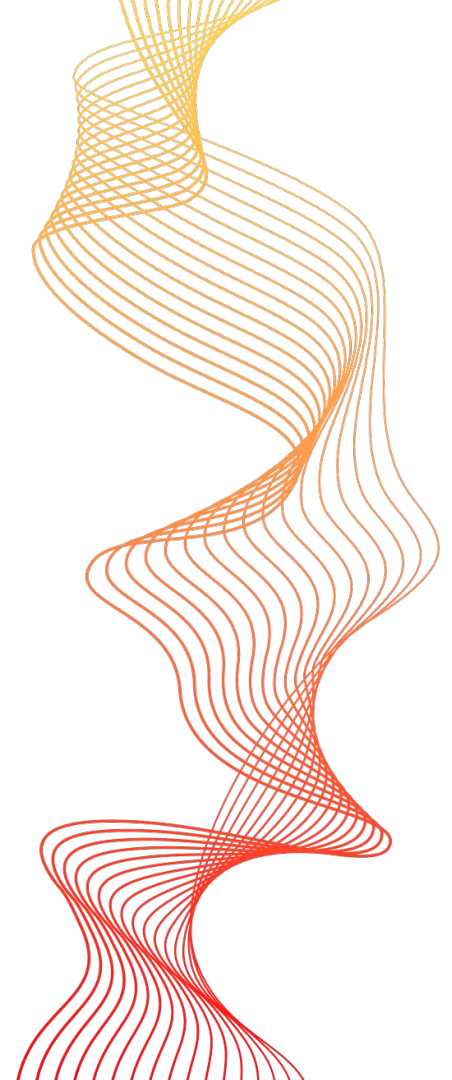




Automated Rhyme Detection in Text

Nguyen Thanh Vinh - 21020673



Introduction

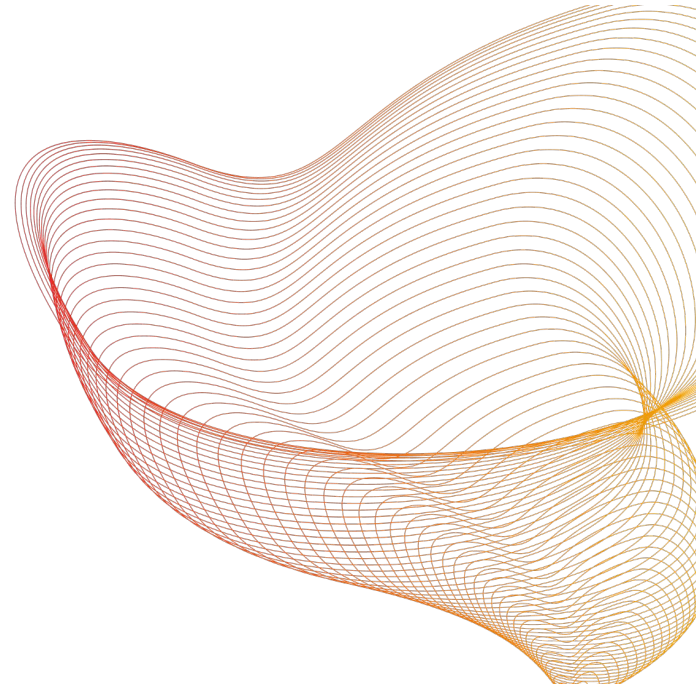


- 01 Rhyme plays a crucial role in various fields such as poem and lyric composition.
- 02 Machines face challenges in identifying rhymes due to the lack of inherent knowledge and context possessed by humans.



What is a rhyme?

- A rhyme is a phonetic similarity between words or phrases, particularly in their ending sound.
- Rhymes are used to create rhythm, musicality, and aesthetic appeal in language.





Types of Rhyme

- Perfect Rhyme: Occurs when the final stressed syllables of two or more words have identical sounds.

Example: cat, hat

- Imperfect Rhyme: Occurs when the final sounds of two or more words are similar, but not identical.

Example: thumb, gun

- Multisyllabic rhyme: Involve more than one syllable in each rhyming word.

Example: Elephant, relevant

- End rhyme: Occurs at the end of lines in a poem or song. Most common type of rhyme.
- Internal rhyme: Occurs within a line of poetry, rather than at the end.



