



**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	10
1.4 Generation tools	12
1.4.1 Rule-based generating	12
1.4.2 Generating using AI	13
2 Data	15
2.1 Preprocessing	15
2.2 Structure of the data	17
3 Rhyme detection	19
3.1 Using available tools for detection	19
3.2 Defining the requirements	20
3.3 Pronunciation	21
3.4 Syllabification	22
3.5 Calculating rating for one rhyme	22
3.6 Calculating song rating	24
4 Evaluation	26
5 Visualization	27
6 Generation	28
Conclusion	30
Bibliography	31
List of Figures	34
List of Tables	35
Glossary of literary and technical terms	36
A Attachments	37
A.1 IPA and ARPAbet transcription table	37

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 písni (oproti básním), krom toho, že jako data máte písni. U písni 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 schválne omezujete jen na ty (písňové) texty a ignorujete zvukovou stránku.

na začátku nezapomenout na motivaci a tam vyjmenovat, kdo jsou cíloví uživatelé, mezi nimiž budou (začínající) autoři básní/písní.

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 \[2017\]](#), [Wikipedia¹](#)). 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 /flai/ – sky /skai/
before /bi.fɔ:r/ – explore /iks.plo:r/ ² ₃

¹<https://en.wikipedia.org/wiki/Rhyme>

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

³Stressed 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 /textipakreɪ.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 /hju: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 /pem.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 /kɑ:r.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 /sɪn/ – scene /sɪ:n/

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

compare /kəm.pər/ – despair /dɪs.pər/.⁴ (Baldick [2008])

It is generally considered not as good as perfect rhyme because it is too predictable for the listener⁴. 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 /gri:n/ – fiend /fi:nd/

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 to 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 /fri:/ – breeze /bri:z/

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 /kən.t्रə/

hickory dickory dock /hɪ.kə.ri dɪ.kə.ri də:k/

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 /pru:v/" ([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 dictionary⁵.

Pronouncing and CMU dictionary

Pronouncing⁶ 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

Enhancing the naive approach with various similarity measures or other features allows us to find more subtle rhymes like *imperfect* or *forced*.

Rhyme Genie

Although this tool is not free, it is the most commercial⁷ and perhaps the most used by general public, so it is worth mentioning. Rhyme Genie⁸ is a desktop application for Windows and MacOS that suggests 30 different rhyme types for a given word. Additionally, it includes sayings, clichés, idioms, and a very unique feature - adjustable rhyme similarity. However, its use case a the reverse of what we are looking for - rhymes are not found, only suggested.

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

⁵<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

⁶<https://pypi.org/project/pronouncing/>

⁷It was included in Grammy Awards gift bag.

⁸<https://www.rhymegenie.com/rhyme-genie.html>

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, assonance, etc.), poetic (metrical structure, rhyme schemes, acoustic length, etc.), and semantic (sentiment, metaphorically 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.

Datamuse

Datamuse API⁹ combines the advantages of rhyming dictionary and semantic analysis. It uses CMUdict for phonetic transcription, analyzes CommonCrawl¹⁰ web data repository for forced rhymes, Google Books Ngrams ([Weiss \[2015\]](#)) for building language model, and WordNet 3.0 ([Pearson et al. \[2005\]](#)) for semantic relations. Users can send complex queries, e.g. “words that rhyme with *grape* that are related to *breakfast*”. Similarly to Rhyme Genie, it focuses more on rhyme suggestion rather than rhyme detection.

1.2.3 Similarity (distance) based approach

Generally, substituting arbitrary consonant in perfect rhyme does not necessarily create forced rhyme – corresponding phonemes must have similar sound. Objective criteria to measure this similarity can be phoneme’s features e.g. plosive, nasal, fricative, voiced, etc. Rhyme detectors can use these features and other specific sound rules to calculate sound distance between two words. We found no specific tool focusing on this approach, but some listed tools like Rhyme Genie incorporated it into their algorithm.

However, not all phonetic features contribute to similar sound the same. Furthermore, speakers from different areas perceive sound differently and may have distinct boundaries for what is similar and what is not. Consequentially, there are no written rules for similarity so other researchers tried to make AI create them.

1.2.4 Machine learning

AI is inherently very good at solving rule-based problems so it is no wonder that the majority of recent tools took this approach. We will describe one tool that uses **LSTM** neural network and two tools that use EM algorithm, with more focus towards Rhyme Tagger, as we will use this tool later for inspiration and comparison with our detector.

