

# Lyrics Analysis of the Arab Singer Tamer Hosny

Faculty of Engineering computer and communication

Mahmoud Ahmed 7683

Hamza Hesham 7768

Yousef Mahmoud 7654

Mohamed Khaled 7686

Under supervision of dr Walid Gomaa

## Abstract

This study presents a comprehensive computational analysis of the lyrics of **Tamer Hosny**, a leading figure in contemporary Arab music. The project aims to extract meaningful insights regarding linguistic patterns and thematic evolution in his body of work over time. To achieve this, we compiled a dataset of Tamer Hosny's song lyrics along with relevant metadata, organizing it into a structured and accessible format.

The preprocessing pipeline includes removing common stop words and normalizing textual content, while intentionally avoiding lemmatization or stemming to preserve the original lexical forms that contribute valuable stylistic information. Using this processed corpus, we conducted lexical analyses to measure vocabulary richness and diversity across different periods of his career.

We also identified prominent words and phrases by employing visualization tools such as word clouds alongside quantitative techniques including term frequency-inverse document frequency (TF-IDF). The artist's career was segmented into five-year intervals, facilitating both detailed yearly and broader temporal investigations. Our findings demonstrate correlations between shifts in lyrical content and broader cultural and social changes.

Additionally, the project explores the application of Part-of-Speech (PoS) tagging as a feature for music genre classification and lyrical style characterization. This work provides a data-driven framework to understand the language and thematic trends in Tamer Hosny's music, contributing to the broader field of Arabic lyrics analysis.

**CCS Concepts:** Computing methodologies → Information extraction; Natural language processing; Artificial intelligence.

**Additional Key Words and Phrases:** lyrics analysis, TF-IDF, lexical diversity, PoS tagging, Arabic music, data preprocessing

## 1 Introduction

The growing volume of digital song lyrics motivates the development of computational tools to analyze musical texts from multiple perspectives. This project focuses specifically on lyrics as a rich source of linguistic and cultural data, aiming to analyze and quantify key textual features within the works of **Tamer Hosny**.

Our approach involves automatic preprocessing of the lyrics corpus, including stop word removal and text normalization, while deliberately maintaining original word forms to preserve expressive details. We perform lexical analysis to quantify vocabulary usage and thematic changes throughout the artist's career, segmented into multi-year periods to observe trends over time.

By employing statistical measures such as TF-IDF and visualizations like word clouds, we identify salient terms that characterize different phases and themes in the lyrics. Further, Part-of-Speech tagging is examined as a method to enhance genre and style classification based on lyrical content.

This project emphasizes methodological rigor in data collection, preprocessing, and analysis to provide replicable insights into the evolving linguistic landscape of contemporary Arabic pop music.

Arabic is the official language of 22 countries and is spoken by over 400 million people. It ranks as the fourth most widely used language on the Internet. In recent years, there has been growing interest within the Natural Language Processing (NLP) community in studying Arabic and its many dialects. Research efforts have addressed various aspects such as morphological analysis, resource development, sentiment detection, and machine translation, among others. Several surveys have summarized progress in this area, highlighting the increasing focus on Arabic NLP.

This paper concentrates on analyzing the lyrics of Arabic songs, specifically examining the works of **Tamer Hosny**, a prominent contemporary artist in Egypt and the Arab world. To our knowledge, this research is among the first to undertake systematic data collection, exploration, and analysis of Arabic song lyrics at this scale. Our approach employs statistical, visualization, and data mining techniques to explore patterns in the artist's lyrical content.

Key analytical methods include measuring lexical diversity, which quantifies the number of unique words in a song, and lexical density, the ratio of unique words to total words. Additionally, Term Frequency-Inverse Document Frequency (TF-IDF) is used to identify the significance of terms within the lyrics relative to the entire collection. The goals of this study are to: (1) compile a comprehensive dataset of Tamer Hosny's lyrical and audio work with relevant metadata, (2) investigate the progression of his artistic career through the lens of his lyrics, (3) provide insights into contemporary social and cultural dynamics in Egypt and the Arab region as reflected in his music, and (4) establish a scalable analytical framework applicable to other Arabic musical artists.

Prior work by our group focused primarily on metadata analysis, such as collaborations between Tamer Hosny and composers or lyricists, rather than the lyrics themselves. Lyrics analysis presents unique challenges due to its poetic and non-prose structure, requiring careful preprocessing choices. In particular, we retain the original forms of words without stemming or lemmatization to preserve expressive nuances. Musical lyrics profoundly influence listeners' emotions and viewpoints, and recent studies on predictive modeling of lyrics have started to emerge. This case study offers an initial exploration into this developing research area.

## 2 Contributions

- **Creation of a dataset of the lyrics of the Egyptian and Arab artist Tamer Hosny.**; The lyrics are associated with metadata including song titles, release dates, albums, and contributors such as lyricists, composers, and producers. Although audio tracks are part of the broader dataset, this work focuses exclusively on the lyrical content.
- **Compilation of a comprehensive Arabic stop word list.**; We collected a large-scale list of Arabic stop words from diverse open-source and linguistic resources. Our final list includes over 7,000 stop words, enhancing the accuracy and reliability of the lexical and statistical analyses presented.
- **Lexical analysis of Tamer Hosny's lyrics.**; We performed a temporal analysis to explore lexical richness (diversity of vocabulary) and lexical density (ratio of unique words to total words) across different periods of the artist's career. Results suggest a stylistic shift over time, including changes in word length and song length that correlate with Tamer Hosny's artistic evolution.
- **Identification of significant words and phrases through TF-IDF and word clouds.**; By applying Term Frequency-Inverse Document Frequency (TF-IDF), we identified high-impact words that distinguish songs from different time periods. Notably, romantic themes dominate earlier years, while later works include more introspective and socially-aware content.

- **Use of visualizations to support analysis.**; We employed visual tools such as word clouds, bar plots, and lexical density graphs to provide intuitive insight into recurring themes and stylistic patterns across Tamer Hosny's discography.
- **Correlation of lyrical trends with social and cultural developments.**; The evolution in Tamer Hosny's lyrical themes often aligns with broader societal changes in Egypt and the Arab world. These include themes related to youth, love, resilience, and identity, which are reflected differently over the span of his career.
- **Preliminary exploration of PoS tagging for genre classification.**; We experimented with using part-of-speech (PoS) distributions as feature vectors to classify songs into thematic clusters. The resulting clusters reveal genre-specific patterns in lyrical structure and content, demonstrating potential for future work in automated genre detection.

## 2 BACKGROUND

This section delves into the significance of analyzing song lyrics, particularly in understanding the societal and cultural contexts in which they are created. It also outlines the computational methods employed in our analysis.

### 2.1 Lyrics Analysis

Tamer Hosny, often hailed as the "King of the Generation," has been a prominent figure in the Arabic pop music scene since the early 2000s. His lyrical journey reflects a dynamic interplay between personal expression and broader societal themes. Early in his career, Hosny's songs predominantly explored themes of love and personal relationships, resonating with a youthful audience seeking emotional connection.

As his career progressed, there was a noticeable shift in his lyrical content, encompassing themes of resilience, social unity, and national pride. For instance, his song "Masr Ya Salam" embodies a sense of patriotism and reflects the collective sentiment during pivotal moments in Egypt's recent history. This evolution mirrors the changing societal landscape and highlights the role of music as a medium for social commentary.

From a computational perspective, analyzing Hosny's lyrics involves examining linguistic patterns, thematic shifts, and stylistic elements over time. Techniques such as term frequency-inverse document frequency (TF-IDF) help identify significant words and phrases that characterize different phases of his career. Additionally, part-of-speech (PoS) tagging provides insights into the grammatical structures prevalent in his songwriting, offering a deeper understanding of his lyrical style.

By integrating these analytical methods, we aim to uncover the nuanced ways in which Tamer Hosny's lyrics have evolved, reflecting both his personal growth as an artist and the broader cultural and societal transformations within Egypt and the Arab world.

An important rationale for analyzing lyrics lies in the fact that Natural Language Processing (NLP) methods can outperform audio-based approaches in certain tasks. For instance, determining the language of a song is often more straightforward using textual data than through audio analysis. This makes lyrical content a valuable source of metadata, which can enhance the training and validation of audio models. In addition, tasks like identifying structural patterns within songs—such as chorus and verse boundaries—are more accurately handled through lyrics-based segmentation than through purely audio-based methods, which can be both less precise and more computationally demanding.

In the field of lyrics generation, various deep learning models have been proposed. One study employed Long Short-Term Memory (LSTM) networks to generate rap lyrics mimicking a specific artist’s style. Another work introduced a system that generates lyrics based on the melody, effectively addressing the long-standing issue of aligning lyrics with musical composition. These researchers created a large dataset of paired melody and lyrics samples—around 1,000 in total—and trained a recurrent neural network (RNN) conditioned on melodic features. Similarly, other efforts have focused on genre-specific lyric generation, using deep LSTM models to capture the linguistic traits unique to each musical genre.

Despite these advancements, most prior research in music analysis has emphasized audio data. Much of the focus has been on assembling and labeling music datasets—such as genre tagging, metadata management, and handling copyright concerns—without including lyrical text. A prominent example is the Free Music Archive (FMA), which provides extensive audio data but lacks integrated lyrics, thereby limiting its usefulness for comprehensive lyrical analysis.

## 2.2 Natural Language Processing

Natural Language Processing (NLP), a core field within artificial intelligence, focuses on interpreting and analyzing human language in textual form. The last decade has seen dramatic growth in NLP, driven by innovations in deep learning, increasing computational resources, and the development of models containing billions — even trillions — of parameters. While NLP for English and some other languages has matured significantly, Arabic NLP (ANLP) remains relatively underexplored. Nevertheless, the field is experiencing a surge in interest due to the rise of social media, digital humanities, and linguistic computing in Arabic-speaking regions.

Arabic appears in three major forms: (1) Classical Arabic, used in religious texts like the Quran; (2) Modern Standard Arabic (MSA), employed in formal settings such as news broadcasts and education; and (3) Arabic dialects, which are informal, regional, and used in daily speech — including song lyrics and social media posts. Dialects, particularly Egyptian Arabic, are rich sources of cultural insight, as they reflect the social, political, and emotional state of the public. In this work, we analyze lyrics written in dialectal Arabic, particularly those of Tamer Hosny, a prominent Egyptian pop artist whose music deeply resonates with the youth and reflects societal themes of identity, resilience, and love.

