Title
Language-Based Music: Cognition and Computation

Author
Ackerman, Jordan Alexander

Publication Date
2022

UNIVERSITY OF CALIFORNIA,
MERCED

Language-Based Music: Cognition and Computation

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive and Information Sciences

by

Jordan Alexander Ackerman

DissertationCommittee:
David C. Noelle, Chair
Paul E. Smaldino
Ramesh Balasubramaniam

2022

The dissertation of Jordan Alexander Ackerman
is approved and is acceptable in quality and form for
publication on microfilm and in digital formats:

Professor Paul E. Smaldino

Professor Ramesh Balasubramaniam

Professor David C. Noelle, Committee Chair

University of California, Merced
2022

DEDICATION

To my parents and grandparents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would first like to thank my family (Farrell, Jan, Molly) and friends (Drew Doallas-Baxter, Sam Spevack, Laura Kelly, Tim Shea, Chelsea Gordon, Ben Falandays, Josh Clingo, etc.. you know who you are). Special thanks to my wonderful and supportive partner, Sabina Sloman, who keeps me inspired and smiling.

I would also like to acknowledge the generous support from members of the Cognitive & Information Sciences program at UC Merced, particularly my committee members David C. Noelle, Paul Smaldino, and Ramesh Balasubramanium.

Thanks also to the Cognitive Science Society for allowing me to reproduce some content published in their proceedings in the current Chapters 4.1, 5.1, and. 6.1.

I have been lucky enough to be supported by a NSF GRFP fellowship in NLP and a NSF NRT fellowship in Intelligent Adaptive Systems during my graduate education, enabling me to pursue work from language science to computational social science to complex systems. These interdisciplinary domains have inspired and informed my perspective, both methodologically and philosophically, and I am grateful for such a remarkable opportunity.

VITA

Jordan Alexander Ackerman

EDUCATION

PhD in Cognitive & Information Science **2022**
University of California, Merced *Merced, California*

BA in East Asian Studies **2011**
New York University *New York, New York*

RESEARCH EXPERIENCE

NSF GRFP Fellow - Natural Language Processing **2017–2022**
University of California, Merced *Merced, California*

NSF NRT Fellow - Intelligent Adaptive Systems **2017–2018, 2021–2022**
University of California, Merced *Merced, California*

TEACHING EXPERIENCE

Teaching Assistant - Linguistics **2016–2017**
UC Merced *Merced, California*

REFEREED JOURNAL PUBLICATIONS

Pokemonikers **2018**
Proceedings of the Linguistic Society of America

Why Don't Cockatoos have War Songs? **2021**
Behavioral & Brain Sciences

REFEREED CONFERENCE PUBLICATIONS

Entropy of Sounds: Sonnets to Battle Rap **July 2020**
Proceedings of the Cognitive Science Society

Of Pieces and Patterns: Modeling Poetic Devices **July 2021**
Proceedings of the Cognitive Science Society

What's Different about Improvised Rap? **July 2022**
Proceedings of the Cognitive Science Society

SOFTWARE

Phonesse `https://github.com/jordanmasters/phonesse`
Toolkit for extracting, manipulating, and visualizing language sounds from text

Phonesse Docs `https://jordanmasters.github.io/phonomials`
Phonesse Documentation

Rhymable `http://rhymable.com`
Interface for visualizing sound from text samples

rhyme-data `https://github.com/jordanmasters/rhyme-data`
Data Repository

ABSTRACT OF THE DISSERTATION

Language-Based Music: Cognition and Computation

By

Jordan Alexander Ackerman

Doctor of Philosophy in Cognitive and Information Sciences

University of California, Merced, 2022

David C. Noelle, Chair

Repeated sound sequences in language occur all the time, but we reliably notice them in popular poetic devices like alliteration, assonance, and rhyme. We leverage their sound structures in acquiring our vocabularies as children, and in our most prized literary works. Verbal sound patterns can even serve to scaffold music-like structure within language. Indeed, humans seem to find music-like verbal patterns compelling and productive enough to spend effort including them in rhetoric, poetry, lyrics, and advertisements throughout history. Yet, understandings of their forms and dynamics are still quite limited. In particular, rap rhymes often sport dense phonological patterns whose complexity has been shown to increase over time [11], yet commensurate analysis has not followed. Lyrical data from rap will therefore serve as the target of much of the current investigation. Much like producin'g language or music, producing complex phonologically patterned speech is a skill that, in and of itself, is worthy of investigation. I will introduce visualization and computational tools to explore various exemplar data and their sound structures, framing many of the findings in terms of cognition.