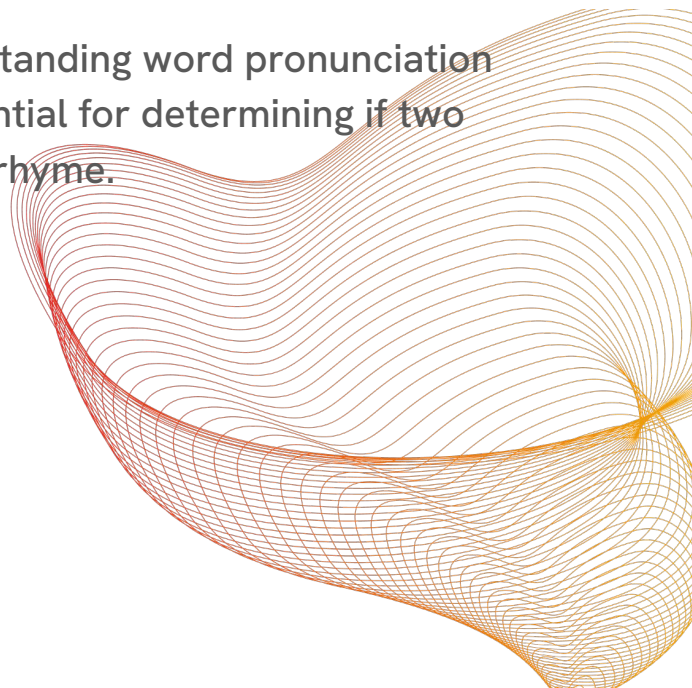


Word Pronunciation

English words often have unpredictable pronunciations compared to Vietnamese and other languages

Example: 'tough', 'though', 'through' and 'thought'

Understanding word pronunciation is essential for determining if two words rhyme.





Word Pronunciation

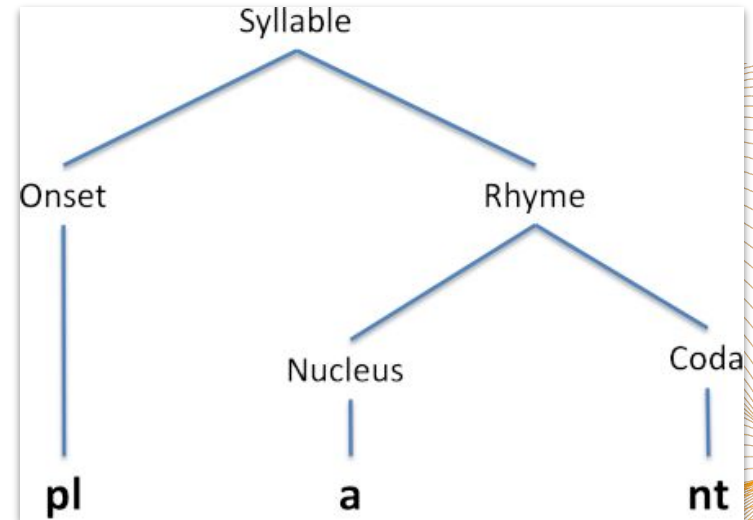
- A word is composed of syllables.
- Syllables are formed by combining one or more phonemes
- Phonemes are the smallest distinct units of sound in a language
- Phonemes can be categorized into vowels and consonants.