⁹<https://www.datamuse.com/api/>

¹⁰<https://commoncrawl.org/>

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.¹¹

RhymeTagger

Plecháč [2018] came with an open-source collocation-driven alternative named RhymeTagger. It uses the same dataset as the previous approach with addition of a larger Czech poetry corpus¹². Each line-final word is transcribed into phonetic transcription and split into two types of 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 əʊ:əʊ will be very high but for consonant component pair k:r 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 during the Shakespearean era they were written in, therefore using current pronunciation dictionaries may ruin the original rhyme, e.g. original near /nɛ:r/ – there /ðɛ:r/ vs. contemporary near /nr:r/ – there /ðɛ:r/. 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 figure 1.1.

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

$$PRECISION = \frac{true\ positives}{true\ positives + false\ positives}$$
$$RECALL = \frac{true\ positives}{true\ positives + false\ negatives}$$
$$F - SCORE = \frac{2 * PRECISION * RECALL}{PRECISION + RECALL}$$

¹¹<http://versologie.cz/talks/2017basel/>

¹²<https://github.com/versotym/corpusCzechVerse>

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

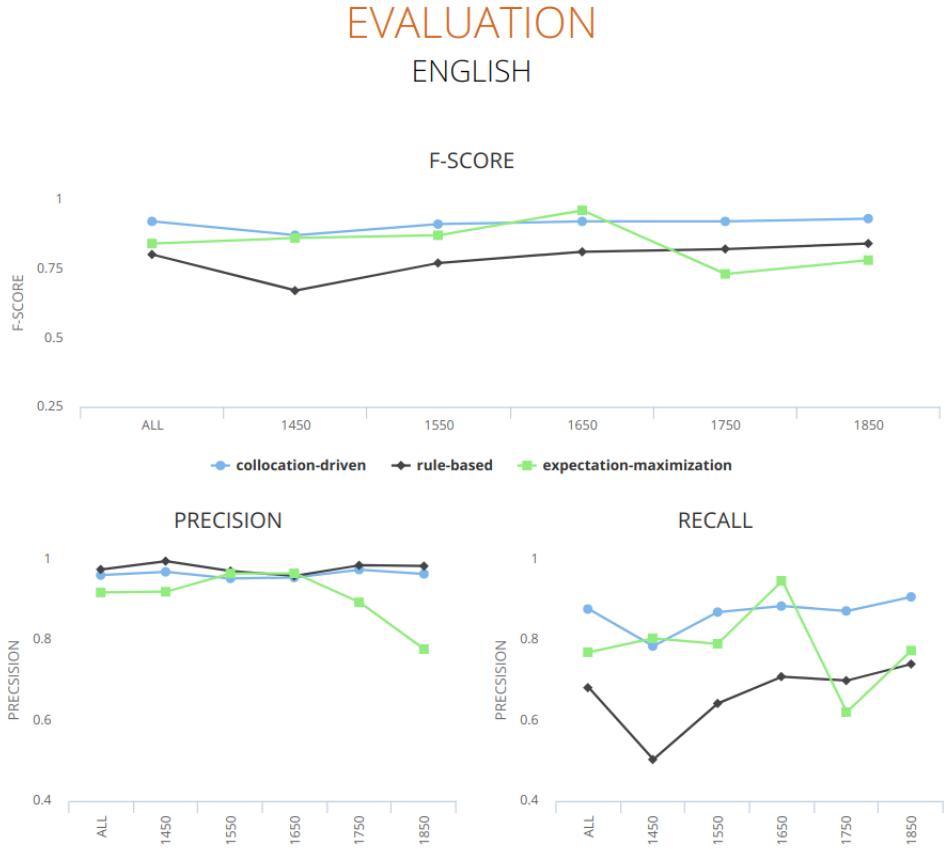


Figure 1.1: Evaluation of RhymeTagger on English corpus in comparison with EM algorithm and simple rule-based approach. The x axis is the year when the tested poem was written, the y axis are the evaluation scores as described above. Reproduced from Plecháč [2017].

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 ¹³

¹³<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

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.

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

Overall this software, although very elaborate, feels overwhelming 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.

