FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Patrícia Březinová

Computational analysis and synthesis of song lyrics

Institute of Formal and Applied Linguistics

Supervisor of the master thesis: Mgr. Martin Popel, Ph.D.
Study programme: Computer Science
Study branch: Artificial Intelligence

Prague 2021

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

Dedication.

Title: Computational analysis and synthesis of song lyrics

Author: Patrícia Březinová

Institute: Institute of Formal and Applied Linguistics

Supervisor: Mgr. Martin Popel, Ph.D., Institute of Formal and Applied Linguistics

Abstract: Abstract.

Keywords: key words

Contents

Introduction **2**

1 Related work **3**
1.1 Rhyme types and literary devices . . . 3
1.1.1 Basic rhyme types . . . 3
1.1.2 Other literary devices . . . 5
1.2 Rhyme detection tools . . . 6
1.2.1 Naive rule-based approach . . . 6
1.2.2 Advanced rule-based approach . . . 6
1.2.3 Similarity (distance) based approach . . . 7
1.2.4 Machine learning . . . 7
1.3 Visualization tools . . . 9
1.3.1 Poem Viewer . . . 9
1.3.2 SPARSAR . . . 10
1.3.3 ProseVis . . . 10
1.3.4 Poemage and RhymeDesign . . . 11
1.3.5 Ambiances . . . 12
1.4 Generation tools . . . 12
1.4.1 Generating using AI . . . 12

2 Data **14**
2.1 Preprocessing . . . 14
2.2 Structure of the data . . . 16

3 Lyrics evaluation - nechat si nazov pre kapitolu evaluation **18**
3.1 Rhyme detection . . . 18
3.1.1 Pronunciation . . . 18
3.1.2 Syllabification . . . 19
3.1.3 Calculating rating for one rhyme . . . 20
3.1.4 Calculating song rating . . . 22

4 Visualization **25**

5 Generation **26**

Conclusion **28**

Bibliography **29**

List of Figures **32**

List of Tables **33**

Glossary of literary and technical terms **34**

A Attachments **35**
A.1 IPA and ARPAbet transcription table . . . 35

Introduction

This works may include some literary or technical terms that the reader is not familiar with. For their definition, please see the chapter "Glossary of literary and technical terms" at the end of this thesis.

I can cite this Greene et al. [2010] when saying machine-generated poetry automatic evaluation is hard

Je otázka, zda bude vaše práce něČím specifická pro písně (oproti básním), krom toho, že jako data máte písně. U písní by samozřejmě šlo dělat analýzu zvukové stránky (melodie, harmonie, rytmus) a studovat korelace s textovou stránkou, ale to je zřejmě mimo rámec této práce. Mohla byste to ale zmínit někde na zaČátku, v zadání, že se schválně omezujete jen na ty (písňové) texty a ignorujete zvukovou stránku.

To round up the analyses and provide additional source of data for comparison, we produced computer generated data using current state-of-art pre-trained GPT-2.

1. Related work

This chapter gives a basic overview of all relevant tools and background information researched during work on this thesis. Firstly, it gives literary background necessary to make reader familiar with rhyme and its different types, to know what to look for before we start detecting them. Secondly, it describes existing tools for rhyme detection and visualization, and the different approaches they took. Lastly, it explains current state-of-art tools for lyrics generation.

1.1 Rhyme types and literary devices

There are many different definitions for what a rhyme is. It is described as "a word that has the same last sound as another word" by Cambridge Dictionary (Walter [2008]) or a "literary device, featured particularly in poetry, in which identical or similar concluding syllables in different words are repeated" by LiteraryDevices Editors [2020]. The definition of what a good rhyme is even changes for different languages and time periods (Zhirmunsky and Hoffmann [2013]). For example, full identity in sound is highly valued in French (*rime riche*), but less valued in English (perfect rhyme requires leading consonant sounds to differ). Some authors refrain from giving an exact definition and instead leave it to reader's intuition (Plecháč [2018]). We will define rhyme through its different types, what will be helpful for detection later.

1.1.1 Basic rhyme types

Perfect rhyme (also true rhyme, or sometimes just "rhyme") is the most common and superior type of rhyme. It requires two conditions to be met:

- last stressed vowel and all following sounds are identical
- immediately preceding sounds differ

It is also the only rhyme for which the definitions are consistent (for example, see Bain [1867], van der Schelde [2020], Bergman [2017], Wikipedia[1]). It can be further distinguished depending on how many syllables are involved:

- **Masculine** (also single, monosyllabic) – "the commonest kind of rhyme, between single stressed syllables at the ends of verse" (Baldick [2008]). Examples:

  fly /flaɪ/ – sky /skaɪ/

  before /bi.fɔːr/ – explore /ɪks.plɔːr/ [2] [3]

[1] https://en.wikipedia.org/wiki/Rhyme

[2] For the examples, we are using IPA transcriptions because it is more comfortable for human readers.

[3] Stressed syllables are underlined. Syllables are separated with a dot.

- **Feminine** (also double) – "a rhyme on two syllables, the first stressed and the second unstressed" (Baldick [2008]). Examples:

  bitten /bɪ.tən/ – written /rɪ.tən/

  lazy /leɪ.zi/ – crazy /textipakreI.zi/

- **Dactylic** (also triple) – "a rhyme on three syllables, the first stressed and the others unstressed"(Baldick [2008]). Examples:

  amorous /æ.mər.əs/ – glamorous /glæ.mər.əs/

  vanity /væ.nɪ.ti/ – humanity /hjuː-mæ.nɪ.ti)/

Imperfect rhyme (also slant or half rhyme) rhymes "the stressed syllable of one word with the unstressed syllable of another word" (Bergman [2017]). Examples:

cabbage /kæ.bɪdʒ/ – ridge /rɪdʒ/

painting /peɪn.tɪŋ/ – ring /rɪŋ/

In other sources, definitions differ – for example LiteraryDevices Editors [2020] calls this effect "feminine rhyme". On the other hand, Baldick [2008] and The Editors of Encyclopaedia Britannica [2014] use the term "imperfect rhyme" for end-line consonance (see definition below) and van der Schelde [2020] uses it for end-line assonance (see definition below). For the purpose of this thesis, we would like to keep rhyme types disjoint. Therefore we will require the sounds in the imperfect rhyme to be identical, except for the stress. This will differentiate it from *forced rhyme* (see below).

Unaccented rhyme (also weakened rhyme) "occurs when the relevant syllable of the rhyming word is unstressed" (The Editors of Encyclopaedia Britannica [2014]). Examples:

hammer /hæ.mər/ – carpenter /kɑːr.pən.tər/

The difference opposed to imperfect rhyme is that here rhyming parts of both words are unstressed. However, for simplicity, in the scope of this thesis we will include this category under *imperfect rhymes*.