Chapter 1

Introduction

1.1 Language-Based Music

> “Spontaneity is a meticulously prepared art.” -Oscar Wilde

> “It should be noted that the broader problem of measuring the information connected with creative human endeavor is of the utmost significance.” [12] -Kolmogorov

Music and language exhibit many evident similarities. For example, they both display organized structure, involving sound pattern production and recognition. They can also demonstrably borrow from each other. In a practice that has gained deserved attention, referred to as *musical surrogate languages* [13, 14, 15], meaningful messages are communicated through sound patterns played on musical instruments, like drums or flutes. This enables linguistic communication, but using the tools of music (rhythms, sequence of notes). This is linguistic communication, but in a musical form, which may called music-based language.

It has recently been suggested that the reverse phenomenon also exists, that is, music, but within a linguistic form. McPherson calls this phenomenon language-based music, or ‘musical genres that involve the adaptation of linguistic form to musical settings’ [15]. She points to two main exemplars of this phenomenon, rhyme and text-setting to metrical grids. Language-based music may also comprehend a range of other rhythmic and sound patterns (e.g. imperfect rhyme, assonance, alliteration). This dissertation focuses on the phonological patterns of language, from normal speech and literature to language-based music across domains like poetry and lyrics. I utilize tools from

NLP and computation to ask cognitive and linguistic questions about language-based music phenomena.

The ultimate goal of this work is to broaden the scope of investigation into language-based music and make steps towards informing pedagogical and AI applications in language learning. For instance, anyone seriously attempting to rap improvisationally (myself included) quickly learns that there is no codified understanding of the elements involved, nor is there pedagogy to facilitate learning this popular and elusive skill. Indeed, only a few dozen rappers in the world might be reasonably considered 'fluent' in improvised rhyme. Despite the fact that phonological complexity in rap has grown over time [11] this has not spurred commensurate formal analysis or curricula that incorporate insights obtained from research. What is necessary is a more formal understanding of this complex skill, like one that has already been established for Jazz music. But unlike in music, where notation systems, analysis, and pedagogy are abundant, in the creative language arts, there are no comprehensive frameworks for describing and learning complex structures. Consequently, an interdisciplinary approach is needed to address this gap and make the educational and creative benefits of improvised rhyming more accessible. Inspired by the richness of musical formalizations and language learning pedagogy, I combine elements from computational linguistics, complexity science, and quantitative and experimental methods in cognitive science, to build a solid foundation for identifying (universal) patterns of sound organization in creative language use that is reflective of learning dynamics and human culture.

In this dissertation, I will show how, in both size and complexity, the forms and usages of phonological patterning have outgrown the traditional vocabulary of poetic devices normally used to describe them. Scientists across fields have graduated from studying only relatively simple or toy examples, to examining more ecologically valid data from the wild, in order to identify the nature of their complexity. In much the same way, I will look in between and across the conventional poetic or lexical classifications to capture the dynamics of the under-explored and complex phenomenon of repeated verbal sounds. This dissertation will introduce visualization techniques and use methods from natural language processing and complexity science to describe repeated verbal sounds, not only from poetic literature, but across language genres (e.g. speeches, conversations, confusable phrases). Throughout, I will also discuss why it is important to integrate what is known about cognition and language in order to achieve a better understanding of verbal sound patterning and language-based music.

1.2 Rhyme

In this section, I briefly review the perspectives from which rhyme has been investigated. Given the interdisciplinary nature of language-based music, it is important to begin by highlighting the breadth of academic fields that may be relevant for a comprehensive description of this phenomenon. A more detailed discussion of these topics can be found in Chapter 3.

In the literature, scholars have explored rhyme (and even rap) from various disciplines including, poetics, music, culture, linguistics, and computation.

In poetics, efforts range from focuses on structure, utility, and aesthetics [16, 17, 18], to individual case studies (such as a Pharoahe Monch hip-hop album) [19], to cognitive poetics [20, 21]. For example, neuropsychological studies have begun to confirm suggestions from early cognitive poetics that popular constraints, such as meter and rhyme, do impact our aesthetic and emotional perception of poetry [21]. This work largely represents an effort to understand the composition and aesthetics of poetic materials.

In music, rhyme is examined in songs from antiquity [22] to modern genres like rap [23]. In this domain, efforts focus on rhyme and its relation to meter, rhythm, and flow [24], pattern placement, and syntactic units [25]. Studies on the syntax of improvised scatt vocalization (jazz) are also relevant to the mechanics of verbal sound patterning [26]. Generally, this manifests as an integrated approach where sound patterns, and their connected linguistic elements, are considered in relation to musical dynamics.