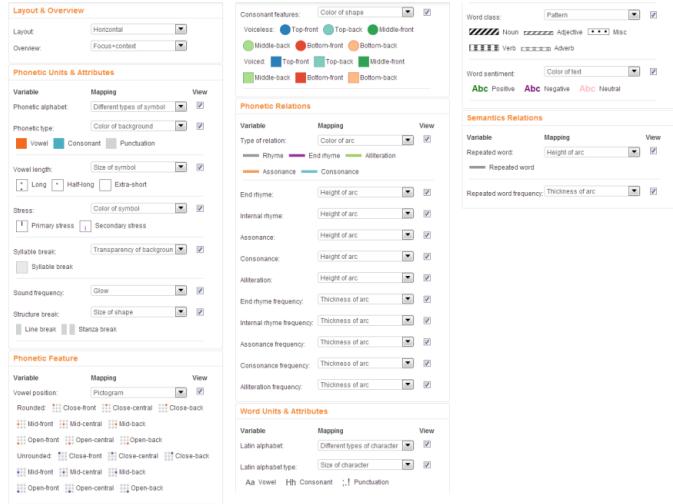


Figure 1.3: Available options and their default mappings in Poem Viewer.

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]).

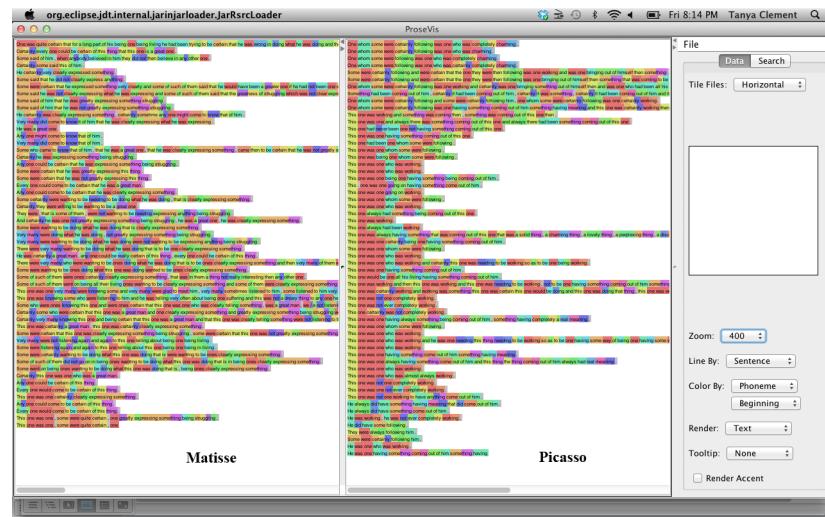


Figure 1.4: Comparison of two poems in ProseVis. Reproduced from ProseVis [2014].

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. Poemage¹⁴ focuses on complex structures of words connected through sonic or linguistic resemblance across the space of the poem. It

¹⁴<http://www.sci.utah.edu/~nmccurdy/Poemage/>

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.

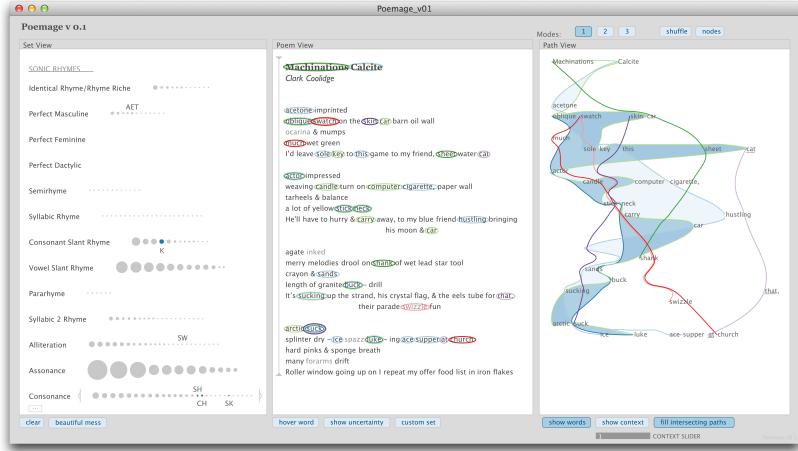


Figure 1.5: An example analysis in Poemage.

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 input 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 publicly available.

1.4 Generation tools

Generation of text, especially artistic, is a very challenging task for a machine. As we observed the outputs of existing 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.

