*. Vol. 2. Athens: Katarti.
- . 2010. *Τα Ρεμπέτικα [The Rebetika]*. Vol. 11. Athens: Ta Nea.
- Lemnitzer, Lothar / Zinsmeister, Heike. 2015. *Korpuslinguistik*. Narr Francke Attempto Verlag.
- Levinthal, Charles F. 2012. *Drugs, Behavior, and Modern Society*. Boston: Pearson.
- Liavas, Lambros. 2009. *Το Ελληνικό Τραγούδι από το 1821 έως τη Δεκαετία του 1950 [The Greek Song from 1821 till 1950s]*. Athens: Emboriki Trapeza tis Ellados.
- Lukei. 2019. "Dealing with Contractions in NLP." *Medium*, accessed February 19, 2022. https://medium.com/@lukei_3514/dealing-with-contractions-in-nlp-d6174300876b.
- Maniatis, Dionisios. 2001. *Οι Φωνογραφήτσηδες. Πρακτικών Μουσικών Εγκώμιον ή Το Γραμμόφωνο, η Εποχή του και οι Γραμμοφωνήσεις των Ελληνικών Λαϊκών (κυρίως Ρεμπέτικων) Τραγουδιών [The Phonographists. In Praise to Practicing Musicians Or the Gramophone, its Era and the Gramophone Records of Greek Laiko (mainly Rebetiko) Songs]*. Athens: Pitsilos.
- Manolis Triantafyllidis Foundation. 1998. "Λεξικό της Κοινής Νεοελληνικής [Dictionary of Standard Modern Greek]." (Thessaloniki), accessed February 7, 2022. https://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/search.html.

- Moustakas, Petros. 2019. “Το «Τρίχορδο» και το «Τετράχορδο» Αρμάτωμα των Μπουζουκιών του Αναστάσιου Σταθόπουλου στις Αρχές του 20ου Αιώνα. [Three- and Four-course Bouzouki Variants by Anastasios Stathopoulos at the Beginning of the 20th Century].” *TAR. Μουσικό διαδικτυακό περιοδικό με αφορμή την κιθάρα*, accessed February 7, 2022. https://www.tar.gr/to_trixordo_kai_to_tetraxordo_armatwma_twn_mpozoykiwn_toy_anast_stathopoyloy-article-5914.html?category_id=90.
- Papadopoulos, Gerasimos Sofoklis. 2016. “Από την Αργκό του Ρεμπέτη στο Καθημερινό Λεξιλόγιο [From the Argot of a Rebetis to the Everyday Language].” *Μελέτες για την ελληνική γλώσσα (Studies in Greek Linguistics)* 36:509–524.
- Passow, Arnold, ed. 1860. *Τραγούδια Ρωμαϊκά: Popularia Carmina Graeciae Recentioris*. Leipzig: Teubner.
- Pennanen, Risto Pekka. 1997. “The Development of Chordal Harmony in Greek Rebetika and Laika Music, 1930s to 1960s.” *British Journal of Ethnomusicology* 6 (1): 65–116.
- . 2004. “The Nationalization of Ottoman Popular Music in Greece.” *Ethnomusicology* 48 (1): 1–25.
- Petropoulos, Elias. 1991. *Ρεμπέτικα Τραγούδια [Rebetiko Songs]*. Athens: Kedros.
- Politis, Nikos. 2006. “Μία Παρουσίαση για τον Αμανέ [A Presentation on Amanes].” *Το λαϊκό τραγούδι (Athens)* 16. Accessed November 18, 2021. <https://rembetiko.gr/t/%CE%BC%CE%AF%CE%B1-%CF%80%CE%B1%CF%81%CE%BF%CF%85%CF%83%CE%AF%CE%B1%CF%83%CE%B7-%CE%B3%CE%B9%CE%B1-%CF%84%CE%BF%CE%BD-%CE%B1%CE%BC%CE%B1%CE%BD%CE%AD/37538>.