Identical rhyme (also rime riche) is "a kind of rhyme in which the rhyming elements include matching consonants before the stressed vowel sounds." This includes "rhyming of two words with the same sound and sometimes the same spelling but different meanings e.g.:

seen /sin/ – scene /siːn/

The term also covers word-endings where the consonant preceding the stressed vowel sound is the same:

compare /kəm.pɛr/ – despair /dɪs.pɛr/." (Baldick [2008])

It is generally considered not as good as perfect rhyme because it is too predictable for the listener[4]. However, all rhyme detection tools as well as gold data that we will be using (annotated by professionals) include identity in perfect rhymes. To make the comparison and evaluation with our tool easier, we will do so as well.

[4] https://literaryterms.net/rhyme/

Forced rhyme (also near rhyme) "includes words with a close but imperfect match in sound in the final syllables" Bergman [2017]. Examples:

green /griːn/ – fiend /fiːnd/

hide /haɪd/ – mind /maɪnd/

This includes the case when spelling is changed in order to make the rhyme work, e.g.:

truth /truθ/ – endu'th /en.duθ/ (a contraction of "endureth")

It can also refer to using unnatural word order to get the rhyming word at the end of the line (Bergman [2017]) but we will not make use of this interpretation in this thesis.

1.1.2 Other literary devices

This is a short overview of other literary devices that closely correlate with forced rhyme and may, according so some sources, be considered a rhyme. We will conservatively exclude these from our classification and focus solely on rhymes occurring at the end of verse.

Assonance is "repetition of stressed vowel sounds within words with different end consonants" (The Editors of Encyclopaedia Britannica [2014]). Examples:

quite /kwaɪt/ – like /laɪk/

free /friː/ – breeze /briːz/

When used at the end of verse with ending consonants having a similar sound, it is equal to forced rhyme. However, the term itself defines a literary device applicable anywhere in the poem, even in the middle of the verse. Some sources classify it as rhyme, giving it various names (van der Schelde [2020], Bergman [2017], and others).

Consonance is "the recurrence or repetition of identical or similar consonants" (The Editors of Encyclopaedia Britannica [2014]). Examples:

country (kən-tri) / contra (kɑn-trə)

hickory dickory dock (hɪ-kə-ri dɪ-kə-ri dɑk)

Similarly as assonance, it applies to repetition of consonants in any part of the verse. When seen at the end of verse, it can be considered a rhyme and again, various terms are used – perhaps the most common is "pararhyme" (The Editors of Encyclopaedia Britannica [2014], Baldick [2008]).

The last two terms may seem as more of a tool for poets than songwriters. Surprisingly, they have found their way into song lyrics and have become a standard in genres like hip hop according to van der Schelde [2020]. From the creative point of view, it is not less sophisticated rather it enriches rhyme as we know it (Brogan [2016]).

Other rhyme types exist e.g. eye rhyme where "the spellings of the rhyming elements match, but the sounds do not, e.g. love /ləv/ – prove /pruːv/" (Baldick [2008]). We do not consider them relevant for song lyrics or the purpose of this thesis.

1.2 Rhyme detection tools

We have defined what to look for, and now we will focus on how to do it. According to Plecháč [2017], there are 4 methods for rhyme detection. We will describe each one and evaluate existing tools. For our use case, it is important that we can use the tool for automatic evaluation - there must be a way to run it with code whether it would be an API, or an executable script, or as a library/module. Another requirement is for it to be able to run on a block of text and generate rhyme scheme as a result. Lastly, it has to be free and preferably open-source.

1.2.1 Naive rule-based approach

The simplest approach is to compare for identity of phonemes at the end of lines. Noticeably, this only detects perfect rhymes and identity. Nevertheless the result will seem decent because it has 100% recall and notices all "obvious" rhymes. Another downside of this approach is its limitation to the size of the dictionary. There are many rhyming dictionaries (of various quality and size) to choose from but the vast majority of them uses or enhances CMU dictionary[5].

Pronouncing and CMU dictionary

Pronouncing[6] is a Python library providing an interface for CMU Pronouncing Dictionary. One possibility is to install CMUdict directly, search for pronunciations for both words and compare then. *Pronouncing* searches the dictionary automatically for a given word and returns a list of rhyming words. However, the list is truncated and probably better suited for writer's inspiration only.

1.2.2 Advanced rule-based approach

Datamuse

Rhyme Genie

SPARSAR

SPARSAR (Delmonte and Prati [2014]) is also a very interesting tool for poetry analysis and expressive Text-to-speech conversion. It is originally designed for a thorough examination of a very strictly structured Shakespeare's sonnets. To achieve this, it has to run analyses on many levels – and these results can be used to analyze any poem. It looks at the poem on three levels: phonetic (pronunciation, consonant and vowel tongue position, assonances, etc.), poetic (metrical structure, rhyme schemes, acoustic length, etc.), and semantic (sentiment, methaphorically linked words, anaphora, etc.).

User can choose between a window application with graphs and diagrams or a headless mode with .xml output files. Its main disadvantage for our use case is that it is written in Prolog and therefore is very strict on the input format and runs only under a specific older version of Ubuntu.

[5] http://www.speech.cs.cmu.edu/cgi-bin/cmudict
[6] https://pypi.org/project/pronouncing/

1.2.3 Similarity (distance) based approach

1.2.4 Machine learning

EM algorithm

Reddy and Knight [2011] proposed a language-independent model for finding rhyme schemes in poetry. They created an unsupervised model based on EM algorithm that assigns the most probable rhyme scheme for each sequence of line-final words. It achieved good results when tested on annotated English and French corpus with poetry from 15th to 20th century. However its big pitfall lies in the fact that it is biased towards the rhyme schemes from golden data. It has a predefined set of all rhyme schemes found in tested data and those are the only ones it chooses from. For illustration, in a 14-line stanza it can choose from 90 schemes which is only 0.00005% of all possible options. In 29% of cases from French corpus it has only one choice.[7]

RhymeTagger

Plecháč [2018] came with a collocation-driven alternative named RhymeTagger. It uses the same dataset as the previous approach with addition of a larger Czech poetry corpus[8]. Each line-final word is transcribed into phonetic transcription and split into components – syllable peak for each syllable and consonant cluster in between. In the "expectation" step, probabilities for each component pair are calculated based on their co-occurrence in line-final words, e.g. conditional probability of rhyme based on peak component pair əʊːəʊ will be very high but for consonant component pair kːr quite low. These statistics for component pairs are then used in the "maximization" step to calculate the probability for line-final word pair as a combined probability of all their components (paired by means of usual association measure). If the probability of two words is above a given threshold they are considered a rhyme. After all such pairs are classified, probabilities are iteratively recalculated in the EM cycle.