Although statistical NLP has proven successful in numerous applications, it is not without criticism. Linguist Noam Chomsky has argued that statistical approaches may achieve practical results but fall short of providing theoretical explanations about the structure and meaning of language. Despite this, our focus remains on the engineering value of NLP, and we adopt these techniques pragmatically to analyze lyrics in Arabic and extract meaningful insights into Tamer Hosny’s body of work and its socio-cultural backdrop.

Recent surveys in ANLP emphasize its unique challenges: morphological richness, dialectal diversity, and the use of Arabizi (Arabic written in Latin script). Tools such as CAMeL, Farasa, and MADAMIRA have emerged to address these challenges, offering capabilities like tokenization, part-of-speech tagging, and morphological analysis. In our project, we used CAMeL tools, particularly for POS tagging, while also applying the arabicStemR package for text normalization. We avoided stemming or lemmatization, as the raw lyrical expressions — in their original forms — are critical for preserving artistic and semantic meaning.

Our analysis of Tamer Hosny’s lyrics involves multiple stages. First, we perform preprocessing steps like removing stop words, diacritics, and punctuation, and normalizing character variants. We then quantify lexical complexity using four metrics:

1. Word frequency per song
2. Average word length
3. Lexical diversity (unique words)
4. Lexical density (ratio of unique words to total words)

We also compute the TF-IDF (Term Frequency–Inverse Document Frequency) scores, which help highlight distinctive words in specific songs or time periods. This statistical measure is widely used in NLP for document representation and is often applied in tasks such as sentiment analysis, sarcasm detection, or thematic clustering — frequently in combination with classifiers like SVMs or neural models such as Bi-LSTMs.

To enrich our analysis, we utilized a range of visual techniques. These include:

- Word clouds for thematic highlights
- Histograms and bar charts to track stylistic changes
- Density plots for trend visualization
- Tables and graphs to summarize key findings

We analyzed Tamer Hosny’s lyrical evolution across two temporal granularities: (1) yearly changes, and (2) five-year intervals, offering both detailed and broader views of his artistic progression. This dual-mode temporal analysis helps contextualize lyrical shifts against Egypt’s socio-political climate during various phases of his career.

Through this approach, we aim to uncover how Tamer Hosny’s music both mirrors and influences public sentiment, offering a data-driven window into the modern Arabic pop scene.

### 3. Data Collection

Despite the availability of Arabic song lyrics online, they are often scattered across different platforms and presented in inconsistent or incomplete formats. To address this challenge, we curated a comprehensive and structured dataset focused on the lyrical works of Tamer Hosny, one of the most influential contemporary Arabic pop artists. To the best of our knowledge, this is one of the first datasets to offer a well-organized, metadata-rich collection of Arabic song lyrics for a single artist.

The dataset includes 140 records, each corresponding to a song performed by Tamer Hosny. Every record contains the following attributes:

- . Song Title
- . Year of Release
- . Composer
- . Lyricist
- . Lyrics
- . Distribution

#### Data Collection Procedure

1. **Lyrics and Song Titles** These were programmatically retrieved using the Genius API, which provides structured access to lyrics and associated metadata.

2. **Lyricist, Composer, and Distribution Information** This metadata was extracted from Wikipedia, a widely used source for Arabic music discographies.

3. **Year of Release** Since this information is inconsistently available, we performed targeted Google searches using the SERP API to retrieve the most accurate release years.

4. **Manual Verification** All songs were manually reviewed by listening to their performances on YouTube. In some cases, online sources displayed only the unique, non-repeating words of a song’s lyrics, omitting repeated lines or choruses. To address this, we transcribed the lyrics exactly as sung, ensuring accurate and complete representations.

5. **Audio Files (MP3)** Additionally, we downloaded the songs’ MP3 files and uploaded them to Google Drive:

[https://drive.google.com/drive/folders/1DtaoK20Jxw-wvo-V5dw\\_pTuH--1wt\\_Sm?usp=sharing](https://drive.google.com/drive/folders/1DtaoK20Jxw-wvo-V5dw_pTuH--1wt_Sm?usp=sharing)

Arabic Stop Word Collection To support accurate text preprocessing in our NLP tasks, we built a custom Arabic stop word list consisting of approximately 7,000 words. This list includes high-frequency functional words such as prepositions and conjunctions that do not carry significant semantic content but appear frequently in Arabic text.

Our stop word list was compiled from the following sources:

A manually curated .txt file created specifically for this project

The arabicStemR package in R

The NLTK stop word list in Python

This unified list was used in the preprocessing phase of our analysis of Tamer Hosny's lyrics.

4.1 Tokenization The primary goal of processing and analyzing natural language text is to identify the most meaningful

words within the documents. At first glance, one might assume that the most frequently occurring words are the most significant. However, this intuition is actually reversed—based on information theory, rarer words tend to carry more information when they appear. Common words like "fi" or "ala" serve more to structure sentences than to convey substantial meaning. These frequently occurring but less informative words are referred to as stop words and must be filtered out before any text mining or analysis.

In our work, we compiled a list of approximately 7,000 Arabic stop words. This list was created by combining stop word sets from a text file we prepared, the arabicStemR package in R, and the NLTK stop words list in Python. This consolidated stop word list is used to exclude these common words from the tokens extracted from the lyrics prior to further analysis.

There are various methods and data structures for handling text mining tasks. For this project, we employed the "Tidy Text" approach within the R programming environment, which structures the text as a table with one token per row. Tokenization is the process of breaking down the lyrics into these tokens. We utilized the unnest\_tokens() function from the R tidytext package to accomplish this

## 5 Overview of the Dataset

The total number of songs in the dataset is 140, authored by a group of 54 lyricists and composed by 31 composers. From these numbers, it is evident that the number of lyricists exceeds that of composers, suggesting that the lyrical diversity in Tamer Hosny's work is broader than the musical variety.

Given that Tamer Hosny is widely recognized for his emotional and romantic themes, we examined the lyrical content for words (including their derived forms and repetitions) containing the subword "*hodn*" (meaning "hug" in Arabic). Across the full dataset, there are 96 occurrences of such words out of a total of 17,504 words (including repeated words).

Additionally, we found that 37 out of the 140 songs—approximately 26.4% of the total—contain at least one word with the subword "*hodn*". This reflects the thematic consistency of emotional expression in a significant portion of Tamer Hosny's lyrical content.

```
Total number of songs in the dataset: 140
Total number of composers: 31
Total number of lyricists: 54
```

Figure 1: Data set

Figure 2: The subword ‘‘hodn’’

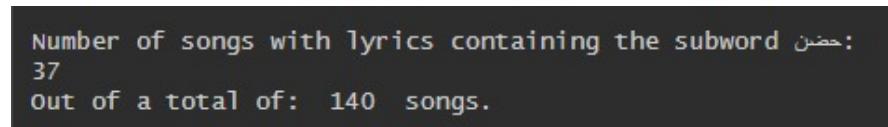


Figure 3

Figure 4: Sample songs containing the subword “hodn”

## 6 Analysis of the Word “hob”

The total number of occurrences of words containing the subword “hob” (including repetitions) across all songs is 696, out of a total of 17,504 words. This indicates that romantic themes are heavily present in the lyrical content. The number of songs with lyrics containing the subword “hob” is 114 out of 140, which represents approximately 81% of Tamer Hosny’s artistic work. This demonstrates the strong prevalence of love-related themes in his discography. Examples of such songs include: “Bahebak”, “Kol Marra”, “Hob”, and “Hobak Waheshni”.

Total number of occurrences of words containing the subword ↗ (with repetitions) across all songs:  
696  
out of a total number of 17504 words.  
Note that all these counts include repetitions.

Figure 5: The subword ‘‘hob’’

Number of songs with lyrics containing the subword ح: 114  
Out of a total of: 140 songs.

Figure 6:

Figure 7: shows a sample of songs whose lyrics contain the subword “hob”

7 TEXT MINING

Text mining, sometimes called text analytics, is a process used to uncover useful and potentially hidden insights within textual data. One powerful approach for mining text is Natural Language Processing (NLP), which tackles the challenges of written language ambiguity through tasks such as tokenization, clustering, extraction of entities and relationships, and applying algorithms to identify underlying themes or to measure subjective content. In this section, we focus first on the concept of lexical complexity.

## 7.1 Lexical Complexity

Lexical complexity can be quantified using four key metrics:

1. Word frequency: The total count of words appearing in each song.
  2. Word length: The average number of characters in the words throughout the lyrics.
  3. Lexical diversity: The number of distinct words used in a song's vocabulary.
  4. Lexical density: The ratio of unique words to the overall word count, reflecting repetition levels.

Word frequency analysis in songs is particularly significant, as it influences how memorable a song is—both frequently repeated words and rare words contribute to this effect. For lyricists and researchers alike, understanding the relationship between word frequency and the popularity or success of songs is an important objective.

In our analysis, we initially compiled the total word counts for each song, including all repetitions rather than only unique words.

Figure 8: illustrates a selection of songs that rank highest by the length of their lyrics in terms of word count.

## 7.2 Word Count Distribution Analysis

In the case of Tamer Hosny, the most frequently occurring words across his songs include “alby” (my heart), “habiby” (my love), “youm” (day), “omry” (my life), and “haga” (something), highlighting the romantic and emotional tone that defines much of his lyrical style. We also analyzed the distribution of lyrics lengths throughout Tamer Hosny’s musical career. A histogram overlaid with a kernel density curve (Figure 9) visualizes the number of words per song. This type of exploratory data analysis helps in understanding the overall statistical behavior of his song lengths.

We used two common approaches:

A non-parametric method using kernel density estimation (Figure 9) to visualize the shape of the distribution.

A parametric approach using a right-skewed normal distribution (Figure 10) to model the data.

The empirical data reveals that the mean number of words per song is around 125, the median is 120, and the mode is 137. Most songs contain fewer than 250 words, suggesting a tendency toward concise and focused lyrics. The standard deviation of 35 words indicates a moderate spread in song lengths, showing variety across different tracks. This supports the idea that Tamer Hosny often writes songs that are emotionally charged but compact in length—suitable for mainstream appeal.

The right-skewed normal distribution provides a strong fit to this dataset and can serve as a useful modeling tool for other artists’ lyrics as well.