1.4.1 Rule-based generating

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 Generator¹⁵. In this thesis, we will focus on generation using artificial intelligence (AI).

1.4.2 Generating using AI

The go-to approach for generating song lyrics is certainly AI in its many forms. It can be proved by the large number of AI lyrics generators created by enthusiasts, e.g. *These Lyrics Do Not Exist*¹⁶, *Freshbots Lyrics Generator*¹⁷, *Random Lyrics Generator*¹⁸, *DeepBeat - Rap Lyrics Generating AI*¹⁹, *BoredHumans Generator*²⁰, *RapPad Lyrics Generator*²¹, and many more. We will further describe Deep-speare (mentioned earlier in section 1.2) and current state-of-art GPT-2 and GPT-3.

Deep-speare

Deep-speare (Lau et al. [2018]) is a joint neural network architecture that generates only a specific type of poems with strict form and meter – **sonnet quatrains**. It consists of three models: language model generates one word at a time, pentameter model samples meter-conforming sentences, rhyme model enforces rhyme, and they are all trained together in multi-task learning setting. They present very good results – generated poems are mostly indistinguishable from human-written ones, apart from expert evaluation, where they report lack of emotion and worse readability.

GPT-2

Generative pre-trained Transformer version 2 (GPT-2) (Radford et al. [2019]) is an unsupervised **transformer model** capable of various text-processing and generating tasks such as answering questions, translating, summarizing, writing coherent paragraphs, etc. 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 on sites like Google, Blogspot, GitHub, NYTimes, BBC, eBay, etc. It offers 4 models of different sizes increasing in the number of parameters: 124 million (small), 355 million (medium), 774 million (large), and 1.5 billion (XL) parameter models.

Although this model was only trained for the general task of predicting the next word, given all of the previous words within some text, it can be further fine-tuned for a more specific task to suit user’s needs. However, even without fine-tuning, it can quickly adapt to the style of the input and continue in the same manner. One example of lyrics generated by GPT-2 is *Keywords To Lyrics*²².

¹⁵<https://www.song-lyrics-generator.org.uk/>

¹⁶<https://theselyricsdonotexist.com/>

¹⁷<https://www.freshbots.org/lyrics-generator>

¹⁸<http://www.anticulture.net/RandomLyrics.php>

¹⁹<https://deepbeat.org/>

²⁰https://boredhumans.com/lyrics_generator.php

²¹<https://www.rappad.co/songs-about/>

²²<https://lyrics.mathigatti.com/>

GPT-3

During writing of this thesis, an even larger model GPT-3 (Brown et al. [2020]) with 175 billion parameters was officially introduced. It is considered the largest artificial neural network created to date. It was trained on five different corpora: Common Crawl, WebText2, Books1, Books2 and Wikipedia. The architecture is the same as in GPT-2, only number of layers and other parameters increased. It is capable of writing articles indistinguishable from human-written ones, even produce functional javascript code for natural-language formulated task.

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.

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.

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 Genius¹. 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 and can be linked with the corresponding item of the dataset via the *url* attribute.

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
- other song details: *producer*, *featured artist*, *recording location*, *charts*, *writer*, *samples*, *sampled in*, *has featured video*, *has featured annotation*

¹<https://genius.com/>

- 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.

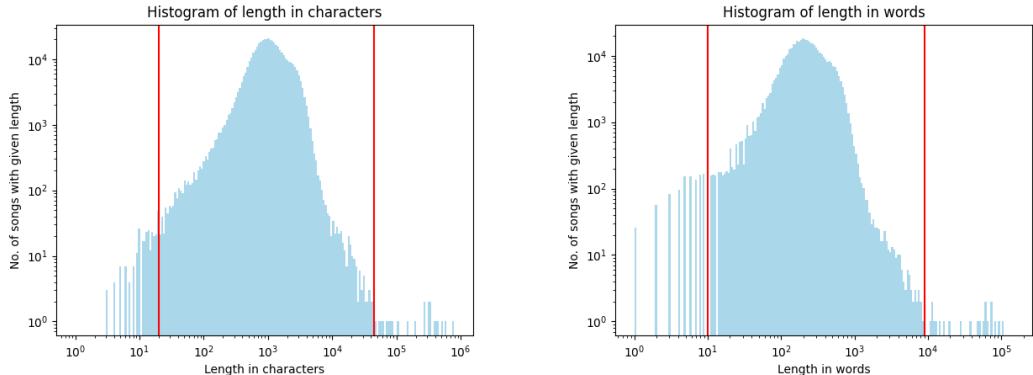


Figure 2.1: Histogram of number of characters in songs of our dataset.