For words that were not successfully classified with this method, there is a fallback. The author observed that some words are now pronounced differently than they were during the Shakespearean era they were written in and therefore using current pronunciation dictionaries may ruin the original rhyme, e.g. original near (nɛːr) / there (ðɛːr) vs. contemporary near (nɪːr) / there (ðɛːr). Original pronunciation can be therefore inferred from words with similar orthography. He calculated rhyme probability given final character trigrams, which helped achieve higher recall. Although other methods may have better precision, collocation-driven approach wins in recall as seen in 1.1. This method does not take into account meter.

For evaluation, they used precision, recall and F-score calculated as follows:

$$PRECISION = \frac{true\ positives}{true\ positives + false\ positives}$$

$$RECALL = \frac{true\ positives}{true\ positives + false\ negatives}$$

[7] http://versologie.cz/talks/2017basel/

[8] https://github.com/versotym/corpusCzechVerse

$$F - SCORE = \frac{2 * PRECISION * RECALL}{PRECISION + RECALL}$$

For an intuitive view, the reader can imagine precision as how many of algorithm's rhymes were actually rhymes and recall as how many rhymes were discovered. F-score describes the trade-off between the two.

Figure 1.1: Evaluation of RhymeTagger on English corpus in comparison with EM algorithm and simple rule-based approach. The x axis is the year when the tested poem was written, the y axis are the evaluation scores as described above. Reproduced from Plecháč [2017].

Popisat preco nejdu ich data pouzit na testovanie mojej detekcie rymov

Deep-speare

As a part of their sonnet quatrain generating model, Lau et al. [2018] have implemented a Rhyme component that identifies and generates rhymes. It is a unidirectional forward LSTM (Hochreiter and Schmidhuber [1997]) that learns to separate rhyming word pairs from non-rhyming. They generate input by pairing one line-final word with the other three from the same quatrain. Since the rhyme scheme of a sonnet quatrain is always *abab* this will result in one rhyming pair and two non-rhyming. Additional non-rhyming pairs are generated with random word sampling. Then the model with margin-based loss learns the margin

separating the best pair from all the others. It returns a cosine similarity score that estimates how well do two words rhyme.

To evaluate this model, authors used phoneme matching with CMUdict [9] and the EM model from Reddy and Knight [2011] trained on their own data and they were able to outperform both based on F1 score.

1.3 Visualization tools

In the following section, we will describe existing visualization tools for poetry. Software mentioned below focuses on poems, however song lyrics can be considered just a more structurally relaxed version of a regular poem.

1.3.1 Poem Viewer

Quite complex and comprehensive visualization tool is Poem Viewer Abdul-Rahman et al. [2013]. With no need for complicated installations it is easily available for the writers as a web-based application as shown in Figure 1.2. Unfortunately, at the time of writing this thesis the upload of custom text was not working. Luckily, this is still an ongoing project so this might be just a temporary issue. Nevertheless there are some default poems available to demonstrate this software's capabilities.

Figure 1.2: Screenshot from Poem Viewer tool – visualizing Love by Elizabeth Barrett Browning.

Most of the analyzed features (shown in Figure 1.3 focus on the phonetic aspects of the poem. After phonetic transcription to IPA users can analyze consonant features, vowel length and position, stress, syllables, word classes and sentiment using color codes and markers. A second layout offers six different graphs/animations of tongue positions during each verse. Arcs are used to mark

[9] http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Figure 1.3: Available options and their default mappings in Poem Viewer.

end rhyme, alliteration, assonance, consonance, their particular frequencies and repeating words.

Overall this software, although very elaborate, appears crammed and confusing for an inexperienced user. Moreover, it is perhaps better suited for its original use case – a well-structured poem – than less regular song lyrics.

1.3.2 SPARSAR

SPARSAR (Delmonte and Prati [2014]) is also a very interesting tool for poetry analysis and expressive Text-to-speech conversion. It is originally designed for a thorough examination of a very strictly structured Shakespeare's sonnets. To achieve this, it has to run analyses on many levels – and these results can be used to analyze any poem. It looks at the poem on three levels: phonetic (pronunciation, consonant and vowel tongue position, assonances, etc.), poetic (metrical structure, rhyme schemes, acoustic length, etc.), and semantic (sentiment, methaphorically linked words, anaphora, etc.).

User can choose between a window application with graphs and diagrams or a headless mode with .xml output files. Its main disadvantage for our use case is that it is written in Prolog and therefore is very strict on the input format and runs only under a specific older version of Ubuntu.

1.3.3 ProseVis

This Java desktop visualization tool by Clement et al. [2013] analyzes text through parts-of-speech, phonemes, stress, tone, and break index. These features are extracted using OpenMary Text-to-speech system (Schröder et al. [2006]) and predictive classification. The authors believe their visualization will present the features to user in a more human readable form (ProseVis [2014]).

Figure 1.4: Comparison of two poems in ProseVis. Reproduced from ProseVis [2014].

1.3.4 Poemage and RhymeDesign

Poemage (McCurdy et al. [2015a]) and RhymeDesign (McCurdy et al. [2015b]) are both open-source applications with focus on analysis of sonic devices and sonic topology in poetry. Poemage[10] focuses on complex structures of words connected through some sonic or linguistic resemblance across the space of the poem. It is available for MacOS or Windows with a web version currently under development. In MacOS application RhymeDesign – which also provides the backend for Poemage – users can enter their poem and query for one of the default rhyme types or choose a custom rhyme type.

Figure 1.5: An example analysis in Poemage.

[10] http://www.sci.utah.edu/~nmccurdy/Poemage/

1.3.5 Ambiances

This software is unique in the fact that the analysis is integrated in the process of writing. As described in the paper Meneses and Furuta [2015], writers enter the poem, receive a visualization and can control this visualization with body and hand gestures which in turn influence the poem. By such interconnection the authors aim to make Ambiances a part of the writing process and give it a chance to influence the final result. However, the actual software does not seem to be openly available.

1.4 Generation tools

Generation of text, especially artistic, is a very challenging task for a machine. As we observed the outputs of exiting tools, we found that the most common flaws include lack of creativity, unnatural and frequent change of the subject, and incoherence. Generation of song lyrics is basically a more specified branch of text generation. Similarly, it can be distinguished into two main methods – rule-based and machine learning.

Rule-based models are inherently very complex and the output is often limited to certain structure or topic. They usually require the user to input starting configuration, whether it is genre, topic, time period, amount and type of rhymes, phrases, etc. They tend to be less creative but better at rhyming and following the form and structure. To achieve that, they may use rhyming dictionaries (see section 1.2). Simpler ones just use a large set of pre-written templates, e.g. MasterPiece Generator[11]. For this thesis, we will focus on generation using artificial intelligence (AI).