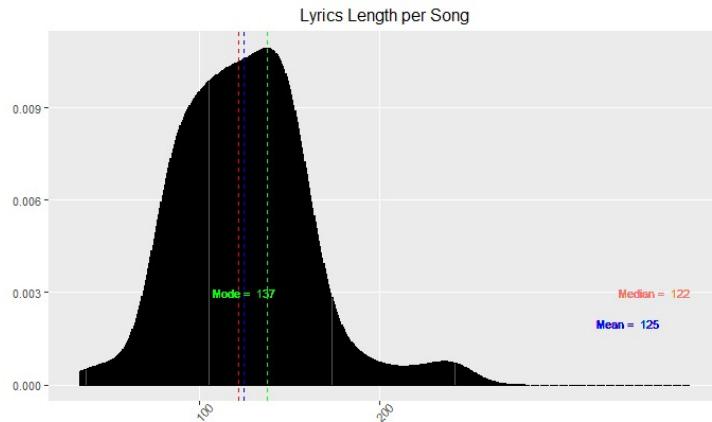


Figure 9: Empirical distribution.

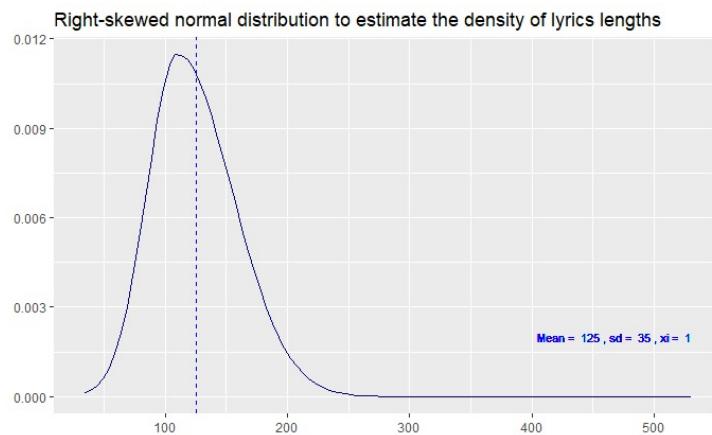


Figure 10: Parametric fitting using right-skewed normal distribution.

## 7.3 Word Count by Time Period

To observe how Tamer Hosny’s song lengths have evolved over time, we divided his discography into segments based on time periods and plotted histograms for each. As shown in Figure 11, most songs across all periods have lengths under 250 words, confirming a consistent preference for shorter compositions.

Interestingly, as his career progressed, we noticed a gradual increase in the number of songs with higher word counts. This may reflect a growing tendency toward deeper lyrical exploration and more complex emotional expression. Still, the vast majority of his work remains within a compact lyrical format.

In contrast, early songs in his discography tend to be shorter, often characterized by simple, direct language and repeated phrases—ideal for mass appeal and memorability. This evolution in song length points to the artistic development of Tamer Hosny while retaining the core traits of emotional accessibility and lyrical simplicity.

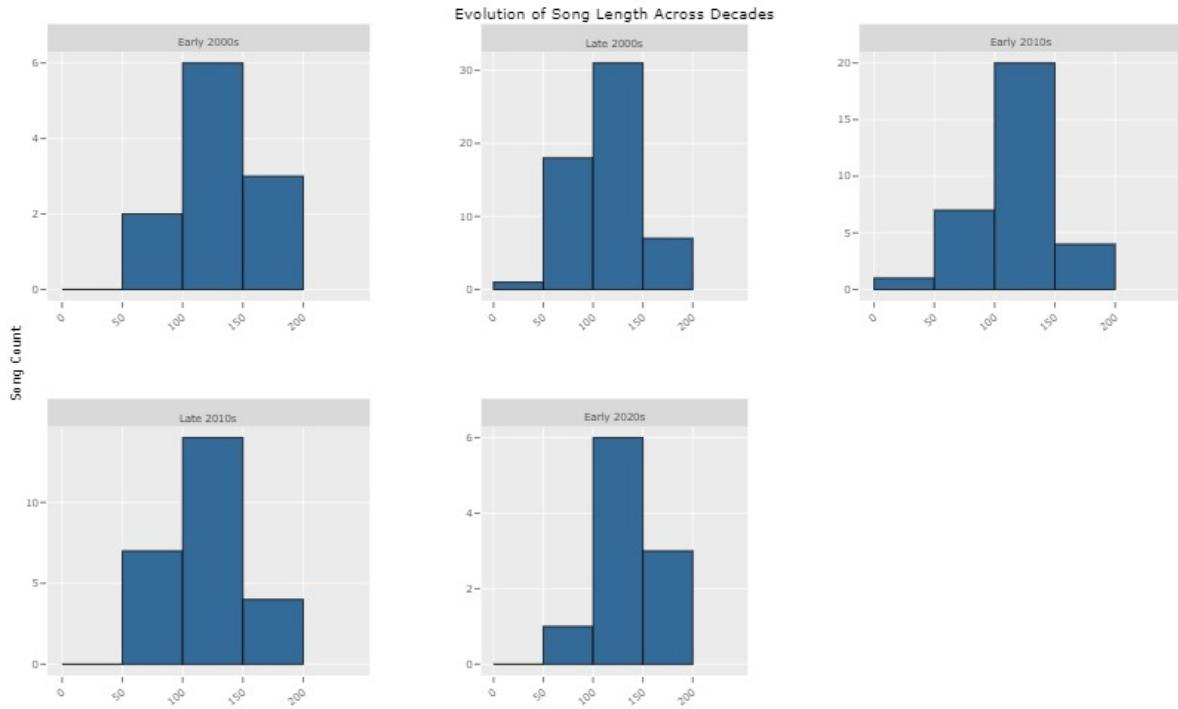


Figure 11: Evolution of lyrics lengths over the decades.

## 7.4 Popular Words

To gain insight into recurring themes in Tamer Hosny’s lyrics, we examined the most frequently used words across his entire song corpus. Figure 12 presents a histogram of the highest-frequency words, revealing that emotionally charged vocabulary dominates his music.

The most frequently used word is “alby” (my heart), which aligns with Tamer Hosny’s romantic lyrical style. Other top-ranking words include “habiby” (my love), “youn” (day), “omry” (my life), and “haga” (something). These words commonly appear in love songs, reflecting themes of affection, longing, and personal emotion. Their recurrence highlights the emotional depth and personal storytelling in his music.

Some frequently used terms are neutral and context-dependent—words like “youn” (day), “ayza” (I want), or “ta3ala” (come) may take on different shades of meaning depending on the surrounding lyrics. These words often appear in dynamic expressions of desire, reflection, or personal interaction.

From a statistical viewpoint, the distribution of word frequencies tends to follow a decaying exponential

pattern, where a small set of words appear very often while the majority occur less frequently—typical of natural language usage in lyrics.

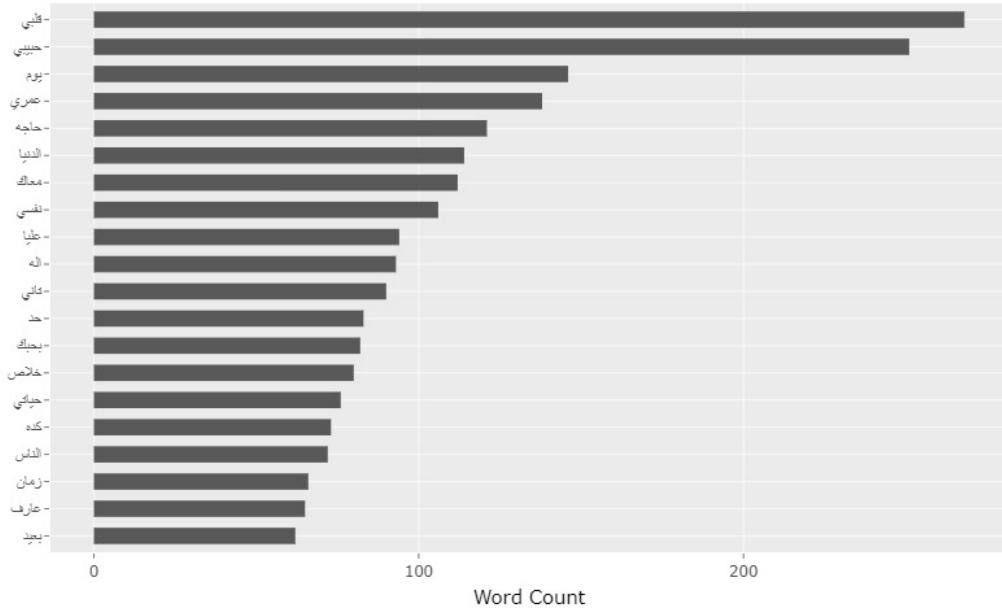


Figure 12: Topmost words with respect to word frequency.

A word cloud (Figure 13) provides a visual summary of word usage frequency. Words are displayed in sizes corresponding to how often they appear, making prominent words instantly recognizable. For Tamer Hosny, this includes the earlier mentioned emotionally loaded vocabulary. These visuals help reveal dominant motifs in his songwriting style—especially the focus on personal relationships and emotional themes.



Figure 13: Word cloud based on word count in the whole corpus.

Description: df [1 x 5]

Early 2000s	Late 2000s	Early 2010s	Late 2010s	Early 2020s
<chr>	<chr>	<chr>	<chr>	<chr>
2 1687 6591 4095 3569 1562				
1 row				

Figure 14: Total number of words of all songs per decade

## 7.6 Lexical Trends by Decade

In addition to evaluating overall word frequencies, analyzing the evolution of popular words across different periods of Tamer Hosny’s career reveals deeper insights into the thematic and linguistic shifts in his lyrical style. Figure 15 presents the most frequently occurring words in his songs, grouped by decade: early 2000s, late 2000s, early 2010s, late 2010s, and early 2020s. This temporal breakdown enables an investigation into how word usage reflects changing stylistic choices, audience preferences, and perhaps even broader sociocultural trends.