- . 2014. “A not so Widespread Perspective of Panos Toundas: The “European Style” Recordings.” *6ο Σεμινάριο-Συνάντηση για το Ρεμπέτικο. Το ρεμπέτικο τραγούδι, ως μία από τις βασικές διαχρονικές συνιστώσες του αστικού λαϊκού μας πολιτισμού [6th International Seminar-Meeting on the Rebetiko. The Rebetiko Song as one of the Basic Timeless Components of our Urban Popular Culture]* (Skyros), 37–46.
- Redhouse, James William. 1880. *Redhouse’s Turkish Dictionary: In Two Parts, English and Turkish, and Turkish and English*. Quaritch.
- Sarandakos, Nikos. 2013. “Η Ρεμπέτα, οι Ρεμπέτες και το Ρεμπέτικο [Rebeta, Rebetes and Rebetiko].” *Το ιστολόγιο του Νίκου Σαραντάκου, για τη γλώσσα, τη λογοτεχνία και... όλα τα άλλα [Blog of Nikos Sarandakos about language, literature and... all other things]*, accessed February 14, 2022. <https://sarantakos.wordpress.com/2013/12/11/rembeta/>.
- . 2015. “Μόρτηδες και Απόλοιμοι (μια Συνεργασία του Spiridione) [Mortides and Apoloimoi (a Collaboration with Spiridione)].” *Το ιστολόγιο του Νίκου Σαραντάκου, για τη γλώσσα, τη λογοτεχνία και... όλα τα άλλα [Blog of Nikos Sarandakos about language, literature and... all other things]*, accessed February 14, 2022. <https://sarantakos.wordpress.com/2015/09/15/mortu/>.
- Savvopoulos, Panos. 2006. *Περί της Λέξεως “Ρεμπέτικο” το Ανάγνωσμα... και άλλα [On the Word “Rebetiko”... and more]*. Athens: Odos Panos.
- Schorelis, Tasos. 1988. *Ρεμπέτικη Ανθολογία [Anthology of Rebetiko]*. Vol. I. Athens: Plethron.

- Sipl, Colin / Fuchs, Florian / Burghardt, Manuel. 2018. "Digital Dylan—Computergestützte Analyse der Liedtexte von Bob Dylan (1962–2016)." *Dhd 2018. Kritik der digitalen Vernunft*. (Köln), 379–383.
- Smaragdi, Marianna. 2012. *Fire, Poison, and Black Tears: Metaphors of Emotion in Rebetiko*. Lund: Centre for Languages / Literature, Lund University.
- Tait, Casey / Tabain, Marija / Sykes, Ingrid. 2014. "Stress-Meter Alignment in American Hip Hop." *Proceedings of the 15th Australasian International Conference on Speech Science and Technology* (Melbourne), 189–192.
- Temple, Shelby. 2019. "Word Clouds Are Lame." *Towards Data Science*, accessed February 13, 2022. <https://towardsdatascience.com/word-clouds-are-lame-263d9cbc49b7>.
- Tragaki, Dafni. 2009. *Rebetiko Worlds*. Newcastle: Cambridge Scholars Publishing.
- Triantafyllidis, Manolis et al. 2018. *Νεοελληνική Γραμματική (της Δημοτικής) [Grammar of the Modern Greek (Demotic) Language]*. Athens: Οργανισμός Εκδόσεως Διδακτικών Βιβλίων.
- Trypanis, Constantine Athanasius. 1981. *Greek Poetry: from Homer to Seferis*. Chicago: University of Chicago Press.
- Turpin, Myfany / Stebbins, Tonya. 2010. "The Language of Song: Some Recent Approaches in Description and Analysis." *Australian Journal of Linguistics* 30 (1): 1–17.
- Viti, Carlotta. 2008. "Genitive Word Order in Ancient Greek: A Functional Analysis of Word Order Freedom in the Noun Phrase." *Glotta*, 203–238.
- Webster, William. 1864. *The Syntax and Synonyms of the Greek Testament*. Rivingtons.
- Zeleps, Ioannis. 2001. *Rebetiko: Die Karriere einer Subkultur*. 203. Köln: Romiosini Verlag.
- "Εικονικό Μουσείο Αρχείου Κουνάδη [Virtual Museum of the Kounadis Archive]." 2019. Last accessed 17 November 2021. <http://www.vmrebetiko.gr>.