In socio-cultural analysis, two broad areas of investigation have emerged. First there are efforts to understand popular rhymes in terms of sociology [27], cognitive psychology [28] and education [29, 30, 31, 32]. Forms such as rap also command special attention in domains like African-American studies [33]. Second, the cultural evolution of verbal sound patterning has been touched on at many scales. Covered topics range from the debate over the origins of language and music (e.g. "Which came first?") [34, 35, 36], to observations about the commonality of phonological patterning in rhetoric among tribal leaders [37], to iterated learning models of sounds from which syllabic templates emerge [38].

In linguistics, there has been a particular focus on the named forms of rhyme [39]. Some have suggested that rhyme should be rigidly defined [40, 41], while others are more integrationist [42], 'viewing rhyme itself as in constant interplay with other sound patterning' [43]. An important line of research on imperfect rhyme has also developed, which explores variation and similarity in rhymes that are perceptually acceptable, yet not 'perfect' segmental matches. These efforts span from rock [44] to hip-hop [45], and often focus on perception and similarity [46, 47, 48, 44, 49, 50, 51]

Finally, in computation, verbal sounds have been used in stylometrics [52, 53, 54, 55, 56, 57, 58], rhyme identification [9, 59, 60, 61], and lyric generation [62, 63, 64, 65, 66, 67]. Recently, a corpus analysis on rap lyrics demonstrated effects such as increases in rap rhyme complexity between 1980 and 2000 [11], as well as general tendency for more variability between individual songs than between artists [68]. Even as far back as 1965, it was noted that while English characters (at the time) had an estimated source entropy of 1.9 bits per character, it is likely that works from artistic disciplines, such as sonnets, would have more constraints (predictability) and therefore, an entropy closer to 1.0-1.2 [12].

1.3 An Interdisciplinary Approach

Given the universality of creative phonological patterns in language, it is natural that this domain has been explored from so many disparate fields, and it is my belief that all of them are needed to inform a more holistic picture of this popular human process and form.

But there are significant gaps in the literature. There is a striking dearth of theory, tagged corpora, corpus studies, and cognitive studies on rhyme. Most efforts in neuroscience or behavioral psychology, while insightful, employ rhyme peripherally and as a proxy for phonological processing, in contrast to syntactic or semantic processing. Unlike music, creative verbal sound patterning has an underdeveloped notation system (outside of phonetics), and almost no pedagogy or robust formalizations of the sequences of elements involved. Consider that musical culture provides not just formalizations of the elements or notes (pitches) to its community, but also scales, chords, progressions, melodies, and various other complex forms for which there are no explicit analogies in the (phonological) language arts. The absence of discriminating tools does not necessarily mean these patterns are absent in language, but rather, that they may not be noticed or labeled; interesting phonological patterns can often be obscured by the other salient layers of information and complexity humans pack into language.

Broadly following Jakobson's integrationist view of rhyme, I explore sound patterns of language-based music that transcend boundaries of rigid poetic classifications in order to begin uncovering the true complexities of language-based music. For this purpose, I use NLP, machine learning, complexity science, and interactive visualizations to discover patterns in primary sources that are hard to see using conventional methods. This is an early step towards a more dynamic view of 1) the landscape of language-based music, 2) the structure of repeated verbal sounds, and 3) the learned skill that these patterns reflect.

1.4 What has Rhyme Become?

Although the target of this investigation has much broader scope than any definition of rhyme can capture, it is instructive to point out the growing influence and scope of the term. In this section I will introduce a few common definitions of rhyme, the range of the cultural phenomenon, and the learned skill that it represents.

First, rhyme, or rime, is a phonological unit within the syllable, Encompassing the nucleus and coda. Second, rhyme has become a catch-all term to refer to many classes of similarity between words. There are hundreds of specific forms that are referred to as rhyme. Some are narrow, some are positional, some are overlapping and under-specified. And third, rhyme is often the term used to refer to the cognitive process of perceiving, producing, and learning these forms. The second and third definitions are intimately connected; cognitive mechanisms constrain the way humans produce repeated sound patterns like rhyme, and the rhyming artifacts that other humans produce are the objects that our cognitive systems perceive. These different types of rhyme are illustrated in Figure 1.1.