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>

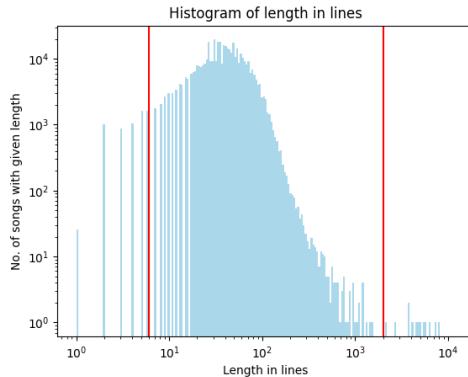


Figure 2.3: Histogram of number of lines in songs of our dataset.

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.

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.

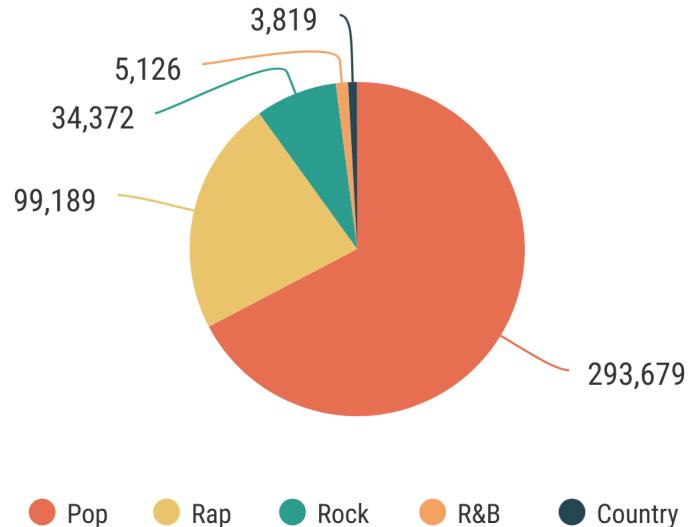


Figure 2.4: Distribution of genres in the dataset.

Add maybe something about gold data from reddy

Add about my annotated data.

3. Rhyme detection

Detecting rhymes may seem like a simple task at first, but looking into details one discovers many problems that need to be addressed. As we have seen in chapter 1, it is not a well-defined task so boundaries need to be set.

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

pri popise taggeru dat typy rymov, ktore nezvlada

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 Using available tools for detection

The simplest approach would be to use one of the tools described in section 1.2. We want a detector that is free, strong (detects more than perfect rhymes), and offers headless mode (we could run it automatically from code). Ruling out unsuitable tools, we are left with Rhyme Tagger and SPARSAR. Rhyme Tagger is easy-to-use but only outputs rhyme scheme. To be able to automatically evaluate the rhyming quality of a song, we would need more information like stress or rhyme type.

SPARSAR

SPARSAR, on the other hand, has a very rich and detailed output. Although it is lacking documentation, the *xml* output format is quite descriptive to understand what most of the values represent. It seemed promising so we attempted to pursue this path.

To bridge the outdated system requirements, we contacted the authors for a newer build (for Ubuntu 19.1). They were very helpful and soon we were able to run it on our computer. Nevertheless, we encountered several issues, mostly stemming from the fact that SPARSAR was written in Prolog. Firstly, the xml output was difficult to parse, as the values were written in Prolog syntax, using the same delimiter for different levels of separation. Secondly, SPARSAR parses the text in sentences but lyrics are usually written without punctuation. We added punctuation using *punctuator2* (Tilk and Alumäe [2016]) which also added space for error.

Lastly, when SPARSAR encountered an unknown word, it failed for the entire song. Since our data was crowd-sourced, it contained many unusual words, so in the beginning it failed on 80% of our data. We iteratively worked with the authors for months on fixing bugs and adding words (mainly contractions, such as I'mma, y'all, yo', 'em, etc.) into their dictionary. In the end, we were able to successfully run it on 95% of our data.