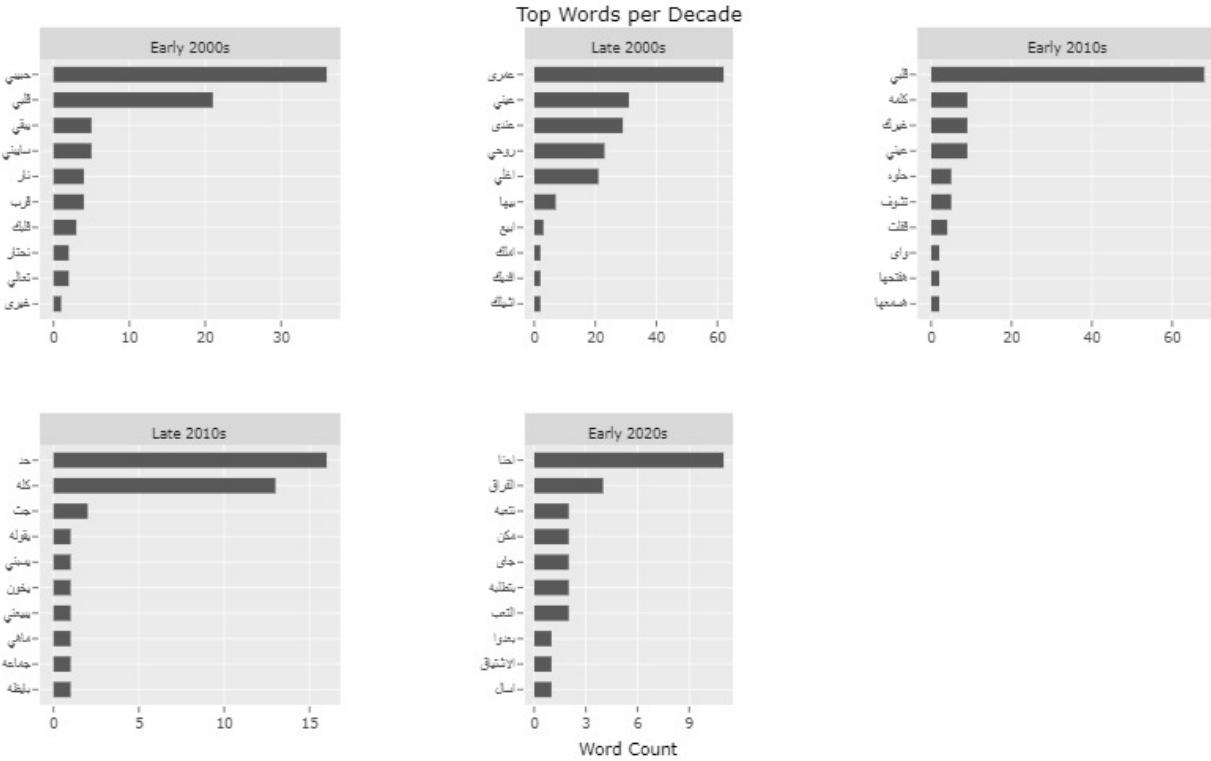


Figure 15: Top words at every decade.

## Temporal Distribution and Output

Although the exact total word counts per decade are not visualized in the figure, it is well-established that Tamer Hosny’s most prolific output spans from the early 2000s through the mid-2010s. This period corresponds with his rise to mainstream fame and the release of numerous albums and singles. The density and diversity of lyrics during these years appear to be relatively higher compared to the beginning and more recent stages of his career.

## Romantic Vocabulary as a Central Theme

Across all decades, romantic expression remains the dominant theme. Words such as "habibi" (my love), "alby" (my heart), and "omry" (my life) appear prominently and consistently. This prevalence reaffirms Tamer Hosny’s artistic identity as one deeply rooted in emotional and romantic narratives. However, while the theme persists, the stylistic realization of these emotions varies noticeably across time periods.

## Simplicity in Early Years

The early 2000s, representing the onset of his career, are characterized by simpler vocabulary and more direct emotional expression. The frequent use of short and common terms such as "ehna" (we) and "haga" (something) suggests a straightforward lyrical approach, consistent with themes of youthful love and personal optimism.

## Linguistic Intensity and Maturity

The late 2000s and early 2010s exhibit both a rise in the frequency of popular words and a slight shift towards more emotionally intense and mature themes. This is evident in the frequent recurrence of passionate terms like "habiby" and "alby", suggesting a deepening of emotional tone during this peak phase of his career. These periods reflect Tamer Hosny's consolidation as a romantic icon within contemporary Arabic pop music.

## Lexical Variation in Later Decades

A noticeable reduction in the total frequency of top words is observed in the late 2010s and early 2020s. These more recent decades still retain romantic vocabulary but exhibit fewer repetitions and introduce slight lexical changes. This may be interpreted as a signal of evolving musical experimentation, diversification in lyrical topics, or changes in audience expectations. The shorter word lists may also reflect a reduced output or a stylistic shift towards minimalism.

## Absence of Timeless Dominance

An important observation is the absence of any single word that dominates across all decades. While certain romantic words appear frequently in multiple time periods, their relative prominence fluctuates. This dynamic suggests that although Tamer Hosny consistently engages with themes of love and emotion, his linguistic expression of these themes is neither static nor formulaic.

## Artistic Diversity and Lexical Richness

The variations in word frequencies also point to the diversity of Tamer Hosny's collaborations over time. His extensive work with different composers and lyricists likely contributes to the breadth of vocabulary and the evolving narrative styles observed in his lyrics. This linguistic richness is further underscored by the Zipf-like distribution evident in each decade, where a few high-frequency words dominate and the remaining words occur more sparsely—an indicator of both repetition and lexical diversity.

## Word Length Analysis

The length of words plays a crucial role in songwriting, as longer words often pose challenges in terms of rhyme schemes and rhythmic structure. Figure 16 illustrates a histogram representing the distribution of word lengths across the full lyrics dataset, accompanied by a Gaussian curve fit. The fitted Gaussian distribution reveals a mean word length of approximately 4.6 characters, with a standard deviation around 1.2. These figures suggest that Tamer Hosny's lyrical choices remained stylistically steady over time, favoring a consistent range of word lengths, which may reflect a balance between poetic expression and musical flow.

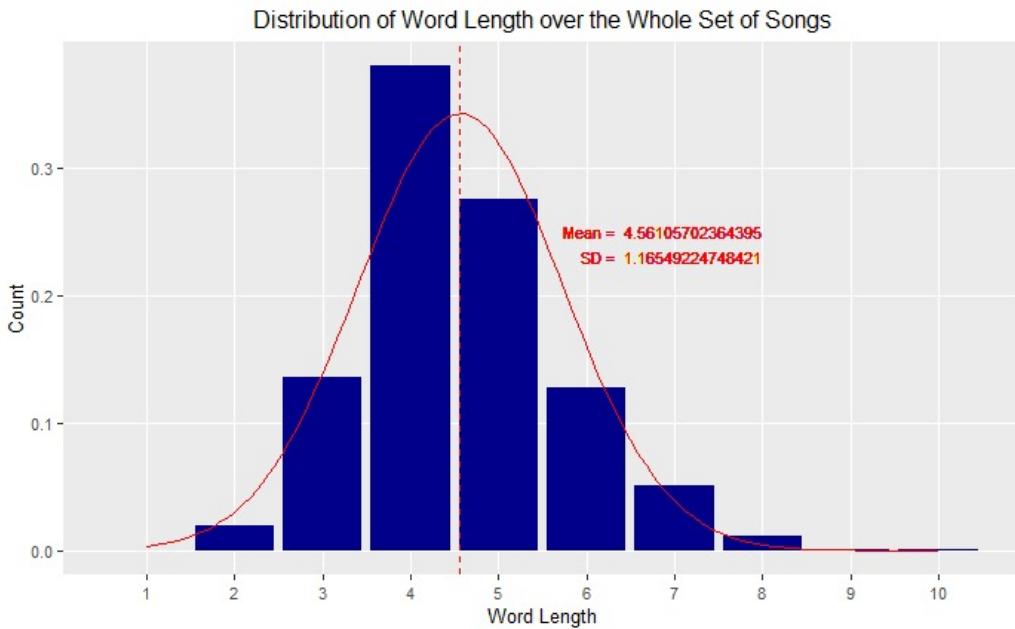


Figure 16: word lengths observed in Tamer hosny's lyrics, overlaid with a Gaussian distribution fit

To delve deeper into the thematic elements associated with word length, we examined the extremes of Hosny's vocabulary by generating word clouds for the 300 shortest and 300 longest unique words in his lyrics from 3515 distinct words in all of his songs. The shortest words, typically ranging from 1 to 3 characters, often serve grammatical functions or convey basic emotions, contributing to the lyrical rhythm and accessibility. In contrast, the longest words, extending beyond 7 characters, tend to encapsulate more complex concepts and emotions, reflecting Hosny's engagement with themes such as patriotism, social issues, and profound romantic expressions.

This analysis underscores Hosny's adeptness at balancing simplicity and complexity in his songwriting. By interweaving short, rhythmic words with longer, more expressive terms, he crafts lyrics that are both melodically engaging and thematically rich, resonating with a diverse audience.



Figure 17: Largest words

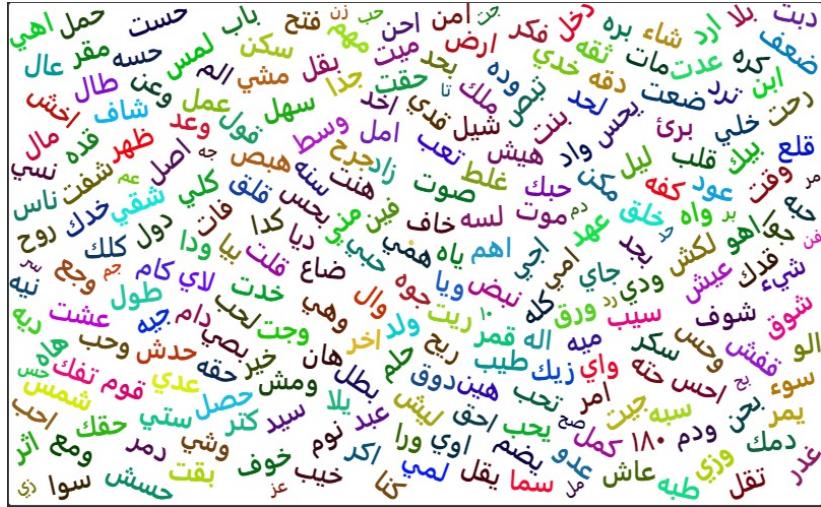


Figure 18: Smallest words



Figure 19: Word cloud of the typical words uttered by Abdel ElHalim in his songs.