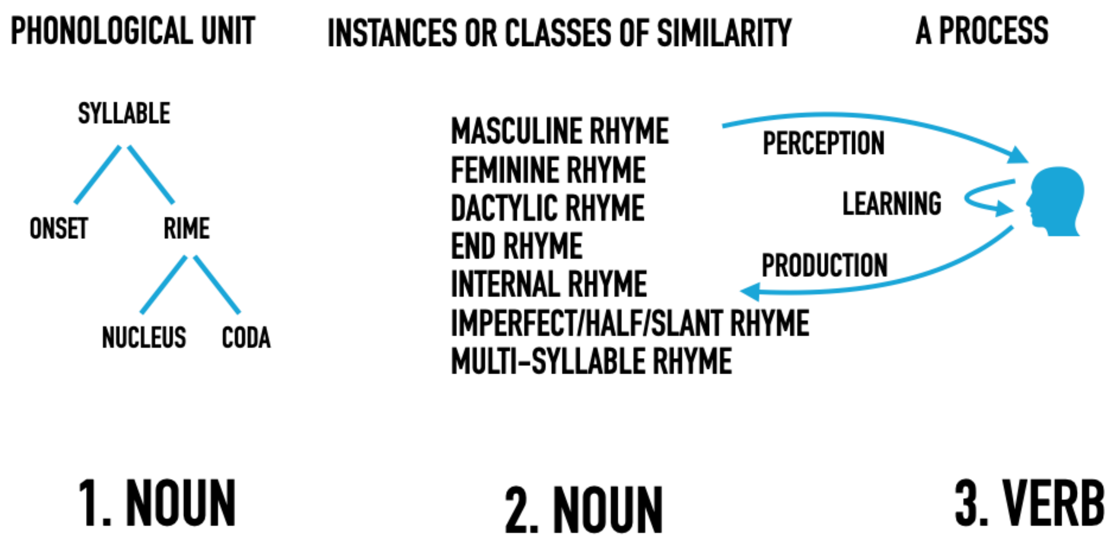


Figure 1.1: 3 Definitions of Rhyme

I will sometimes refer to the first definition, and will always indicate this usage with the spelling rime. I identify these other broader definitions of ‘rhyme’ (Defs. 2 & 3) in order to highlight two high-level senses of the word that I will be focusing on in this project. Admittedly, these are both under-specified and imprecise umbrella terms and I will attempt to disambiguate these notions by framing them as either form (recorded cultural output) or process (learned skill, cognition).

1.4.1 A Cultural Phenomenon

Prehistory

Whether humans developed musical or verbal patterning first is an open question that has motivated much debate [69, 35, 36]. Indeed, it seems plausible that group selection and culture niches may explain the diversity seen in music [70]. And, in Chapter 3, I will explore a related question; what cognitive mechanisms are responsible for the common human practice of verbal sound patterning?

It has long been assumed that music is a human universal [71], and this claim was recently verified across many different known cultures [72]. Is rhyme also a human universal? Thus far, it is unclear. However, rhyme is certainly a prominent feature of verbal art practices all around the world. It is possible that it exists in all known languages, but its relative frequency and use likely depends on the variety of cultural, grammatical, lexical, and phonological attributes of any given language. Determining the universality of rhyme is a question for future cross-linguistic analyses.

Although evidence of musical instruments only dates back 40,000 years [73], it is believed that the phenomenon of music is much older. Indeed, evidence of the music-related elements of perception and production far predates this period. On the one hand, it is suspected that human voice boxes developed into their modern form some 530,000 years ago, which suggests that early humans (and neanderthals) could have had the physical ability to speak and even sing [74]. The evolutionary development of rhythmic capabilities, particularly those specific to human musical and linguistic domains, is still debated [75], but the literature suggests that many of the basic elements needed for the perception and production of musical and verbal forms, namely, the appropriate anatomical parts and cognitive abilities, have been around for thousands of years.

Early History

The earliest written evidence of rhyming comes from a set of Chinese works called the Book of Songs, collected between 1200-600 B.C.E. [76, 77]. As the title suggests, this work contains the rhyming lyrics of songs, but it also contains verse that is believed to be independent of music (i.e. poetry). Across cultures, rhyme has developed in various ways. In western culture, many poetic forms, like iambic pentameter and end-rhyme, were made popular during the renaissance. In the Persian tradition, the science of rhyme has been discussed since at least the 14th century [78].

Anthropological studies have also begun to reveal early cross-cultural pressures for "unusual linguistic knowledge and rhetorical skill in individuals," particularly in 'big-men' or tribal leaders. These rhetorical devices have been present for thousands of years in many traditions. They often rely on "subtle uses of alliteration, rhyme, and consonance, and the development of a kind of sprung rhythm after an initially regular metrical form". These skills are commonly displayed by men in public settings, or in exhibitions such as verbal duels, which pit individuals against each other, armed with insults, humor, and various types of evidence for cleverness. Social displays like this are believed to reward proficient competitors with "enhanced dominance and new opportunities for sex" [37].