WORD	cat	ship	light
PHONEMES	/k/ /ă/ /t/	/sh/ /ī/ /p/	/l/ /ī/ /t/

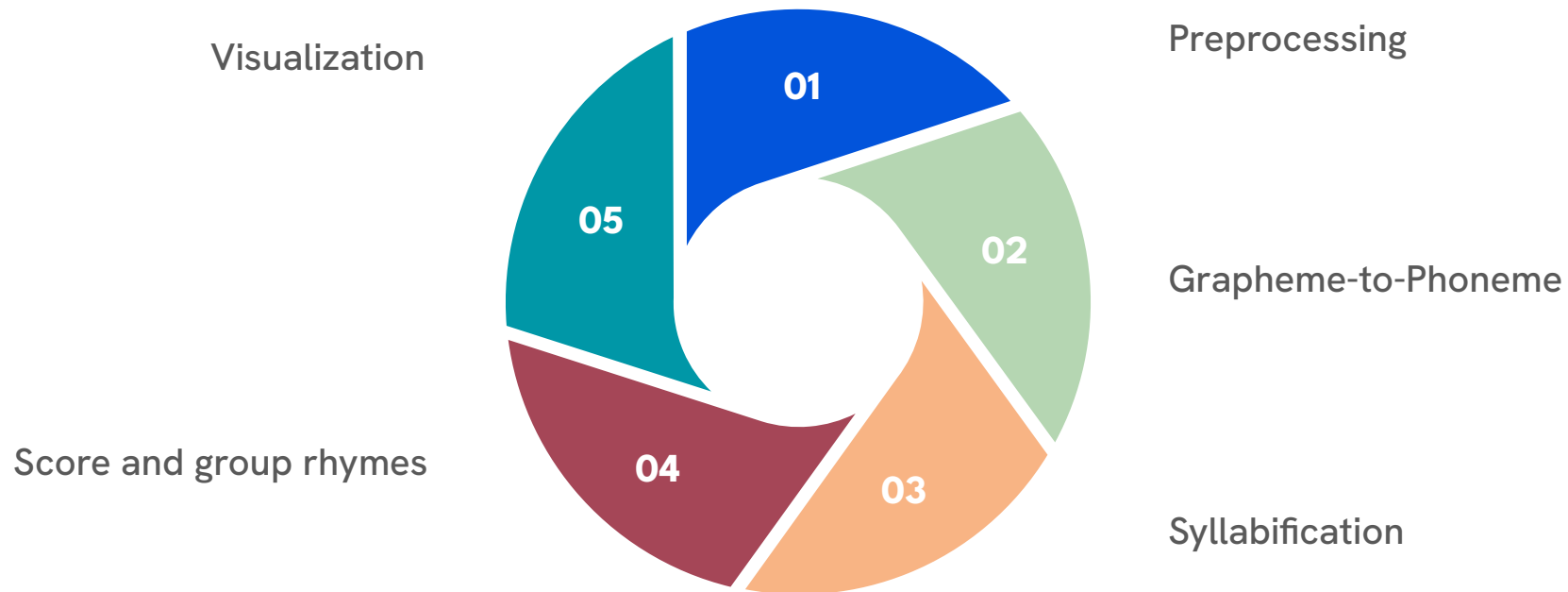


Word Pronunciation

- A syllable consists of three parts: onset, nucleus, and coda.
- The onset is the initial consonant
- The nucleus is the central vowel sound and is the main part of a syllable.
- The coda is the final consonants
- The onset and coda can be empty.



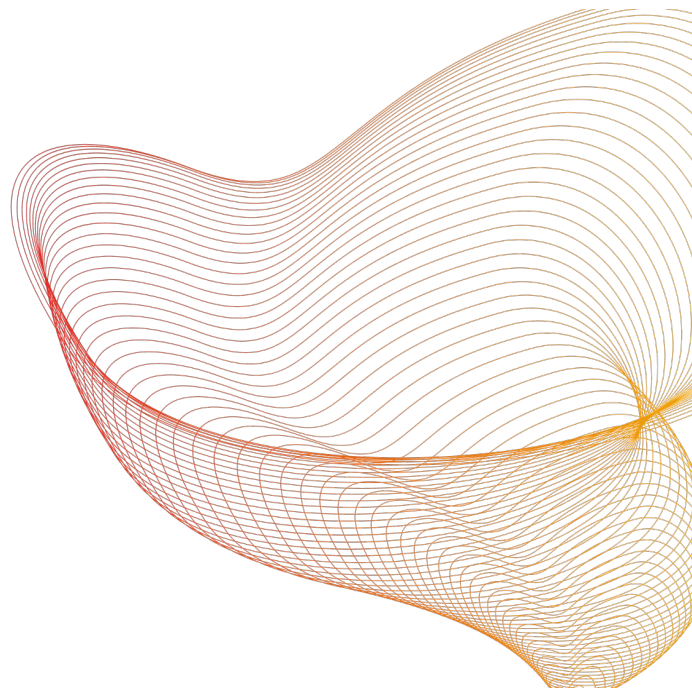
Rhyme Detector Pipeline





Preprocessing

- Text preprocessing is the initial step in rhyme detection.
- It involves removing punctuation, converting text to lowercase, and tokenizing the text into words.





Grapheme-to-Phoneme

- G2P is the process of converting a word from its written form to its spoken form.
- The CMU Pronouncing Dictionary is used for G2P, containing over 134,000 words with their pronunciations and stress patterns.
- The CMU Pronouncing Dictionary uses a modified form of the ARPABET system to encode pronunciations.
- The ARPABET system is preferred over the IPA system for readability.
- For words not found in the CMU Pronouncing Dictionary, a neural network model is used to predict the pronunciation.