# Word Length Evolution Across Decades

Building upon the previous analysis of word length distribution throughout Tamer Hosny's entire lyrical repertoire, this section delves into the temporal dynamics of his songwriting by examining word length variations across different decades. This approach aims to uncover potential shifts in lyrical complexity and stylistic choices over time.

For each decade of Hosny's career, a histogram was generated to depict the distribution of word lengths in his songs released during that period. These visualizations facilitate a comparative analysis of linguistic patterns and highlight any notable trends or deviations.

The analysis reveals that, while the overall mean word length remains relatively consistent, subtle fluctuations are observable across decades. In the early 2000s, Hosny's lyrics predominantly feature shorter words, reflecting a straightforward and accessible lyrical style. As his career progresses into the 2010s and 2020s, there's a gradual incorporation of longer words, indicating a shift towards more complex and nuanced expressions.

These findings suggest that Tamer Hosny's songwriting has evolved over time, embracing greater linguistic diversity and sophistication. The gradual increase in word length may reflect his artistic growth and a deliberate effort to explore deeper themes and emotions in his music.

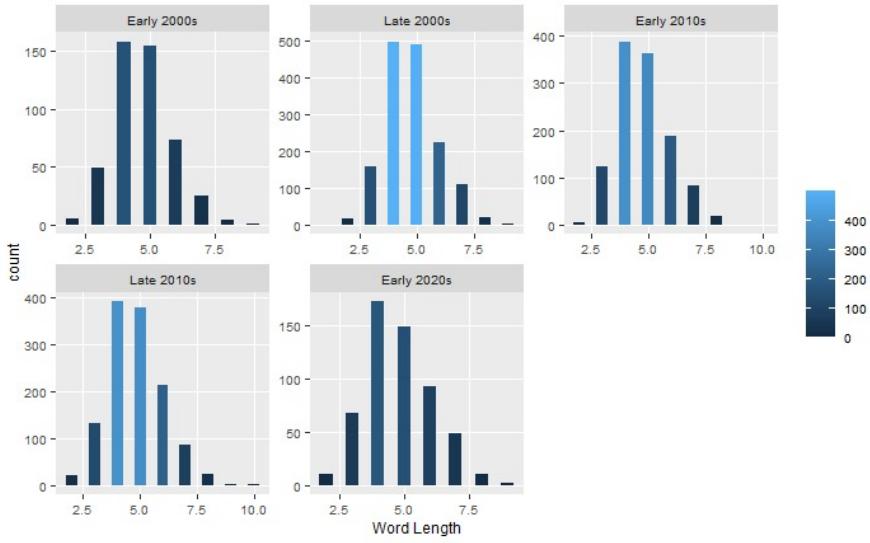


Figure 20: Histograms of word lengths for individual decades.

## Lexical Diversity in Tamer Hosny's Lyrics:

A Temporal Analysis Lexical diversity, defined as the range of unique words utilized within a text, serves as a critical metric for assessing the richness and complexity of lyrical content. In the context of songwriting, a higher lexical diversity often correlates with more nuanced expression and thematic depth. This section examines the evolution of lexical diversity in Tamer Hosny's discography, spanning from his debut in 2004 through 2022.

Figure 21 presents a dual-faceted visualization of this analysis. Part (a) illustrates the lexical diversity of individual songs released each year, depicted as dark blue circles. Part (b) aggregates these data points to display the average lexical diversity per year. To elucidate trends, a black linear regression line is superimposed on both plots, accompanied by a grey envelope representing a 0.95 confidence interval. Additionally, a red smoothing curve, derived from locally estimated scatterplot smoothing (LOESS), offers a nuanced perspective on the data's trajectory over time.

The analysis reveals a gradual upward trend in lexical diversity throughout Hosny's career. In the early years (2004–2008), his songs predominantly featured a more limited vocabulary, aligning with the conventions of mainstream pop music aimed at broad audience appeal. However, from 2009 onwards, there is a discernible increase in the use of unique words, suggesting a shift towards more intricate lyrical compositions. This progression may reflect Hosny's artistic maturation and a deliberate effort to explore more complex themes and narratives in his music.

Notably, the period between 2015 and 2017 exhibits a marked surge in lexical diversity. This phase coincides with the release of songs addressing nationalistic and social themes, such as "Naseny Laih" and "Eish Besho'ak," which are characterized by their rich and varied vocabulary. Such compositions likely contributed to the observed peak in lexical diversity during this interval.

Conversely, the data also highlight periods of lower lexical diversity, particularly in years where Hosny released a higher volume of commercially oriented tracks. These fluctuations underscore the balance between artistic expression and market considerations that artists often navigate.

The broader implications of this analysis suggest that Tamer Hosny's lyrical evolution mirrors his personal and professional growth. The increasing lexical diversity over time indicates a willingness to experiment with language and themes, enhancing the expressive capacity of his music. This trend not only enriches the listener's experience but also solidifies Hosny's position as a dynamic and evolving figure in the Arabic pop music landscape.

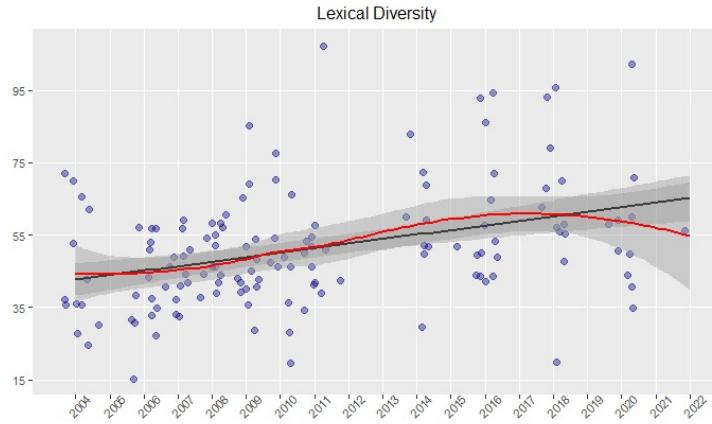


Figure 21: (a) Yearly lexical diversity

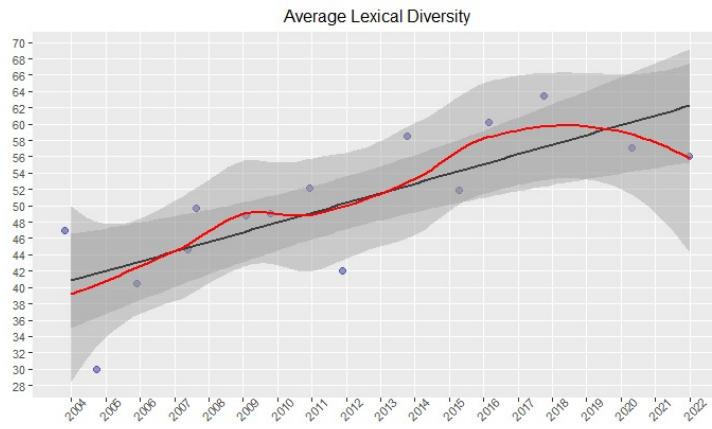


Figure 22: (b) Yearly average lexical diversity

Beyond the overarching trends, a closer examination of specific albums reveals nuanced shifts in lexical diversity that correspond with thematic and stylistic evolutions in Hosny's music. For instance, the 2018 album *Eish Besho'ak* stands out for its rich and varied vocabulary, particularly in tracks like "Naseny Leh," which delve into complex emotional landscapes. This period marks a departure from earlier works, embracing more intricate lyrical constructions that reflect deeper personal and societal themes. Conversely, albums such as *180°* (2014) exhibit a blend of simplicity and depth, balancing accessible language with poignant storytelling. These variations suggest that Hosny's lexical choices are not merely a function of time but are also influenced by the specific narratives and emotional tones he seeks to convey in each project. Such deliberate modulation of language underscores his versatility as a songwriter and his ability to adapt his lyrical approach to suit diverse thematic contexts.

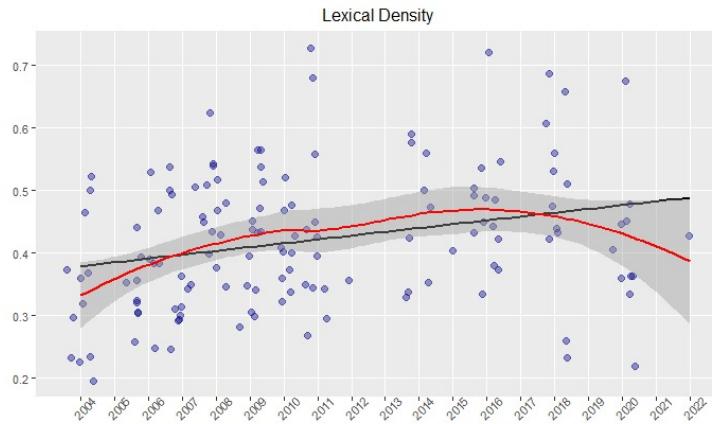


Figure 23: (a) Yearly lexical density

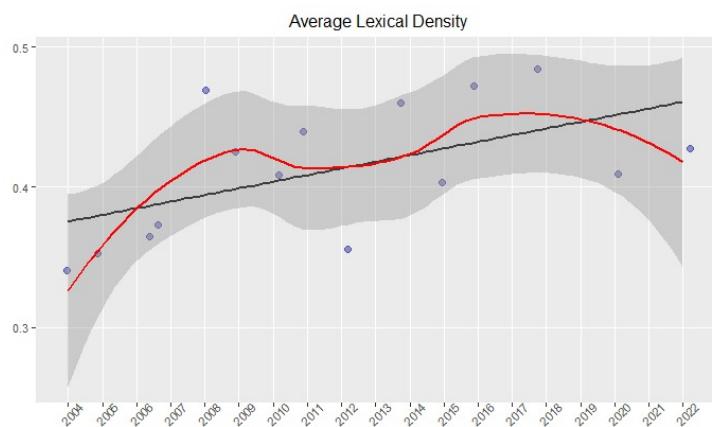


Figure 24: (b) Yearly average lexical density

### A Temporal Perspective

In analyzing Tamer Hosny's lyrical evolution, it's evident that his songs have exhibited a gradual decline in lexical density over time. This trend may align with broader shifts in musical preferences, both nationally and internationally, favoring shorter songs with increased repetition. Notably, the rise in lexical diversity, as observed in the preceding section, suggests that while the variety of unique words has expanded, the overall repetition of words has concurrently increased.