A more recent ritual also follows in this tradition. The Dozens (a.k.a Dirty Dozens), a game of rhyming insults, was often played by African-American men in the early 20th century [79, 80].

In most cases, the primary use and focus of rhyme has been on end-rhyme: rhyme that occurs at the very end of a line. This type of structure can be captured by the 1-dimensional representation scheme traditionally used to annotate rhyme patterns. This notation works well whether the lines are successive (AABB, AABBA, etc.) or alternating, as in common in sonnets and limericks (ABAB...). This kind of representation well-represents the single syllable patterns that typified rhyme for much of recorded history. Even subcultures like hip-hop, which are now known for their multi-syllabic rhymes, were dominated 1-syllable end-rhymes early in their development.

These historical considerations point towards a clear human capacity for musical and verbal art. But, while many reviews have been conducted on musical forms and their underlying cognitive mechanisms, intentional phonological patterning in rhetoric and art have yet to draw such scholarly attention. This gap is only highlighted by the fact that humans seem to use creative or intentional sound patterning in nearly all contexts that involve language (e.g. prose, verse, humor, conversation, advertising, journalism, marketing). This is anecdotally demonstrated in Figure 1.2 which shows that the frequency of alliteration, assonance, and rhyme used in 2020 Economist article titles was higher than the base rates for those devices collected from random 3 and 4 word phrases from the Brown Corpus of sentences.

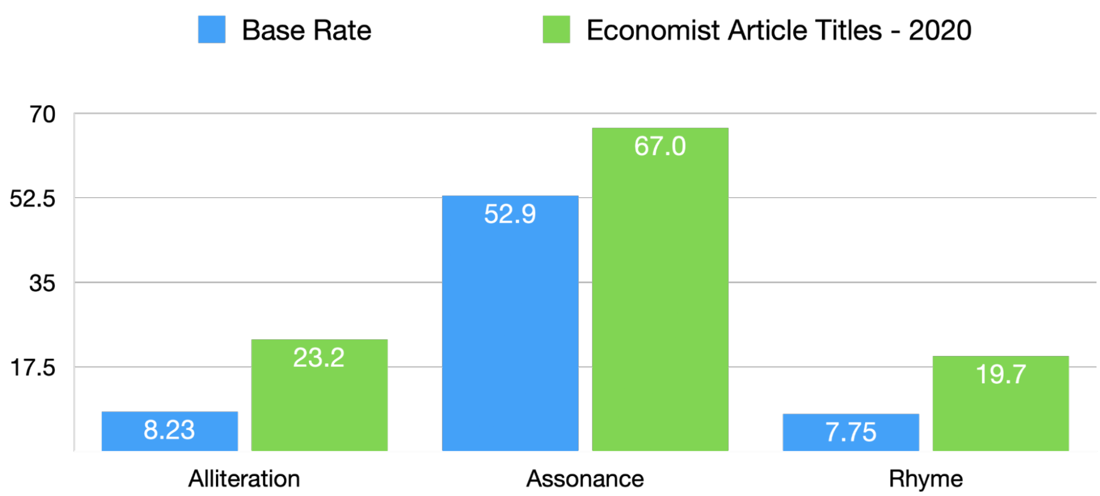


Figure 1.2: Poetic Device Usage in Magazine Article Titles. Article titles represent all 2020 issues of The Economist

Modern History

It may seem natural that rhyme has not motivated as much research as music. After all, rhyme schemes and related devices are often perceived to be much less complex than musical forms. In addition, language involves more than just sound patterns. The usual use of language is constrained at various layers, including semantics, syntax, compositionality, etc. But as newer forms of verbal expression, like hip-hop and improvisational rapping, have become increasingly popular, the sound patterns they produce, improvisationally and in writing, have also grown in size and complexity (e.g. 2-10+ syllable imperfect patterns, often repeated dozens of times). As mentioned earlier, this trend of increasing rhyme complexity in hip-hop has been demonstrated in the period from 1980 to 2000 [11]. Hip-hop continues to display enormous growth in popularity [81].

Rappers often use clever punchlines and sound patterning as marks of achievement or demonstrations of skill in the hip-hop community. In general, as the 1990's progressed, it became much more common to hear 2-syllable and 3+ syllable rhymes in songs. But hip-hop rhyme did not develop within the context of music alone.

In 1993, a style of competitive rapping, battle rap, began to be formalized in a yearly competition called the Rap Olympics. Before that, battle rap had existed largely as an informal activity or display where people would try to impress onlookers by insulting each other with clever rhyming verse. This is reminiscent of the 'big-men' and 'the dirty dozens', mentioned earlier. This competitive verbal performance also has roots in the early hip-hop tradition of DJ battles, where DJs would set up across from each other (e.g. on either end of a basketball court) and try to attract the largest crowd with their record selections and mixing strategies.