1.4.1 Generating using AI

There are many algorithms in AI that can be used for lyrics generation and

Deepspeare

EM algorithm by Reddy and Knight

GPT-2

Generative pre-trained Transformer version 2 (GPT-2) is an ... It was created by AI-based research laboratory named OpenAI.

This model was trained on (and evaluated against) WebText, a dataset consisting of the text contents of 45 million links posted by users of the 'Reddit' social network. WebText is made of data derived from outbound links from Reddit and does not consist of data taken directly from Reddit itself. Before generating the dataset we used a blocklist to ensure we didn't sample from a variety of subreddits which contain sexually explicit or otherwise offensive content.

To get a sense of the data that went into GPT-2, we've published a list of the top 1,000 domains present in WebText and their frequency. The top 15 domains by volume in WebText are: Google, Archive, Blogspot, GitHub, NYTimes,

[11] https://www.song-lyrics-generator.org.uk/

Wordpress, Washington Post, Wikia, BBC, The Guardian, eBay, Pastebin, CNN, Yahoo!, and the Huffington Post. 1.5 billion parameters: the fourth and largest GPT-2 version. We have also released 124 million, 355 million, and 774 million parameter models. **?**

GPT-3

Originally, this thesis intended to focus more on lyrics generation. However, as our research of works in this field shows, there is an abundance of tools for purpose. Users can just choose a tool that fits their needs. Creating one tool that would be better or more versatile would either require more computing power or literary knowledge, and is above the scope of a master thesis. Therefore, we shifted our main focus to automatic rhyme detection and evaluation, which seems to be far less explored and more interesting topic.

Is this a good argumentation or should I leave it out?

Lau et al. [2018] - learns rhyme automatically, for sonnets, results indistinguishable, apart from expert evaluation, missing emotion

generation by Reddy

- deepspeare - ako oni generuju - generovanie pomocou slovnikov - gpt2, gtp3 Brown et al. [2020] - pravidlove generatory - tym sa nebudeme venovat, spomenut priklad

2. Data

A crucial part of every analysis are the data. To be able to conduct an analysis with results that can reasonably represent the domain, we need to have enough of them - the more the better.

Our dataset consist of 658,460 song lyrics scraped from the crowd-sourced website Genius[1]. Sadly, the original author of the dataset is unknown, it has been passed on to us by a colleague as a potentially interesting source for research. However, all song lyrics are publicly available on the Genius website.

Not sure how to articulate this not to sound that bad...

2.1 Preprocessing

In most areas it is very hard to find a dataset of good quality and large quantity. Usually at least one of the two suffers. It is not any different with our data - although the dataset is large, the contents were created by ordinary people and intended for human readers so they are not well suited for automated processing. It is necessary to look closely at the data, remove faulty or redundant items, and clean the rest with preprocessing.

To assess what are the problems in the data and how to address them, we created a very small dataset of only about 10 songs which we cleaned manually. To select these songs, we looked at about 100 random songs and chose the ones that contained the most common faults. We also tried to contain a broad spectrum of errors by focusing on the diversity in the selected dataset. We then iteratively implemented an automated solution for each type of data corruption, comparing the automatically and manually cleaned data, until they matched. We also extracted statistical information that further showed the weak points that needed addressing.

We received the dataset in JSON format, with each song as a separate item, each containing following features:

- *title* - the name of the song
- *lyrics* - the text of the song's lyrics
- *album* - song's album (or null)
- *genre* - one of the following: rap, pop, rock, r-b, country
- *artist* - song's performer
- *url* - the url of the lyrics page on Genius website
- *year* - the year the song was produced
- *is_music* - boolean flag distinguishing song lyrics from other texts

[1] https://genius.com/

- other song details: *producer, featured artist, recording location, charts, writer, samples, sampled in, has featured video, has featured annotation*
- other website specific information: *rg artist id, rg type, rg tag id, rg song id, rg album id, rg created, has verified callout*

The features from the last two points were *null* for all or most of the items. That, and the fact that they do not give us much more information that would contribute to the lyrics analysis, made them useless and so we decided not to keep them. We removed all songs for which the attribute *is_music* was False, indicating it was not a song (often a poem or prose) or they did not contain lyrics - 34,259 songs in total. We further removed one song with invalid (incomplete) JSON. After comparing the lyrics to each other, we found and removed 32,551 duplicates.

Upon further inspection, we found out that our dataset also contains lyrics in different languages. We used the neural network model for language identification by Google (CLD3)[2] for the classification. It showed that our dataset contains exactly 100 different languages, most of them represented only marginally. Since other languages did not have enough data to support a good analysis and implementing them would be above the scope of this thesis, we kept only English lyrics. We further removed 832 items with language detection errors. All of them were under 10 lines long so they would not be a valuable addition anyway. At this point our dataset contained 438,037 items.

Figure 2.1: Histogram of number of characters in songs of our dataset.

Figure 2.2: Histogram of number of words in songs of our dataset.

To learn more about our data, we created histograms with song's length in characters, words, and lines (see Figures 2.1, 2.2, 2.3). Knowing the common issues of extreme values, we manually examined a few of the shortest and a few of the longest items. Confirming our expectations, we found out that they were not valid songs either. The long ones were usually book excerpts or rap improvisation battles, while the short ones were often links to advertisements or motivational quotes. We removed 14 long and 1,838 short lyrics. Although it may not seem as much, it could have strong negative influence mainly during the generation phase. In the Figures 2.1, 2.2, and 2.3, red lines mark the borders for removal - only the items in between them were kept.

[2] https://github.com/google/cld3

Figure 2.3: Histogram of number of lines in songs of our dataset.

Should I move the boundaries to remove more extremes?

2.2 Structure of the data

This section gives statistical information about the dataset after preprocessing. Table 2.1 sums up basic statistics about the data overall and for each genre specifically. Pie chart in Figure 2.4 shows the portions of the data belonging to each genre. All the attributes are listed in Table 2.2. For some songs, not all attributes are available. Number of items for with non-empty values is given as well.

Genre	Songs	Avg. lines per song	Avg. words per line
Pop	293,679	36.73	5.50
Rap	99,189	64.32	6.88
Rock	34,372	38.73	5.30
R&B	5,126	52.82	5.51
Country	3,819	38.43	5.85
Total	436,185	43.36	5.96

Table 2.1: Basic statistics about the dataset.

Add maybe something about gold data from reddy

Figure 2.4: Distribution of genres in the dataset.

Attribute	Non-empty values
lyrics	436,185
title	436,179
album	112,060
genre	436,185
artist	436,184
url	436,185
year	96,491
lang	436,185
id	436,185
word_count	436,185

Table 2.2: Attributes and their counts of non-empty values.