However, we were still not very satisfied with the result. As you can see in the example in table 3.1, it failed to detect the perfect rhyme between 2nd and 4th

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

line, and instead marked a rhyme on line 1 and 3, where there was none. Such errors were not sparse and led us to believe, that the encoded inclination of this tool to look for sonnet-shaped schemes caused it to make errors in the diverse schemes of song lyrics. It may be a great tool for sonnets, but we have found it insufficient for our purpose, so we proceeded to create our own rhyme detector.

Scheme	Line	Last word's pronunciation (as assigned by SPARSAR)
a	Pulled out from the station	s-t-ey-sh-ah-n
h	fifteen after two	t-uw
a	300 miles away from Vegas	v-ey-g-ah-s
b	We had nothin better to do	d-uw

Table 3.1: Example of incorrect scheme assignment by SPARSAR. Excerpt from the song *Good Life*.

3.2 Defining the requirements

Before we dive into algorithm selection and implementation details of our detector, let's define what exactly do we want our detector to do. Additionally, we also establish terms for common cases to keep our further explanations short and clear.

component	a vowel or a consonant cluster of a syllable
rhyming fellows	two lines that rhyme together
rhyming part	the exact components that participate in rhyme (are equal or similar in sound)
rhyme rating	rating of the quality of one rhyme between two lines
song rating	rating of the rhyming quality of the entire song

Input For our input, we expect song lyrics in English, formatted in lines with rhyme always at the end of line. If there will be rhyme in the middle of the line, it will be considered an *internal rhyme* and will not be detected, as we decided in chapter 1 to only focus on end rhymes in this thesis. Optional empty lines between stanzas or chorus will be preserved but skipped.

Output The main element of our output is rhyme scheme. As a single element, it gives the best overview of the song and more importantly, it allows us to compare our detector with others or with the *gold data*. Additionally, we want more information to assess rhyme quality: rhyme rating for each individual rhyme, song rating, rhyme type, pronunciation of the rhyming part, optional modification of stress, etc.

3.3 Pronunciation

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 (the more common) 2-character version where each phoneme is represented by one or more upper-case ASCII characters ([Lea \[1980\]](#))(see Table 3.2 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
story	ɔ	c	AO
butter	r	F	DX

Table 3.2: Short 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 including stress markers for primary, secondary or no stress. For the implementation, we used its Python wrapper package *cmudict*³. To use this we had 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. We manually inspected a small portion of them and found out that they can be mostly split into these 5 categories:

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

Additionally, we replaced the closing quotation mark "" (U+2019) with the typewriter apostrophe '' (U+0027) since only the second variant of apostrophe is accepted by CMUdict. However, this only decreased the percentage of unrecognized words to 5.47%.

²<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

³<https://pypi.org/project/cmudict/>

Clearly, we needed a way to estimate the pronunciation for these words, so we used grapheme-to-phoneme library *g2p* by Park and Kim [2019]. It predicts the pronunciation for out-of-dictionary words using deep learning seq2seq model by TensorFlow (Abadi et al. [2015]).

Even having the pronunciation for every word will not ensure we find every rhyme intended. Song artists may take their liberty in modifying or skewing the pronunciation to make the rhyme work. Sometimes they can also sing two syllables in one beat or use an unusual pronunciation from different culture to convey a message. As we established in the beginning, we will focus only on information retrievable from the text and ignore these possible deviations in pronunciation.

3.4 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 3.3 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. For forced rhymes this inherently becomes a very frequent issue.

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

Table 3.3: 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 (*onset*), vowel (*nucleus*), and trailing consonant cluster (*coda*). Consonant clusters can sometimes be empty. For syllabification we used Python library *syllabify*⁴ which conveniently returns syllables in CVC triplets as described above.

3.5 Calculating rating for one rhyme

Finally, after we have extracted pronunciation and syllables, 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

⁴<https://github.com/kylebgorman/syllabify>

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 than 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.

Mohla byste začít tím, že každému rýmu chcete přiřadit skóre v intervalu 0-1. Pak asi prostě uvést, že (jste se rozhodla, že) perfect rhyme dostane maximální skóre, tedy 1.

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 and as stated in section 3.2, not a target of analysis in this thesis. 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 the output of 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.

...

...

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]

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.

stress penalty

The final formula for a rhyme rating is ..simple multiplication...:

$$\text{average}(\text{stress_multiplication_factor} * \text{weighted_average(similarities)})$$

3.6 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 písničky, 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. Realistically, a rhyme between a line at the beginning of the lyrics and 20 lines later would not have a strong effect on the song listener – it requires a close proximity of rhyme fellows within the poem for the rhyme to be noticed by ear. 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 5.