Battle rap is an arena where rappers can sharpen their lyric writing abilities, especially with regards to crowd response, clever punchlines, and rhyming. Battle rap is fundamentally a turn-taking activity which began in the context of improvised rapping to a beat (hip-hop music or beat boxing). This sets up two important constraints. First, pattern complexity – it is difficult to make lyrics phonologically complex when they are improvised, as a great deal of online cognitive effort must go towards making sense and attending to grammaticality. Second, battle rap lyrics were originally generated and delivered simultaneously with music, which creates its own set of rhythmic constraints and time pressure. The largest of these organizations, such as The Rap Olympics or Scribble Jam, are no longer in operation, but their many of their videos can be found on YouTube.

A notable progression in the development of battle rap was to avoid the constraints that a beat imposes by transitioning to a cappella (but still improvised) rap battles (e.g. World Rap Championships, 2004-2007). This allowed rappers to focus entirely on the lyrics, but it also led to a lot more 'sneaking in' of pre-prepared written content (punchlines, rhymes, insults). In any event, only a handful of rappers were good enough at improvisation to sustain the growing demand for this sort of entertainment. Outside the elite performers, the requirement for improvisation seemed to result in a reliably lower quality of content. Around 2006, battle rap leagues like Elements League and Grind Time Now embraced an a cappella and written format, which allowed for more impressive content in battles. Permitting lyrics to be openly premeditated, rather than improvised, also led to much larger and complex phonological patterns (e.g. rhyme, stress, and alliteration schemes). These changes in the constraints and cognitive affordances of battle rap may also have played a role in the increased usage of multi-syllable rhyming. Today, although all these formats still exist, written a cappella performances dominate this growing subculture (90%+), which boasts almost 800 different leagues, and over 23,000 rappers [82].

It should also be noted that various forms of rhyme have relied on perfect agreement, that is, a matching of the vowel and all consonants following it i.e., the rime, at the end of a word. Rap, in particular, has pushed the boundaries of perfect matching and flexibility.

1.4.2 A Skill

Although there are currently no comprehensive notation systems for, or theories of, language-based music, it is a domain where humans are able to develop skill, even fluency. From language acquisition to rhetoric to literature to poetics, humans learn to manipulate both the overt and subtle elements of sounds patterning and organization in language. While almost all humans are familiar with rudimentary perception and production of rhyme, we are capable of much more expertise and complexity, not only in terms of the premeditated production common in canonized lyrics and poetry, but also in improvisation. Some singular individuals even make it clear that rhyming is a deep skill, like playing music, which can be acquired at a high level of expertise.

If you were presented with the words, “blue sky”, “Denver”, and “Cream”, how might you work them into rap lyrics? How long would it take you? Harry Mack, an improvisational rapper from L.A., observes stimuli and creates improvised raps concurrently. At the same time as he is reading words that his fans write into a YouTube live-stream chat, he is integrating them into brand new improvised lyrics over music – as thousands of people watch live. He does this for hours at a time [83].

The ability to improvise this way is a highly honed skill. Although he is talented as both a drummer and rapper, Harry Mack has learned to produce language in this way through almost 20 years of intentional practice. He teaches improv rap in private lessons, but the sciences of cognition, learning, and language that might explain his acquisition of this skill have yet to be developed. In Chapter 3, I review the cognitive mechanisms that enable both traditional and improvised uses of rhyme. Then, in Chapter 6, I conduct case studies on different forms of improvised rhyming practices (including Harry Mack).

1.5 Getting Started

In English, there are over 200+ specific named forms of rhyme (e.g. Masculine, Feminine, Syllabic rhyme). A perfect masculine rhyme occurs if a pair of words share the same final vowel (nucleus) and coda (the consonants after the vowel in a syllable) — also their syllables must have primary stress. Anything that meets that definition is a masculine rhyme (e.g. can-ran), anything that does not meet that definition is not (e.g. ran-ram). Almost all of the terminology used to classify poetic devices relies on this notion of perfect or exact matching of some conditions.

But not all rhymes are perfect. From the limited work on imperfect rhyme [84, 44, 85, 86] it is known that there are important regularities in the ways humans bend or break these conventional perfect matching rules while still maintaining the illusion of rhyme. Data of this type are usually referred to with the catch-all term 'imperfect rhyme', though they come in a rich variety of alternative forms. Even in a context where perfect rhyming is the dominant form, there are regular and imbalanced amounts of matching in the elements of rhyming syllables.