3. Lyrics evaluation - nechat si nazov pre kapitolu evaluation

sucast zadania ze to bude bez zvukovej stranky hoci to vplyv ma

pri popise taggeru dat typy rymov, ktore nezvlada

3.1 Rhyme detection

Describe the original attempt with SPARSAR.- uviest priklady pesnicky, ktora sa zjavne rymuje ale urcilo ju to zle; priklad slova na ktorom sparsar spadne

[1] - points out most works focus on rhyme schemes in well structured poetry instead of common rhymes in text - where to draw the line between intended rhyme and accidental word similarity

3.1.1 Pronunciation

spomenut ze nevieme zistit ci to spevak nevyslovi inak ,ze moze hltat bezprizvucne slabiky viac naraz

Unlike many other languages, English does not have a straightforward pronunciation rules. Therefore to be able to assess rhymes, we need to transcribe our text into a phonetic alphabet first. There are two commonly used alphabets to choose from – IPA and ARPAbet. The original International Phonetic Alphabet (IPA) used since 1888 uses one UNICODE character to encode each phoneme and it is commonly used for example in dictionaries. Since it uses non-ASCII characters, ARPAbet was developed as an equivalent for computers. It has two versions: 1-character that uses upper-case and lower-case letters and 2-character version where each phoneme is represented by one or more upper-case ASCII characters (Lea [1980])(see Table3.1 for comparison). We will be using the 2-character ARPAbet because it is used by the CMUdict.

Example word	IPA	1-character ARPAbet	2-character ARPAbet
st**o**ry	ɔ	c	AO
bu**tt**er	ɾ	F	DX

Table 3.1: Comparison of different pronunciation alphabets.

Carnegie Mellon University Pronouncing Dictionary (CMUdict) is an open-source pronunciation dictionary.[2] Currently it contains 134,373 words (including their inflections) and their pronunciations in 2-character ARPAbet. For each word, there is one or several possible pronunciations in North American English

[1] http://phylonetworks.blogspot.com/2020/07/automated-detection-of-rhymes-in-texts.html

[2] http://www.speech.cs.cmu.edu/cgi-bin/cmudict

including stress markers for primary, secondary or no stress. For the implementation, we used its Python wrapper package *cmudict* [3]. To use this we need to strip the input of punctuation and convert it to lower case.

CMUdict is a large dictionary and it includes also slang words so it should cover most of our input. To test this, we looked at all last words on each line of our data (since those are the important ones for rhyme analysis) and we found out that 5.52% of them are not in CMU dictionary. These included:

- uncommon words, e.g. superglue, redundantly
- misspelled words, e.g. decsion, girlfren
- numbers
- foreign words, e.g. revoluccion, ecolli
- interjections and onomatopoeia, e.g. shoooshooo, woahwoah

Also, we applied some further data preprocessing that ensured more words in data would be found in the dictionary. We replaced the closing quotation mark "’" with the typewriter apostrophe "'" since only the second variant of apostrophe is accepted by CMUdict. We replaced hyphen "-" with a space " " to separate the hyphen-connected compound words into individual component words that have a higher chance of being found in the dictionary than the full version.

Describe here how we dealt with the ones not in CMUdict. Open MaryTTS?

3.1.2 Syllabification

Once we have the pronunciations, we can start to compare them. When comparing lines for rhymes, we have to establish a system of alignment so that we analyze only relevant pairs of phonemes. Initially, we created a simple rhyme detector that just traversed both verses backwards phoneme by phoneme and compared them. However, rhyming words do not have to have an equal number of phonemes. For example words in the Table 3.2 have a 2-syllable rhyme. However if we compared each phonemes one by one they get misaligned on consonant clusters S-T-R and P-L and we will miss the second syllable rhyme.

Word	ARPAbet transcription
constrain	K AH N - S T R EY N
complain	K AH M - P L EY N

Table 3.2: Example of misalignment when aligning by phonemes.

We need to make sure that we are comparing corresponding parts of verses otherwise we will miss the rhyme. A better approach would be to compare corresponding syllables. Each syllable can be further split into 3 groups ("CVC") – leading consonant cluster, vowel, and trailing consonant cluster. Consonant clusters can sometimes be empty. For syllabification we used python library *syllabify* [4] which conveniently returns syllables in CVC triplets as described above.

[3] https://pypi.org/project/cmudict/
[4] https://github.com/kylebgorman/syllabify

3.1.3 Calculating rating for one rhyme

Finally, we have extracted pronunciation and syllables so we can continue to analyze the rhyme and rate the song. We decided to first calculate ratings for pairs of verses and then create an overall song rating based on these individual ratings. Another approach would be to analyze the complex statistics of the entire song and rate it at the end all at once. The second method could be better at incorporating the high-level properties like repeating of the refrain. We chose the first approach because it is more straight-forward and gives us a number for each rhyme which can be more interesting for the writer. Additional high-level analysis can be added later if necessary.

So let's focus on the rhyme analysis of two verses – or rhyme fellows – as they are typically called. Rhymes are located at the end of each line so there is no need to analyze the entire verse. How far should we look? The first choice would be to look at the last word. However rhymes can extend over more words as we see in

Find an example of multi-word rhyme where the second word is unaccented

. When we look at the rhyme types, the basic ones do not go further then the first stressed syllable (looking at the line backwards). Notably, even if the rhyme does extend further we can ignore the rest because it will not contribute to the rating. According to our research, the most perfect rhyme is perfect rhyme so it should get the perfect score. And if there are more rhyming syllables preceding the perfect rhyme, they cannot make the score better. Similarly, if the rhyme is not perfect, syllables preceding the final stress would already be considered an internal rhyme – which is also used (mainly in rap lyrics) but less valued than the classical end rhyme. We will therefore limit our window to the minimum number of syllables needed to include the stressed syllable in both rhyme fellows. Having a sequence of four unstressed syllables is very unlikely in English language so we limited our word preprocessing (pronunciation + syllabification) to last 4 syllables to speed up the performance.

To determine the rhyme we need to assess the match in sound of individual phonemes. Since we have each syllable separated into three groups we decided to give each group a number between -1 and 1 that would represent how similarly they sound. The numbers are assigned according to following heuristic:

Rewrite this list in a better-readable manner.

- 1 – all phonemes are identical
- 0.75
  - one is a subset of another
  - it is a pair of similar sounding phonemes
- A number between 0 and 1 if both are consonant clusters. This number is calculated:
  1. Add 1/(length of the shorter consonant cluster) for each **shared** sound at the beginning or the end of the consonant clusters.

2. Add 0.75/(length of the shorter consonant cluster) for each **similar** sound at the beginning or the end of the consonant clusters.

- 0 – both are empty
- -0.5 – one of them is empty
- -1 – otherwise, meaning no matching or similar sounds

For an example of similarity evaluation see Table 3.3.