Word	IPA Transcription	ARPAbet Transcription
parsley	[ˈpɑːsli]	[p aa r s l iy]
tarragon	[ˈtæɾəɡɑːn]	[t æ r ax g aa n]
catnip	[ˈkætnɪp]	[k æ t n ix p]
bay	[beɪ]	[b ey]
dill	[dɪl]	[d ih l]
garlic	[ˈɡɑːlɪk]	[g aa r l ix k]
mint	[mɪnt]	[m ih n t]
nutmeg	[ˈnʌtmɛɡ]	[n ah t m eh g]
ginseng	[ˈdʒɪnsɛŋ]	[j h ih n s ix ŋ]
fennel	[ˈfɛn]	[f eh n el]
clove	[ˈkloʊv]	[k l ow v]
thistle	[ˈθɪsl]	[th ih s el]
heather	[ˈhɛðə]	[h eh dh axr]
sage	[seɪdʒ]	[s ey jh]
hazelnut	[ˈheɪzlnʌt]	[h ey z el n ah t]
squash	[ˈskwɒʃ]	[s k w a sh]
ambrosia	[æmˈbrɒʒiə]	[æ m b r ow zh ax]
chicory	[ˈtʃɪkəri]	[ch ih k axr iy]
sage	[seɪdʒ]	[s ey jh]
licorice	[ˈlɪkəri]	[l ih k axr ix sh]
kiwi	[ˈkiwi]	[k iy w iy]
parsley	[ˈpɑːsli]	[p aa r s l iy]
yew	[juː]	[y uw]
horseradish	[ˈhɔːsrædɪʃ]	[h ao r s r ae d ih sh]
uh-oh	[ʊə]	[q ah q ow]
butter	[ˈbʌtə]	[b ah dx axr]
wintergreen	[ˈwɪntəˈɡriːn]	[w ih nx axr g r i n]
thistle	[ˈθɪsl]	[th ih s el]



Syllabification

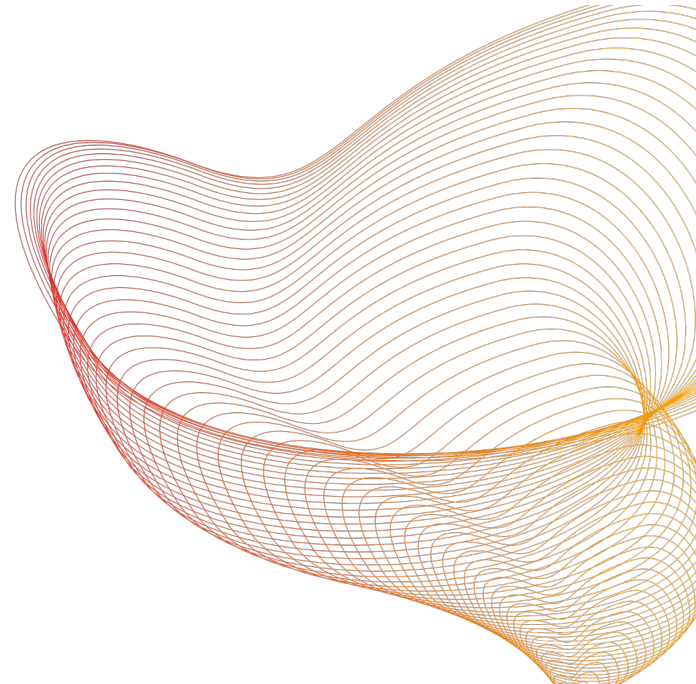
- Syllabification is the process of dividing words into their syllables.
- Factors such as vowel sounds and consonant clusters are considered in determining syllable boundaries.
- Much easier to syllabify word pronunciation than actual words

Syllable Division Rules	
<p>VCCV</p> <p>basket</p> <p>Divide between two middle consonants.</p>	<p>VCV</p> <p>cabin</p> <p>Divide after the consonant when the 1st vowel has a short sound.</p>
<p>VCV</p> <p>basic</p> <p>Divide before the consonant when the 1st vowel is long.</p>	<p>C+LE</p> <p>candle</p> <p>Divide before consonant -le</p>
<p>VCCCV</p> <p>complex</p> <p>With 3 consonants between vowels, usually split after 1st consonant (but keep digraphs and blends together).</p>	<p>VCCCV</p> <p>instruct</p> <p>With 4 consonants between vowels, usually split after 1st consonant.</p>
<p>windmill</p> <p>Divide compound words between the two words.</p>	<p>unkindly</p> <p>Divide syllables before a suffix and after a prefix.</p>



Grouping Rhymes

- Check each syllable pairs of 2 words from one line to the next
- Check end word of a line against the next 3 lines



Scoring Rhymes

01

Remove onsets

02

Align vowel and consonant pairs

03

Scoring rhymes based on vowel and consonant similarity.