For example, 83% of disagreement in imperfect rhymes by British poets are explained by changes in fricative voicing (e.g. love-cuff) [85], whereas in the rock music genre, 49% of disagreement in imperfect rhyme is due to changes in nasal place (e.g. ran-ram) [44]. What drives this difference? Phonological properties of a language or dialect? Cultural conventions or Individual differences? Perceptual biases?

Additionally, multi-syllable rhyming patterns are extremely popular in rap lyrics, which opens up opportunities for variation to occur in many different (consonant) locations (Rollecoaster / Nova Scotia / Coca Cola), not just the final coda (ram/man). And in fact, it is not just the consonant sounds that form imperfect patterns. There are surprising structures and imperfections within the stress and vowel patterns of multi-syllable patterns in language.

So, what are the underlying linguistic constraints on patterns like this? And how can the ways in which cognition shapes those constraints be understood? Many factors likely play a role, but there is not yet a broad or deep enough understanding of the phenomenon of imperfect verbal sound patterning to identify specifically what they are. Ideally, such a research program would be able to concretely answer questions such as the following:

- How does an individual or a culture's use of repeated sounds change over time?
- In what ways, and how much, do the sound patterns used in language change under different cognitive or task constraints?
- What is the role of similarity (perceptual, phonological, acoustic) in constraining these patterns?

- What are the dynamics of sound patterning in rap lyrics (across subcultures), where imperfect rhyming is the norm?
- How can creative verbal sound patterning be better formalized to enable thriving scientific and pedagogical communities (as exists in music)?

I will address these questions, answering some in Chapters 4, 5, 6, and considering others in Chapter 7.

There are at least two fundamental lines of research that must be expanded on in order to address these questions. First, it is important to document sound patterns in relevant data from the wild to capture the true diversity of human language production. Second, these descriptive analyses can be used to drive various perceptual and behavioral studies, testing hypotheses about the role cognition plays in this phenomenon (e.g. similarity, category membership, confusability, learning, expertise, phonological processes).

I focus on the first issue, which is largely an exploratory and descriptive endeavor, framing much of the discussion in terms of the second issue. Specifically, I present a framework for analyzing repeated patterns in speech sounds and discussing their cognitive implications.

There is a need to document sound patterns in the wild to capture the true diversity of human language production. I take both bottom-up and top-down approaches to this problem. From a top-down perspective, I represent all the sounds of language, not just overt devices, in three formats, bag of words, serialized, and as grids (matrices), each capturing different structural assumptions of repeated patterns. I describe these data with computational linguistics and complexity science to quantify their patterns. Finally, I use both supervised and unsupervised machine learning to cluster and classify them. From a bottom-up perspective, I leverage human annotated (overt) data, using linguistic and complexity tools to characterize underlying sound patterns in sets of annotated utterances (e.g. rhyme sets, confusable phrases). These two approaches are outlined below.

1.6 Problem 1 (Top-Down)

How can the repeated sounds of language be measured without first labeling all the patterns?

Conventional or static definitions of poetic devices have been somewhat productive in the linguistic framing and discussion of poetic language. This approach follows in the tradition of Aristotle, who contended that category membership is decided from lists of defining features [87]. But, one problem with this approach is that humans have biases — we only label things that are noticeable or accessible to label. Sometimes patterns are perceptually confusing, complex, or we don't have terminology for them, and they can slip by unnoticed.

Let's look at an example and try to identify the largest poetic sound structures in this poem.

A Flea

A flea and a fly in a flue
Were imprisoned, so what could they do?
Said the fly, "let us flee!"
"Let us fly!" said the flea.
So they flew through a flaw in the flue.

Ogden Nash

You likely noticed two patterns, repeated use of the /f/ + /l/ sound for alliteration, and an AABBA end-rhyme pattern, but did you find a third? The vowels of line 3 and line 4 are identical, a form of assonance. How do we so easily miss large patterns like this? The point is, humans don't always notice repeated sounds in language, even if the patterns are large and in a context that is being actively examined.