1st verse	on	her	front	door
2nd verse	the	pain	no	more
1st pronunciation	_, AA, N	HH, ER, _	F R, AH, N T	D, AO, R
2nd pronunciation	DH, AH, _	P, EY, N	N, OW, _	M, AO, R
similarity	-0.5, 0.75, -0.5	-1, -1, -0.5	-1, -1, -0.5	-1, 1, 1

Table 3.3: Example similarity evaluation for the last 4 syllables of two verses from song Cheatin' Woman by Lynyrd Skynyrd.

To identify similar sounds, we look them up if there is a similarity group containing both of them. These similarity groups were created...

Describe how it was done. The iterative approach? Or maybe Holtman's hierarchy? panPhon? Mortensen et al. [2016]

Words for which we have not found pronunciation cannot be further processed so they are given rhyme rating 0 and skipped. Some words may have multiple possible pronunciations – in that case we evaluate each possible combination of pronunciations for given line pair. After we assign a rating for each combination, we will keep only the best rated combination of pronunciations and discard the rest.

With these similarity values we can proceed to calculate the rating which we decided to be as typically between 0 and 1. We will look at it syllable by syllable a return an average. There are some cases we need to consider individually:

- different – if all similarities are equal to -1, we can definitely say they do not rhyme and return a rating of 0
- identical – since identity is a weaker rhyme than perfect, it will be given a penalty for "little creativity" returning a fixed rating of 0.8
- perfect – perfect rhyme has a specific structure and if that holds, we can return the perfect score of 1.0

For the remaining cases we will create rules based on how rhymes behave. Not all phonemes are equally important so let's assign weights to reflect it. The key role in rhyming plays the vowel so it should have the strongest impact on the rating. Second important is the ending consonant because it is closer to the end. Beginning consonant can add up to a nicer rhyme but it cannot bear the rhyme on its own. Since the vowel itself can be enough to create the rhyming effect it

should have more weight than the rest combined. Therefore we assigned weights as follows:

Beginning consonants: 1, Vowel: 4, Ending consonants: 2

The rating for one syllable is created as normalized sum of weights times similarities. Furthermore, we need to account for stress. We can do that by multiplying the result with a multiplication factor depending on how does the stress match.

- 1.0 for stressed rhyme because it is the strongest
- 0.9 for unstressed rhyme – it is weaker but the stress pattern matches
- 0.8 for an mismatching stress pattern

The final formula for a rhyme rating is an average of syllable ratings and looks like this:

$$average(stress_multiplication_factor * weighted_average(similarities))$$

This formula formatting looks weird but I don't have an idea how to do it better.

3.1.4 Calculating song rating

rating celej piesne je priemer 2. stlpca, okrem '-'

- rozlišovat terminologicky: - (rhyme) rating, tedy rating dvou veršů=řádků, číslo mezi 0 a 1 - assigned (rhyme) rating, tedy rating, který nakonec danému řádku přidělíte(zatím jako nejvyšší možný z předchozích 1-5 řádků; případně 0, jsou-li všechny ratingy <0.8; případně žádný, jde-li o první řádku skupiny) - (rhyme) score, skóre celé písně, tedy průměr všech assigned ratings.

The next step is to combine these rhyme ratings into one final rating for the entire song. In the previous section we created a function that returns rhyme rating for given two verses. To search for rhymes in the full lyrics we need to decide which verse pairs to check. The most straight-forward approach would be "brute force" – try each line with all the other lines. Besides its obvious disadvantage of increased time requirements it also detects rhymes that span across tens of lines. It is not strictly defined how many lines apart can the rhyme fellows be to still be considered a rhyme – the author can even make it a part of his artistic expression e.g. in "Author's Prologue" by Thomas [1952] the 1st line rhymes with 102th, 2nd with 101th and so on. But realistically, a rhyme between a line at the beginning of the lyrics and 20 lines later would not have an effect on the song listener – it requires a close proximity of rhyme fellows within the poem. Since the most common stanzaic form in English is a quatrain, a stanza of four lines (Eastman et al. [1970]), we decided to set the distance to 3. Further than that the effect on the reader gets weaker really fast.

This probably needs a citation but I can't find any.

As we decided, for each line, we will look at 3 preceding lines to look for its rhyme fellow. In case multiple of them rhyme, the one with the smallest distance will be selected – if it is closer it is more probable the reader will associate it with the rhyme in the next line because it is the strongest in the memory. Consequently, a rating representing the score of this rhyming pair will be saved for the second rhyme fellow. It will not be saved for the first to ensure it will not be added to the final rating twice. The first of the pair will either keep a rating it shares with another line before or it will get an "X" (as seen in Table 3.4) that represent no rating.

Let's focus now on the first four line marked with rhyme scheme letter "e" in Table 3.4. They all rhyme with each other and all rhymes except for *are/dark* are perfect. That means the line with "dark" would receive less than perfect score and it would lower the score of the entire song. If we instead ignore this rhyme and associate only the first two and the last two rhymes together we will receive a perfect score. Loosing rating by marking weaker rhymes does not make sense so we must add an exception to only keep the better score.

Using the ratings added to the lines a final score will be calculated as the average of lines that have a rating (not an "X").

Counterexample - we have a 4-line song where 2 lines rhyme and a 24-line song where 2 lines rhyme - they will receive the same score. That does not seem fair?

Rhymes in songs or poems are typically marked using a rhyme scheme. That means each verse gets assigned a letter – lines that share the same letter rhyme and those with different letters do not. We also decided to adapt this common notation. In case the song needs more letters than there are in the alphabet we will add another letter and continue alphabetically – aa, ab, ac, ..., ba, bb, bc, ..., ca, etc.

Since meter plays an important role in rhymes, another relevant property to examine is the number of syllables. To count syllables for each line we used Python package *syllabify* [5] which returns syllables using ARPAbet transcription. For words not in CMUdict we used a simple heuristic – since the nucleus of each syllable is most often a vowel (except for syllabic consonants) we counted the number of (groups of) vowels and used it as an estimation for the number of syllables. Although this gives a wrong estimate for words like *rhythm* or *house*it performed quite well when we tested it on a few out-of-dictionary words from our dataset. We found it unnecessary to try to further improve the heuristic since the words that are not in CMUdict are often foreign words that do not follow the standard pronunciation rules of English so any application of these rules would probably be of little help.

Nemalo by tu mozno byt este nieco o tom ako sme to testovali? Resp. netreba to odtestovat nejak systematickejsie nez od oka?