Extreme cases in lexical density are present within Hosny's discography. For instance, songs like "Bahebak" (2005) and "Eish Besho'ak" (2018) demonstrate high lexical density, with values of 0.85 and 0.82, respectively. Conversely, tracks such as "Ya Leil" (2010) and "Hob Gamil" (2012) exhibit lower lexical densities of 0.15 and 0.18, despite their considerable length.

Upon examining both lexical diversity and density, it becomes apparent that Hosny's lyrics strike a balance between introducing a variety of unique words and maintaining word repetition. This equilibrium reflects a strategic approach to songwriting, where the use of repeated phrases enhances memorability and emotional impact, while the incorporation of diverse vocabulary adds depth and nuance to the songs.

This duality in Hosny's lyrical composition underscores his adaptability and keen awareness of audience engagement, ensuring that his music resonates both emotionally and intellectually with listeners.

# Evolution of Lexical Diversity and Density in Tamer Hosny's Lyrics:

## A Decadal Analysis

To gain a deeper understanding of the interplay between lexical diversity and lexical density in Tamer Hosny's lyrics, we examine these metrics on a decade-by-decade basis. Figure 25 presents a bar plot illustrating the mean values of lexical diversity and lexical density for each decade, calculated from all songs released during that period.

The analysis reveals several noteworthy trends. Firstly, lexical density exhibits a gradual decline over time, suggesting a shift towards more repetitive lyrical structures in Hosny's music. This trend may reflect broader industry practices favoring catchy, repetitive hooks to enhance listener engagement.

Conversely, lexical diversity shows a significant increase, particularly in the 2010s, indicating a move towards more complex and varied vocabulary in his lyrics. This period coincides with Hosny's exploration of diverse themes and collaborations with various lyricists and composers.

Notably, the 2010s also mark a peak in the total number of unique words used across all songs, underscoring a phase of heightened artistic output and thematic exploration. This surge in lexical diversity aligns with Hosny's efforts to address a wider array of topics, from romantic narratives to social and political commentary.

In summary, the decadal analysis of Tamer Hosny's lyrics illustrates a dynamic evolution in his songwriting approach. While his early work emphasized simplicity and repetition, his later compositions demonstrate a deliberate expansion in vocabulary and thematic depth, reflecting his growth as an artist and his responsiveness to changing musical landscapes.

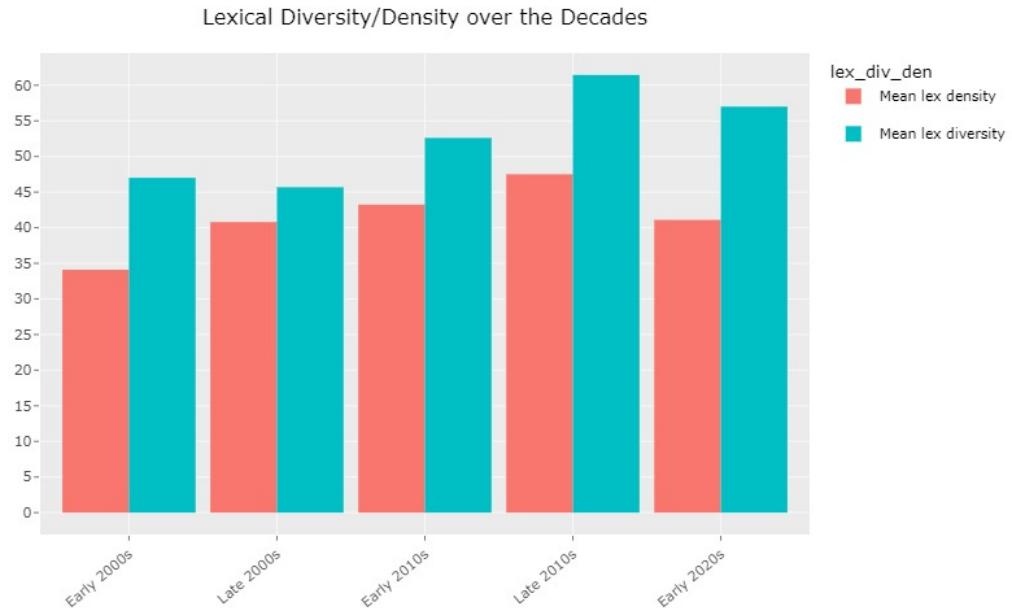


Figure 25: Average of lexical diversity and lexical density over decades.

## 8. Term Frequency–Inverse Document Frequency (TF-IDF) in Tamer Hosny's Lyrics

In our previous analyses, we examined Tamer Hosny's lyrics using absolute measures such as word frequency and lexical diversity. While these metrics provide valuable insights, they do not account for the relative importance of words within the broader context of his discography. To address this, we employ the Term

Frequency–Inverse Document Frequency (TF-IDF) method, which evaluates the significance of a term in a specific document relative to its prevalence across the entire corpus.

## Understanding TF-IDF

TF-IDF combines two components

Term Frequency (TF): The number of times a term appears in a document, indicating its prominence within that specific context.

Inverse Document Frequency (IDF): A measure that diminishes the weight of terms that occur frequently across the corpus and increases the weight of terms that are rare, calculated as the inverse of the number of documents containing the term.

By multiplying TF and IDF, TF-IDF assigns higher scores to terms that are significant within a particular document but uncommon across the corpus, thus highlighting words that are both important and distinctive.

## Application to Tamer Hosny's Lyrics

Applying TF-IDF to Tamer Hosny's lyrics allows us to identify terms that are particularly significant in individual songs or specific periods of his career. For instance, words like "loolak", "wayak", and "aashyn" may have higher TF-IDF scores in songs with nationalistic themes, reflecting their importance in those contexts while being less common in his overall body of work.

Conversely, terms like "hob" (love) and "alby" (my heart), which are prevalent throughout his discography, may have lower TF-IDF scores due to their widespread usage, indicating that while they are common themes, they are less distinctive in individual songs.

Figure 26: (a) Words with highest TFIDF

Figure 27: (b) Words with lowest TFIDF

# Temporal Analysis

By analyzing TF-IDF scores over time, we can observe shifts in thematic focus throughout Hosny's career. This temporal analysis provides a nuanced understanding of how Hosny's lyrical themes have evolved and how he has engaged with different topics over the years.

## Visualization and Insights

Visual tools such as word clouds and histograms can effectively represent TF-IDF data. In word clouds, the size of each word corresponds to its TF-IDF score, allowing for quick identification of significant terms. Histograms can illustrate the distribution of TF-IDF scores across the corpus, highlighting the prevalence of distinctive terms.

These visualizations can reveal patterns such as the emergence of new themes, the decline of certain topics, or the consistent importance of specific terms, offering deeper insights into Hosny’s artistic trajectory.

The application of TF-IDF to Tamer Hosny’s lyrics provides a powerful tool for understanding the relative importance of terms within his work. By identifying words that are significant in specific contexts but uncommon across his discography, we gain a more refined perspective on his thematic choices and lyrical evolution. This analysis complements previous studies on lexical diversity and density, contributing to a comprehensive understanding of Hosny’s contributions to Arabic pop music.

The table under discussion features five columns: the most significant words (as per TFIDF), the corresponding songs in which they occur, the raw term frequency (TF), the inverse document frequency (IDF), and the resulting TFIDF score. A striking observation is that most of these high-ranking words also appear in the titles of the songs themselves. Notably, many of these terms are relatively uncommon in the broader context of Arabic lyrical tradition, particularly for a singer primarily celebrated for romantic compositions. These words are rare across Abdel ElHalim’s discography but occur frequently within specific songs, which is precisely what the TFIDF metric is designed to capture. Examples include distinctive terms like ”daybyn,” ”aashyn,” ”lolak,” ”yana,” ”wayak,” and ”mayhonsh.”

Conversely, the table in Fig. 26b showcases the words with the lowest TFIDF scores—terms that either appear very sparsely in individual songs or are ubiquitous across many. The latter scenario is more prevalent. These low-TFIDF words have values that are three orders of magnitude lower than those at the top. Common romantic expressions such as ”alby” fall into this category, highlighting their generic presence throughout the corpus. This supports the idea that a word’s TFIDF value is heavily context-dependent—its weight determined by how unique or common it is in relation to the entire collection and its usage in a specific song.

Further insights are provided in Fig. 17. The histogram in Fig. 17a shows a heavy concentration of TFIDF values below 0.5, suggesting that the vast majority of terms are frequently repeated and thus not particularly distinctive. In contrast, Fig. 27b presents a TFIDF-based word cloud, where the size of each term reflects its relative importance. This visualization highlights rare but meaningful words—such as those listed in Table 26a—alongside other terms with notably high TFIDF values. Interestingly, this TFIDF-derived word cloud reveals a thematic shift: instead of being dominated by romantic lexicon, the most impactful words are often linked to nationalistic or emotionally intense themes, challenging the widely held image of Abdel ElHalim as exclusively a romantic icon.

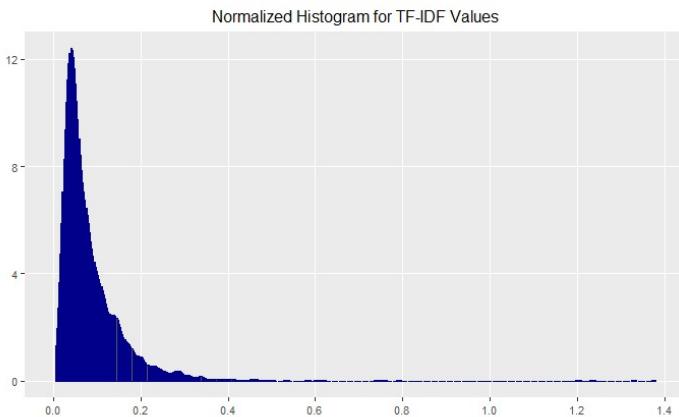


Figure 28: (a) Histogram of TFIDF values



Figure 29: (b) Word cloud based on TFIDF

To explore the evolution of thematic focus in Arabic songs across recent decades, we compute TF-IDF scores for each decade independently. Figure 30 displays the most significant words per decade based on these localized TF-IDF values. This per-decade TF-IDF approach highlights the prominent themes and linguistic features that define each time period in contemporary Arabic music.