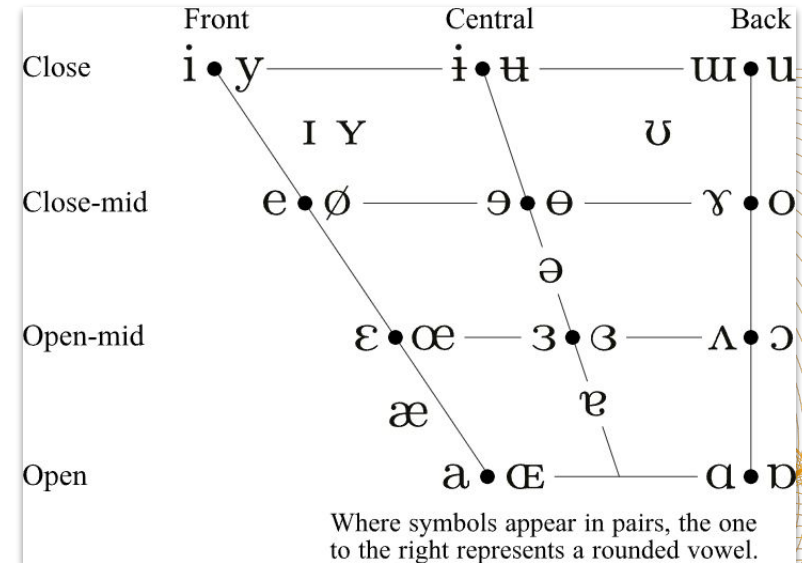
04

Weighting the vowel score over the consonant score to emphasize the importance of vowel sounds in rhyming.



Phonetic Features

- Vowels are classified based on height, backness, rounding.
- Consonants are classified based on place of articulation, manner of articulation, and voicing.
- From these features we classify vowels and consonants into similar sounding groups





Phonetic Features

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b		t d			ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ	n			ɳ	ɲ	ŋ	ɴ		
Trill	ʙ		r						ʀ		
Tap or Flap		ⱱ	ɾ			ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative			ɬ ɮ								
Approximant		ʋ	ɹ			ɻ	j	ɰ			
Lateral approximant			l			ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

Visualization

Tripping off the beat kinda, dripping off the meat grinder

TRIH1-PIH0NG AO1F DHAH0 BIY1T KIH1N-DAH0 DRIH1-PIH0NG AO1F DHAH0 MIY1T GRAY1N-DER0

Heat niner, pimping, stripping, soft sweet minor

HHIY1T NAY1-NER0 PIH1M-PIH0NG STRIH1-PIH0NG SAA1FT SWIY1T MAY1-NER0

China was a neat signer, trouble with the script digits

CHAY1-NAH0 WAA1Z AH0 NIY1T SAY1-NER0 TRAH1-BAH0L WIH1DH DHAH0 SKRIH1PT DIH1-JHAH0TS

Double dipped, bubble lipped, subtle lisp midget

DAH1-BAH0L DIH1PT BAH1-BAH0L LIH1PT SAH1-TAH0L LIH1SP MIH1-JHAH0T

Borderline schizo, sort of fine tits though

BAO1R-DER0-LAY2N SHIY1-TOW0-SOW0 SAO1RT AH1V FAY1N TIH1TS DHOW1

Pour the wine, whore to grind, quarter to nine, let's go

PAO1R DHAH0 WAY1N HHAO1R TUW1 GRAY1ND KWAO1R-TER0 TUW1 NAY1N LEH1TS GOW1

Ever since ten eleven, glad she made a brethren

EH1-VER0 SIH1NS TEH1N IH0-LEH1-VAH0N GLAE1D SHIY1 MEY1D AH0 BREH1DH-RAH0N

Then it's last down, seven alligator seven

DHEH1N IH1TS LAE1ST DAW1N SEH1-VAH0N AE1-LAH0-GEY2-TER0 SEH1-VAH0N

At the gates of heaven, knocking - no answer

AE1T DHAH0 GEY1TS AH1V HHEH1-VAH0N NAA1-KIH0NG NOW1 AE1N-SER0

Slow dancer, hopeless romancer, dopest flow stanzas

SLOW1 DAE1N-SER0 HHOW1-PLAH0S ROW0-MAE1N-SER0 DOW1-PEH2ST FLOW1 STAE1N-ZAH0Z

Results



0.84

Accuracy



0.93

Recall



0.90

Precision

Discussion

01 Speaker can choose to pronounce word in an unexpected way

02 G2P Model is not perfect

03 Doesn't takes the word position within a line into account

04 Machine learning model to optimize weights



Thank you. Please feel free to ask any questions. 😊