[5] https://github.com/kylebgorman/syllabify

Rhyme	Rating	Lyrics	Pronunciation (last 2 syllables)
a	X	Twinkle twinkle little star	'T AH L', 'S T AA R'
a	1.0	How I wonder what you are	'Y UW ', ' AA R'
b	X	Up above the world so high	'S OW ', 'HH AY '
b	1.0	Like a diamond in the sky	'DH AH ', 'S K AY '
a	1.0	Twinkle twinkle little star	'T AH L', 'S T AA R'
a	1.0	How I wonder what you are	'Y UW ', ' AA R'
c	X	When the blazing sun is gone	' IH Z', 'G AO N'
c	0.8	When he nothing shines upon	' AH ', 'P AA N'
d	X	Then you show your little light	'T AH L', 'L AY T'
d	1.0	Twinkle twinkle all the night	'DH AH ', 'N AY T'
e	X	Twinkle twinkle little star	'T AH L', 'S T AA R'
e	1.0	How I wonder what you are	'Y UW ', ' AA R'
e	X	Then the traveler in the dark	'DH AH ', 'D AA R K'
e	1.0	Thanks you for your tiny spark	'N IY ', 'S P AA R K'
f	X	He could not see which way to go	'T UW ', 'G OW '
f	1.0	If you did not twinkle so	'K AH L', 'S OW '
e	X	Twinkle twinkle little star	'T AH L', 'S T AA R'
e	1.0	How I wonder what you are	'Y UW ', ' AA R'
g	X	In the dark blue sky you keep	'Y UW ', 'K IY P'
g	1.0	Often through my curtains peep	'T AH N Z', 'P IY P'
h	X	For you never shut your eye	'Y AO R', ' AY '
h	1.0	Till the sun is in the sky	'DH AH ', 'S K AY '
i	X	Twinkle twinkle little star	'T AH L', 'S T AA R'
i	1.0	How I wonder what you are	'Y UW ', ' AA R'

Table 3.4: Example of song analysis of the nursery rhyme "Twinkle, Twinkle Little Star" with a final rating of 0.989.

4. Visualization

- predpokladame, ze rym je na konci versa

5. Generation

Writing a lyrics for a song is a task that requires creativity. This is one of a few areas where AI has not become better than humans yet. While it is generally capable of keeping similar structure and topic, it still fails use rhymes and other literary devices.

As described in section 1.4, current state-of-art tool for generation is GPT-3 Brown et al. [2020]. During writing of this thesis, it was introduced to public with only limited access. Realizing the power of this tool, the authors did not want to make it available to broad public, fearing it might be misused with bad intentions. Instead they created a form to sign up for access, reviewed individual requests, granting access only to a small portion of them. We have sent several requests for the API but have not received a reply to this day.

The second-best option was therefore its earlier version GPT-2. It offers pre-trained models of sizes small, medium, large, and XL that can be further fine-tuned to the specific task. Since training models of such sizes has very high memory requirements, we would only be able to fine-tune the small model on available school clusters.

However, our colleagues from UFAL institute at Charles University agreed to let us generate on GPT-2 XL that they are using in their ongoing research project THEaiTRE Rosa et al. [2021]. It is a vanilla model, meaning that it has not been fine-tuned to specific task, but as their research confirms, it adapts very well when given a relevant primer. We do not have resources to fine-tune the large model and we decided not to fine-tune the small either. Having available such a large model it would ... Knowing it has been trained on 8 million web pages, it is highly probable that it has already seen our data (or some similar web containing lyrics data).

A chair is still a chair
In most other aspects, I think the two camps in this debate can agree. We're in good company, for instance, as far as the law is concerned. When it comes to the "should" part of the debate, the fact that some states have no-fault insurance mandates, and others have them, is not a matter of opinion, or of opinion alone.

Figure 5.1: Result of generation (in blue) using 1-line primer (in black).

generovanie - z kazdeho zanru 1 pesnicku - dat z pesnicky 1, 2, 4, 8 riadkov ako primer

Original	2-line primer
A chair is still a chair Even when there's no one sittin' there But a chair is not a house And a house is not a home When there's no one there to hold you tight And no one there you can kiss goodnight Woah girl	A chair is still a chair Even when there's no one sittin' there I still think there are people here A chair is still a chair Even when there's no one sittin' thereI still think there are people here A chair is still a chair
4-line primer	**8-line primer**
A chair is still a chair Even when there's no one sittin' there But a chair is not a house And a house is not a home So why do we make up the terms? It's not just us who insist on the word house, we use it in all sorts of other contexts too. For example, if you want to use a single word to refer to a group of people, you might say: If you want to invite everyone to an event, you might say: It's not hard to see why we seem to love using a term like 'house' to refer to a group of people – it's the easiest word for us to remember.	A chair is still a chair Even when there's no one sittin' there But a chair is not a house And a house is not a home When there's no one there to hold you tight And no one there you can kiss goodnight Woah girl A room is a still a room And a still a room is still a home When there's no one there to hold you tight And no one there you can kiss goodnight I'm a little man on a little island And a lonely little bird on a lonely little island But a lonely little bird is not a castle And a lonely little bird is not a home And a lonely little bird is not a world I'm a little man in a little house And a lonely little bird in a lonely little house

Table 5.1: Beginning of "A House Is Not A Home" lyrics by Luther Vandross - genre R&B. Comparison with results generated using 2,4, and 8-line primers.

Conclusion

do future works pripisat napady z mailov a work diary

zistit preco sa pri webovych citaciach neukazuje "accessed on"

Bibliography

A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C.R. Johnson, A. Trefethen, and M. Chen. Rule-based visual mappings – with a case study on poetry visualization. *Computer Graphics Forum*, 32(3):381–390, 2013. doi: 10.1111/cgf.12125. URL http://dx.doi.org/10.1111/cgf.12125.

Alexander Bain. *English composition and rhetoric, a manual*. New York, Appleton, 1867.

Chris Baldick. *The Oxford Dictionary of Literary Terms*, volume 3. Oxford University PressPrint, 2008. ISBN 9780199208272.

Bennet Bergman. Rhyme. 2017. URL https://www.litcharts.com/literary-devices-and-terms/rhyme.

T. V. F. Brogan. *The Princeton Handbook of Poetic Terms*, volume 3. Princeton University Press, 2016. ISBN 9781400880645.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. 2020.

Tanya Clement, David Tcheng, Loretta Auvil, Boris Capitanu, and Megan Monroe. Sounding for meaning: Using theories of knowledge representation to analyze aural patterns in texts. *DHQ: Digital Humanities Quarterly*, 7(1), 2013.

R. Delmonte and A. M. Prati. SPARSAR: An expressive poetry reader. pages 73–76, apr 2014. doi: 10.3115/v1/E14-2019. URL /https://www.aclweb.org/anthology/E14-2019.

Arthur M Eastman, Alexander Ward Allison, Herbert Barrows, et al. *The Norton Anthology of Poetry*. Norton New York, 1970.