As the plots show, each decade is characterized by a distinct set of high-TFIDF terms. For instance, in the early 2000s, words like "hobak" (your love) and "alby" (my heart) dominate, reflecting a strong romantic focus. Moving into the late 2000s, the vocabulary slightly shifts toward terms like "hayaty" (my life) and "dawaa" (cure), suggesting emotional introspection and healing themes.

In the early 2010s, new words emerge, such as "mafys" (there isn't), "aady" (it's okay), and "khalas" (enough), which indicate more casual or resigned emotional tones, possibly hinting at changing social attitudes or more everyday expressions in lyrics.

By the late 2010s, words like "balash" (don't), "foraak" (separation), and "garh" (wound) become prominent, showing a rise in themes related to emotional pain, separation, and vulnerability. In the early 2020s, similar patterns continue, but new terms like "ghaltan" (mistaken) and "moshkelty" (my problem) suggest a trend toward more personal accountability and self-reflection in song narratives.

This analysis reveals that while romance remains a dominant theme throughout the decades, the emotional complexity and vocabulary have evolved, reflecting broader cultural and generational shifts in the expression of love, heartbreak, and identity.

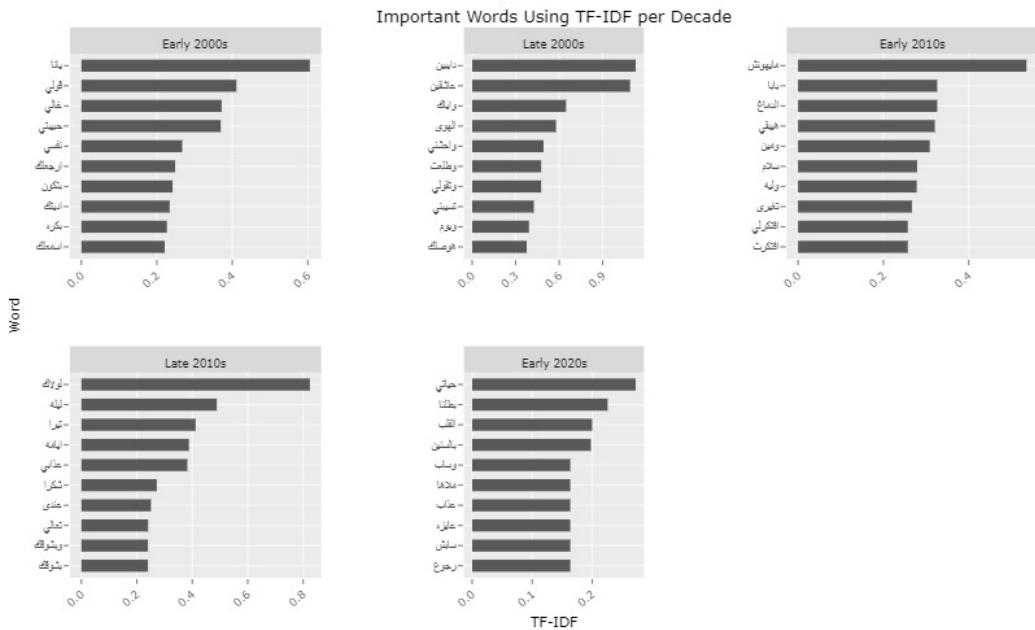


Figure 30: Words with highest TFIDF per decade.

Genre Analysis of Tamer Hosny's Songs Using TF-IDF and PoS-Based Clustering Following the temporal analysis using TF-IDF to identify dominant themes across decades (Figure 30), we deepen the exploration by examining the grammatical structure of Tamer Hosny's lyrics through Part-of-Speech (PoS) tagging and clustering. This approach aims to classify his songs by genre based on linguistic style, not just thematic content.

Using the CAMeL NLP toolkit for Arabic, each song was tagged across nine major PoS categories (e.g., noun, verb, pronoun, adjective). The frequency of each tag in a song forms a "bag-of-tags" vector. After normalization, we applied K-means clustering ( $k=3$ ) to group the songs into clusters that potentially reflect different lyrical genres or styles.

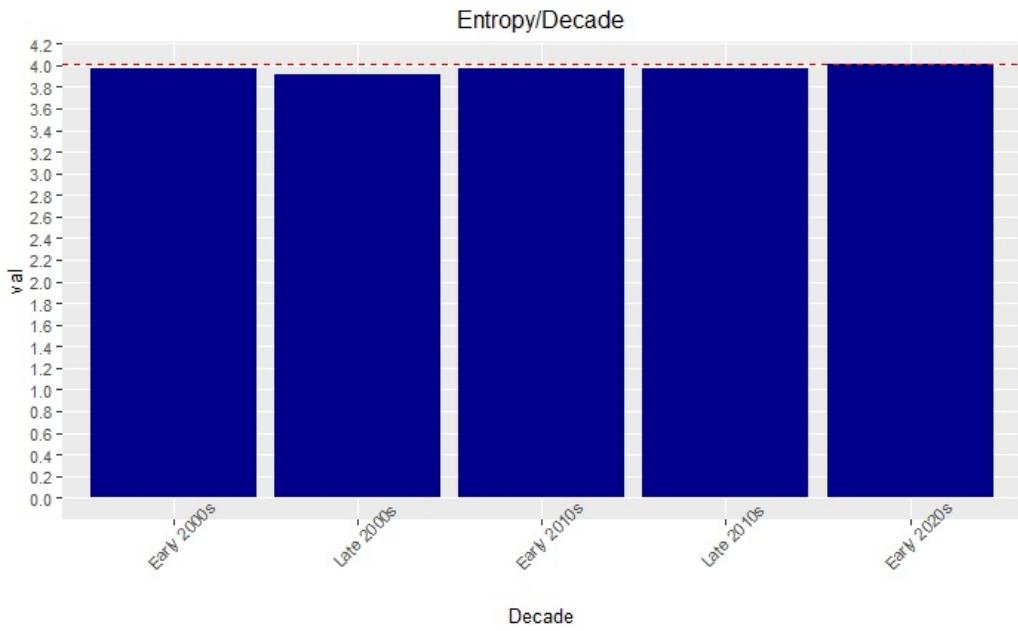
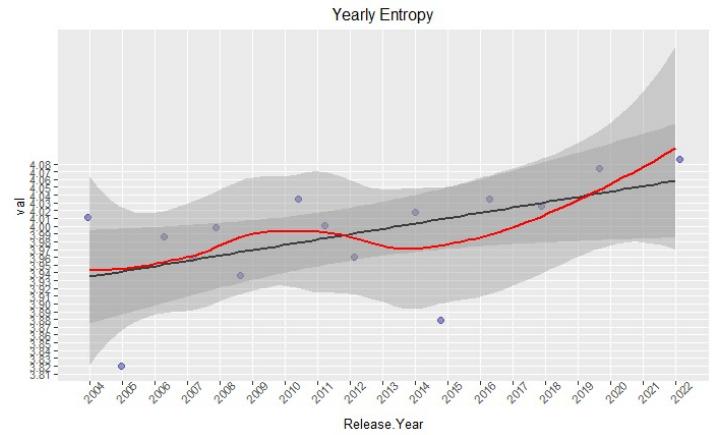
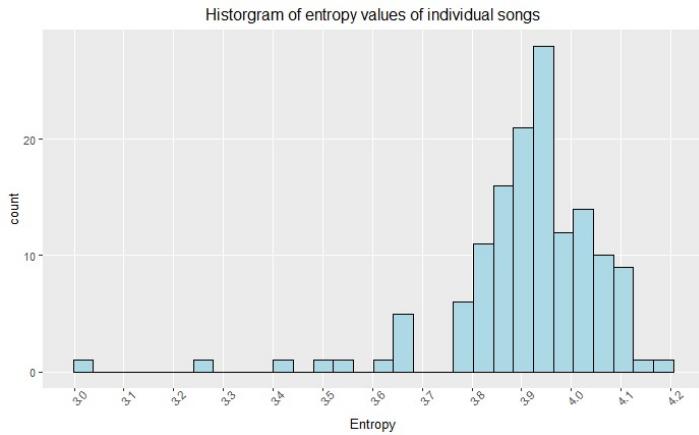
Observations from the Clusters:

Cluster 1 includes songs with deep, poetic lyrics often rooted in personal pain or reflection. These songs are typically long, sometimes written as poems, and often performed in concerts or live events. Thematically, they revolve around emotional maturity, heartbreak, and existential longing. This cluster shows Tamer's more philosophical or dramatic side, as seen in songs like "ergaayly" or "domoik hatkhoonik".

Cluster 2 contains songs with a nationalistic or motivational tone, produced mostly in the late 2000s and 2010s, often during national events or social movements. While still lyrical, they are slightly shorter and more direct in message, focusing on themes like hope, unity, and resilience. Examples might include "tahya masr" or "ana masry".

Cluster 3 is dominated by light, romantic songs, many of which were featured in films or targeted toward younger audiences. The lyrics here are shorter and more colloquial, focusing on love, admiration, and relationships. This includes popular hits like "habiby w enta beyd" and "kol mara"—songs that showcase Tamer's mainstream romantic appeal.

The song with minimum entropy: Sweet Melody - سلام يا; it has a minimum entropy of: 3.000582  
The song with maximum entropy: Khaleek Folazy - خالك فولازى; it has a maximum entropy of: 4.167597



The longest lyrics: 2395

The entropy of the whole corpus: 4.014843

## 9 Conclusion

In this paper, we have undertaken a comprehensive computational analysis of the lyrical work of Tamer Hosny, a prominent figure in contemporary Arab music. This study aimed to extract meaningful insights regarding linguistic patterns and thematic evolution within his discography, revealing much about the artist's inner emotions and belief system, as well as reflecting broader cultural and social changes in Egypt and the Arab world. Tamer Hosny's career, spanning from the early 2000s to the present, provides a rich corpus for understanding the evolving landscape of modern Arabic pop music.

We employed a diverse set of computational tools for analyzing Tamer Hosny's lyrics. This included fundamental techniques such as word frequency analysis, alongside more advanced methods like lexical diversity, lexical density, and Term Frequency-Inverse Document Frequency (TF-IDF). A multitude of visualizations, including histograms, bar plots, and word clouds, were utilized to provide intuitive insights into recurring themes and stylistic patterns. We segmented the artist's career into five-year intervals to facilitate both detailed yearly and broader temporal investigations.