As we decided, for each line, we will look at 5 preceding lines to look for its rhyme fellow. In case multiple of them rhyme,the one with the smallest distance will be selected .

...

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").

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 some words e.g. *rhythm* or *house*, it performed quite well when we tested it on a few out-of-dictionary words

⁵<https://github.com/kylebgorman/syllabify>

from our dataset. We found it unnecessary to try to further improve the heuristic because 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 možno byť este niečo o tom ako sme to testovali? Resp. ne-
treba to odtestovať nejak systematickejsie nez od oka?

4. Evaluation

Popisat preco nejdu reddy data pouzit na testovanie mojej detekcie rymov

5. Visualization

- predpokladame, ze rym je na konci versa - ze tu su pre uzivatela pekne znazornene dodatocke informacie ako kvalita jednotlivych rymov atd.

6. 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 ?. 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 ...

Finish the sentence

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

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 good-night 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' there I 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 good-night Woah girl
	A room is 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 good-night 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 6.1: 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 preto sa pri webovych citaciach neukazuje "accessed on"

Bibliography

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL <https://www.tensorflow.org/>. Software available from tensorflow.org.

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.
- Kyubyong Park and Jongseok Kim. g2pe, 2019. URL <https://github.com/Kyubyong/g2p>.
- Barbara Zurer Pearson, PA de Villiers, K Brown, and E Lieven. Encyclopedia of language and linguistics, 2005.
- 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/>.
- Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

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 Schmidlová, David Mareček, Ondřej Bojar, Tom Kocmi, Daniel Hrbek, David Košt’ák, et al. Theatitre 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.

Ottokar Tilk and Tanel Alumäe. Bidirectional recurrent neural network with attention mechanism for punctuation restoration. In *Interspeech*, pages 3047–3051, 2016.

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.

Andrew Weiss. Google n-gram viewer. *The Complete Guide to Using Google in Libraries: Instruction, Administration, and Staff Productivity*, 1:183, 2015.

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	9
1.2	Screenshot from Poem Viewer tool – visualizing Love by Elizabeth Barrett Browning.	10
1.3	Available options and their default mappings in Poem Viewer.	11
1.4	Comparison of two poems in ProseVis	11
1.5	An example analysis in Poemage.	12
2.1	Histogram of number of characters in songs of our dataset.	16
2.2	Histogram of number of words in songs of our dataset.	16
2.3	Histogram of number of lines in songs of our dataset.	17
2.4	Distribution of genres in the dataset.	18
6.1	Result of generation (in blue) using 1-line primer (in black).	28

List of Tables

2.1	Basic statistics about the dataset.	17
2.2	Attributes and their counts of non-empty values.	17
3.1	Example of incorrect scheme assignment by SPARSAR. Excerpt from the song <i>Good Life</i>	20
3.2	Short comparison of different pronunciation alphabets.	21
3.3	Example of misalignment when aligning by phonemes.	22
6.1	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.	29
A.1	Consonant phonemes - transcription between IPA and ARPAbet.	37
A.2	Vowel phonemes - transcription between IPA and ARPAbet.	38

Glossary of literary and technical terms

consonant cluster A sequence of syllables without a vowel. 8

gold data In data classification, it is the dataset with correct labels already assigned. It can be used for supervised learning or an evaluation of unsupervised learning.. 20

headless mode A mode in which software runs on hardware without a graphic user interface, e.g. a script in terminal. 7

internal rhyme A rhyme that occurs in the middle of lines of poetry, instead of at the ends of lines.. 20

LSTM Long-Sort Term Memory - a type of recurrent neural network. 7, 9

quatrain A type of stanza consisting of four lines. 9, 13

rhyming part A part of word (or multiple words) that rhymes (is identical or similar in sound) with other word/words.. 4

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, 4

sonnet A poetic form traditionally containing 14 lines written in iambic pentameter with rhyme scheme *abab cdcd efef gg*. 9, 13

syllable peak A nucleus of a syllable - either a vowel or a syllabic consonant. 8

transformer model A deep learning model that adopts the mechanism of attention, differentially weighing the significance of each part of the input data.. 13

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 website¹. 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	a	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.