Erica Greene, Tugba Bodrumlu, and Kevin Knight. Automatic analysis of rhythmic poetry with applications to generation and translation. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, pages 524–533, 2010.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Jey Han Lau, Trevor Cohn, Timothy Baldwin, Julian Brooke, and Adam Hammond. Deep-speare: A joint neural model of poetic language, meter and rhyme. *arXiv preprint arXiv:1807.03491*, 2018.

Wayne A Lea. *Trends in speech recognition*. Prentice Hall PTR, 1980.

LiteraryDevices Editors. Rhyme - examples and definition of rhyme as a literary device, Dec 2020. URL https://literarydevices.net/rhyme/.

Nina McCurdy, Julie Lein, Katharine Coles, and Miriah Meyer. Poemage: Visualizing the sonic topology of a poem. *IEEE transactions on visualization and computer graphics*, 22(1):439–448, 2015a.

Nina McCurdy, Vivek Srikumar, and Miriah Meyer. Rhymedesign: A tool for analyzing sonic devices in poetry. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 12–22, 2015b.

Luis Meneses and Richard Furuta. Visualizing poetry: Tools for critical analysis. *paj: The Journal of the Initiative for Digital Humanities, Media, and Culture*, 3, 2015.

David R. Mortensen, Patrick Littell, Akash Bharadwaj, Kartik Goyal, Chris Dyer, and Lori S. Levin. Panphon: A resource for mapping IPA segments to articulatory feature vectors. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3475–3484. ACL, 2016.

Petr Plecháč. A collocation-driven method of discovering rhymes (in czech, english, and french poetry). In *Taming the Corpus*, pages 79–95. Springer, 2018.

Petr Plecháč. Collocation-driven method of discovering rhymes in a corpus of czech, english, and french poetic texts. 2017. URL http://versologie.cz/talks/2017basel/.

ProseVis. Prosevis. 2014. URL https://sourceforge.net/projects/prosevis/.

Sravana Reddy and Kevin Knight. Unsupervised discovery of rhyme schemes. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 77–82, 2011.

Rudolf Rosa, Tomáš Musil, Ondřej Dušek, Dominik Jurko, Patrícia Schmidtová, David Mareček, Ondřej Bojar, Tom Kocmi, Daniel Hrbek, David Košt'ák, et al. Theaitre 1.0: Interactive generation of theatre play scripts. *arXiv preprint arXiv:2102.08892*, 2021.

Marc Schröder, Anna Hunecke, and Sacha Krstulovic. Openmary–open source unit selection as the basis for research on expressive synthesis. In *Blizzard Workshop*, 2006.

The Editors of Encyclopaedia Britannica. Encyclopedia britannica. 2014. URL https://www.britannica.com/.

Dylan Thomas. Author's prologue. *Collected Poems 1934-1952*, 1952.

Jona van der Schelde. Phonological and phonetic similarity as underlying principles of imperfect rhyme. 2020.

Elizabeth Walter. *Cambridge advanced learner's dictionary*. Cambridge university press, 2008.

Viktor Zhirmunsky and John Hoffmann. Introduction to rhyme: Its" history and theory". *Chicago Review*, 57(3/4):121–128, 2013.

List of Figures

1.1 RhymeTagger evaluation 8
1.2 Screenshot from Poem Viewer tool – visualizing Love by Elizabeth Barrett Browning. 9
1.3 Available options and their default mappings in Poem Viewer. . . 10
1.4 Comparison of two poems in ProseVis 11
1.5 An example analysis in Poemage. 11

2.1 Histogram of number of characters in songs of our dataset. 15
2.2 Histogram of number of words in songs of our dataset. 15
2.3 Histogram of number of lines in songs of our dataset. 16
2.4 Distribution of genres in the dataset. 17

5.1 Result of generation (in blue) using 1-line primer (in black). . . . 26

List of Tables

2.1 Basic statistics about the dataset. 16
2.2 Attributes and their counts of non-empty values. 17

3.1 Comparison of different pronunciation alphabets. 18
3.2 Example of misalignment when aligning by phonemes. 19
3.3 Example similarity evaluation for the last 4 syllables of two verses from song Cheatin' Woman by Lynyrd Skynyrd. 21
3.4 Example of song analysis of the nursery rhyme "Twinkle, Twinkle Little Star" with a final rating of 0.989. 24

5.1 Beginning of "A House Is Not A Home" lyrics by Luther Vandross - genre R&B. Comparison with results generated using 2,4, and 8-line primers. 27

A.1 Consonant phonemes - transcription between IPA and ARPAbet. 35
A.2 Vowel phonemes - transcription between IPA and ARPAbet. . . . 36

Glossary of literary and technical terms

consonant cluster A sequence of syllables without a vowel. 6

headless mode A mode in which software runs on hardware without a graphic user interface, e.g. a script in terminal. 9

LSTM Long-Sort Term Memory - a type of recurrent neural network. 7

quatrain A type of stanza consisting of four lines. 7

rime riche A rhyme produced by agreement in sound not only of the last accented vowel and any succeeding sounds but also of the consonant preceding this rhyming vowel. 3

sonnet A poetic form traditionally containing 14 lines written in iambic pentameter with rhyme scheme *abab cdcd efef gg*. 7

syllable peak A nucleus of a syllable - either a vowel or a syllabic consonant. 6

A. Attachments

A.1 IPA and ARPAbet transcription table

Following tables show the transcription between IPA and ARPAbet for consonants (Figure A.1) and vowels (Figure A.2). The ARPAbet phoneme set used by CMUdict is shown, as described on their website[1]. Note, that IPA diphthongs are not transcribed separately but as one two-character ARPAbet symbol.

ARPAbet	IPA	Example
B	b	be
CH	tʃ	cheese
D	d	dee
DH	ð	thee
F	f	fee
G	g	green
HH	h	he
JH	dʒ	gee
K	k	key
L	l	lee
M	m	me
N	n	knee
NG	ŋ	ping
P	p	pee
R	r	read
S	s	sea
SH	ʃ	she
T	t	tea
TH	θ	theta
V	v	vee
W	w	we
Y	j	yield
Z	z	zee
ZH	ʒ	seizure

Table A.1: Consonant phonemes - transcription between IPA and ARPAbet.

[1] http://www.speech.cs.cmu.edu/cgi-bin/cmudict

ARPAbet	IPA	Example
AA	ɑ	odd
AE	æ	at
AH	ʌ	hut
AO	ɔ	ought
AW	aʊ	cow
AY	aɪ	hide
EH	ɛ	Ed
ER	ɜr	hurt
EY	eɪ	ate
IH	ɪ	it
IY	i	eat
OW	oʊ	oat
OY	ɔɪ	toy
UH	ʊ	hood
UW	u	two

Table A.2: Vowel phonemes - transcription between IPA and ARPAbet.