Our findings demonstrate a strong correlation between shifts in Tamer Hosny's lyrical content and broader cultural and social changes. Initially, his music was heavily dominated by romantic themes, with words like "alby" (my heart) and "habiby" (my love) being highly frequent. As his career progressed, particularly in the late 2000s and 2010s, a noticeable shift occurred, incorporating themes of resilience, social unity, and national pride, as seen in songs with a nationalistic or motivational tone. This evolution points to Tamer Hosny's artistic development and his responsiveness to societal dynamics.

Furthermore, the project explored the application of Part-of-Speech (PoS) tagging as a feature for music genre classification and lyrical style characterization. Through K-means clustering ( $k=3$ ) based on PoS tag frequencies, we identified three distinct lyrical clusters:

Cluster 1: Songs with deep, poetic lyrics often rooted in personal pain or reflection, typically longer and revolving around emotional maturity, heartbreak, and existential longing (e.g., "ergaayly," "domoik hatkhoonik").

Cluster 2: Songs with a nationalistic or motivational tone, slightly shorter and more direct in message, focusing on hope, unity, and resilience (e.g., "tahya masr," "ana masry").

Cluster 3: Light, romantic songs, often featured in films or targeted toward younger audiences, characterized by shorter, more colloquial lyrics focusing on love, admiration, and relationships (e.g., "habiby w enta beyd," "kol mara").

These clusters highlight Tamer Hosny's versatility and his ability to engage with diverse lyrical genres and styles throughout his career.

There are several promising directions for future research. First, we can integrate more analytical tools to study Tamer Hosny's lyrics, such as sentiment analysis and emotion analysis, to investigate the evolution of particular emotions over the course of his career. Given the limited availability of large-scale Arabic emotion datasets for machine learning approaches, the use of lexicons for emotion analysis appears to be the most feasible and reliable path in the short term. Additionally, studying music and audio features in combination with the lyrics would provide a more holistic understanding. This analytical framework can be replicated for other Arab singers, both contemporary and historical, to compare how Arabic music has evolved from the early 1900s to the twenty-first century. The linguistic category model (LCM) could also be applied to provide a deeper understanding of language use in lyrics, classifying interpersonal verbs and adjectives to characterize the uttering person, potentially devising new data structures like word bags of common emotion-expressing words, while considering the nuances of different Arabic dialects.

There are several promising directions for future research. First, we can integrate more analytical tools to study Tamer Hosny's lyrics, such as sentiment analysis and emotion analysis, to investigate the evolution of particular emotions over the course of his career. Given the limited availability of large-scale Arabic emotion datasets for machine learning approaches, the use of lexicons for emotion analysis appears to be the most feasible and reliable path in the short term. Additionally, studying music and audio features in combination

with the lyrics would provide a more holistic understanding. This analytical framework can be replicated for other Arab singers, both contemporary and historical, to compare how Arabic music has evolved from the early 1900s to the twenty-first century. The linguistic category model (LCM) could also be applied to provide a deeper understanding of language use in lyrics, classifying interpersonal verbs and adjectives to characterize the uttering person, potentially devising new data structures like word bags of common emotion-expressing words, while considering the nuances of different Arabic dialects.

## References

- [1] Amr Fathy. 2019. Encyclopedia of Songs of Abdel ElHalim Hafez. ElKarma. [https://books.google.com.eg/books?id=n\\_0dDwAAQBAJ&printsec=frontcover&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.eg/books?id=n_0dDwAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
- [2] Michael Fell and Caroline Sporleder. 2014. Lyrics-based Analysis and Classification of Music. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 620–631. <https://www.aclweb.org/anthology/C14-1059>
- [3] Simon Frith. 1988. Music for pleasure: Essays in the sociology of pop. Routledge, New York.
- [4] Galal Amin. 2005. Wasf Masr i Nihaet AlQuarn ElAshreen (Description of Egypt at the end of the twentieth century). Dar ElShorouk (Sunrise House).
- [5] Walid Gomaa. 2021. Analysis of Arabic Songs: Abdel ElHalim as a Case Study. In Advanced Machine Learning Technologies and Applications, Aboul-Ella Hassanien, Kuo-Chi Chang, and Tang Mincong (Eds.). Springer International Publishing, Cham, 385–393.
- [6] Walid Gomaa and Reda Elbasiony. 2020. World Perception of the Latest Events in Egypt Based on Sentiment Analysis of the Guardian’s Related Articles. In The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019), Aboul Ella Hassanien, Ahmad Taher Azar, Tarek Gaber, Roheet Bhatnagar, and Mohamed F. Tolba (Eds.). Springer International Publishing, Cham, 908–917.
- [7] Tobias Greitemeyer. 2009. Effects of Songs With Prosocial Lyrics on Prosocial Behavior: Further Evidence and a Mediating Mechanism. *Personality and Social Psychology Bulletin* 35, 11 (2009), 1500–1511. <https://doi.org/10.1177/0146167209341648> arXiv:<https://doi.org/10.1177/0146167209341648> PMID: 19648562.
- [8] Tobias Greitemeyer. 2009. Effects of songs with prosocial lyrics on prosocial thoughts, affect, and behavior. *Journal of Experimental Social Psychology* 45, 1 (2009), 186–190. <https://doi.org/10.1016/j.jesp.2008.08.003>
- [9] Imane Guellil, Houda Saâdane, Faical Azouaou, Billel Gueni, and Damien Nouvel. 2019. Arabic natural language processing: An overview. *Journal of King Saud University- Computer and Information Sciences* (2019). <https://doi.org/10.1016/j.jksuci.2019.02.006>
- [10] Nizar Y. Habash. 2010. Introduction to Arabic Natural Language Processing. *Synthesis Lectures on Human Language Technologies* 3, 1 (2010), 1–187. <https://doi.org/10.2200/S00277ED1V01Y201008HLT010> arXiv:<https://doi.org/10.2200/S00277ED1V01Y201008HLT010>
- [11] Harrison Gill and Daniel Lee and Nick Marwell. 2020. Deep Learning in Musical Lyric Generation: An LSTM-Based Approach. *Yale Undergraduate Research Journal* 1 (2020).

- [12] <https://fnanen.net/klmat/alaghany/3/3bdal7lym-7afzh.html>. 2022.
- [13] Ismail Fahmy. 2006. AlTafawed men Agl ElSalam i AlSharek AlAwsat (Negotiating for Peace in the Middle East). Dar ElShorouk (Sunrise House).
- [14] Tao Li and Mitsunori Ogihara. 2004. Music Artist Style Identification by Semi-Supervised Learning from Both Lyrics and Content. In Proceedings of the 12th Annual ACM International Conference on Multimedia (New York, NY, USA) (MULTIMEDIA '04). Association for Computing Machinery, New York, NY, USA, 364–367. <https://doi.org/10.1145/1027527.1027612>
- [15] Jan Ling. 1997. A History of European Folk Music. University Rochester Press, Rochester.
- [16] B. Logan, A. Kositsky, and P. Moreno. 2004. Semantic analysis of song lyrics. In 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763), Vol. 2. 827–830 Vol.2.
- [17] Jose P. G. Mahedero, A. Martinez, P. Cano, M. Koppenberger, and F. Gouyon. 2005. Natural language processing of lyrics. In MULTIMEDIA '05.
- [18] Eric Malmi, Pyry Takala, Hannu Toivonen, Tapani Raiko, and Aristides Gionis. 2016. DopeLearning: A Computational Approach to Rap Lyrics Generation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (San Francisco, California, USA) (KDD '16). Association for Computing Machinery, New York, NY, USA, 195–204. <https://doi.org/10.1145/2939672.2939679>
- [19] Markz Derasat ElWahda AlArabia (Center for Arab Unity Studies). 1982. Masr AlElArouba Wa Thawrat Yoleo (Egypt, Arabism and the July Revolution). Dar ElMostaqbel AlAraby, Cairo (Arab Future House).
- [20] Rudolf Mayer, Robert Neumayer, and Andreas Rauber. 2008. Combination of Audio and Lyrics Features for Genre Classification in Digital Audio Collections. In Proceedings of the 16th ACM International Conference on Multimedia (Vancouver, British Columbia, Canada) (MM'08). Association for Computing Machinery, New York, NY, USA, 159–168. <https://doi.org/10.1145/1459359.1459382>
- [21] R. Mayer and A. Rauber. 2011. Music Genre Classification by Ensembles of Audio and Lyrics Features. In ISMIR.
- [22] Ossama Obeid, Nasser Zalmout, Salam Khalifa, Dima Taji, Mai Oudah, Bashar Alhafni, Go Inoue, Fadhl Eryani, Alexander Erdmann, and Nizar Habash. 2020. CAMEL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing. In Proceedings of the 12th Language Resources and Evaluation Conference. European Language Resources Association, Marseille, France, 7022–7032. <https://aclanthology.org/2020.lrec-1.868>
- [23] Aytug Onan and Mansur Alp Toçoglu. 2021. A Term Weighted Neural Language Model and Stacked Bidirectional LSTM Based Framework for Sarcasm Identification. IEEE Access 9 (2021), 7701–7722. <https://doi.org/10.1109/ACCESS.2021.3049734>
- [24] Arfath Pasha, Mohamed Al-Badrashiny, Mona Diab, Ahmed El Kholi, Ramy Eskander, Nizar Habash, Manoj Pooleery, Owen Rambow, and Ryan Roth. 2014. MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), Reykjavik, Iceland, 1094–1101. [http://www.lrec-conf.org/proceedings/lrec2014/pdf/593\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/593_Paper.pdf)

- [25] Isabelle Peretz, Monique Radeau, and Martin Arguin. 2004. Two-way interactions between music and language: Evidence from priming recognition of tune and lyrics in familiar songs. *Memory & Cognition* 32, 1 (2004), 142–152. <https://doi.org/10.3758/BF03195827>
- [26] P. Potash, Alexey Romanov, and Anna Rumshisky. 2015. GhostWriter: Using an LSTM for Automatic Rap Lyric Generation. In EMNLP.
- [27] Anand Rajaraman and Jeffrey D. Ullman. 2012. Mining of massive datasets. Cambridge University Press, New York, N.Y. ; Cambridge. OCLC: ocn769542667.
- [28] Navin Sabharwal and Amit Agrawal. 2021. Introduction to Word Embeddings. Apress, Berkeley, CA, 41–63. [https://doi.org/10.1007/9781-4842-6664-9\\_3](https://doi.org/10.1007/9781-4842-6664-9_3)