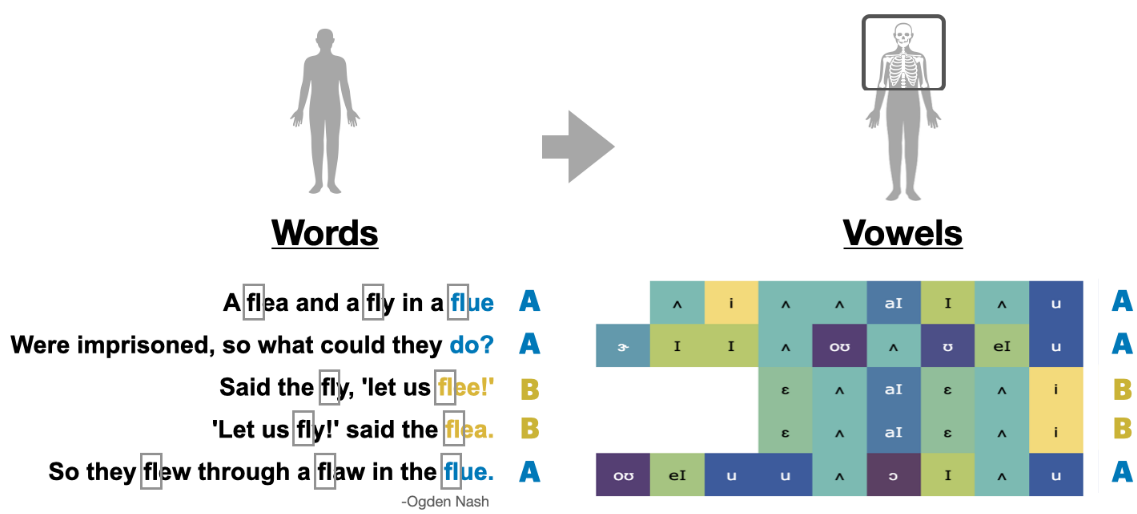


Figure 1.3: Color-coded and right-aligned vowels the poem A Flea

With this in mind, it is important to capture these patterns without having to rely on humans to first decide on where the patterns are. To do this a visualization such as that provided in Figure 1.4, allows us to examine the sound patterns more holistically.

Of all the verbal sound patterns in language, humans consciously attend to some, label a subset of those (e.g. rhyme, alliteration, assonance), and leave the rest unaccounted for. This spectrum, from the natural sound patterns of language, to the surplus sound patterns of poetic and other language, is filled with both undocumented and unclassified phenomena. In this dissertation, I explore samples of language across this space and present methods for quantifying and comparing the underlying structure of their sound patterns.

For this purpose I use natural language processing, machine learning, and interactive visualizations to uncover patterns in primary sources (big and small data) that are otherwise undetectable using conventional methods. A focus on repeated verbal sounds may seem like the study of only poetic devices, but most of the speech humans hear contains sound patterns that are not actively attended to. Indeed, most conventional speech contains repeated sound patterns that the speaker may have, or may not have, been conscious of producing. Many of these patterns are built into phonology (sound systems in language), morphology (system for relating different forms of words), and syntax (a system for clausal organization of words), and it should be expected that most speakers produce speech containing sound patterns consistent with the regularities of a given language. But, if an amount of patterning exceeds some baseline, one might surmise that energy has been exerted to further structure those sounds.

1.7 Problem 2 (Bottom-Up)

How can the sound patterns within overt annotated sets of utterances be measured?

Sometimes it is known where poetic devices occur within a sample because humans (or AI) have identified them. Within the space of identified patterns (e.g. rhyme, alliteration, assonance), there are yet more imperfections, systematic structures, and variations that listeners are unaware of, or cannot describe well (or at all). Moreover, they do not fit neatly into the simple categorical definitions of poetic devices enumerated previously.

Let's take a look at three example sets of phrases. What sound similarities are noticeable?

Sets 1 and 2 do not rhyme, while Set 3 presents some kind of imperfect multi-syllable rhymes where many of the vowels, consonants, and stress are similar. Indeed, Set 1 is a pseudo-random collection of phrases.

If you have a suspicion that Set 2 is a bit less random than Set 1, you are not wrong. A cursory look at the underlying metrical patterns of these sets is revealing. Just like the rhymes in Set 3, all the phrases in Set 2 follow a very similar stress sequence

Set 1	Set 2	Set 3
Sitting on the porch	Bart Simpson Bouncing	Really though, thanks
How are you, mister?	Rotating Pirate Ship	Video tapes
Democratic state	That isn't my receipt	Hillary banks
Communication	Lobsters in Motion	Live in the lake
Let's go for tacos	That is Embarrassing	Gimmie a shank
Eco friendly team	Lactates in Pharmacy	Amphibious face
Computer programs	I'm Chasing Martian	Fill in the blank
Cosmic alignment	Baptism Piracy	Giving me breaks
On the cliffs edge	That isn't Mercy	Obsidian tank

Figure 1.4: 3 sets of phrases with some phonological similarity. Can you notice their internal sound similarities?

profile while set 1 phrases do not (seen through the vertical alignment of stressed vs unstressed syllables). This helps account for the sense of rhythmic or metrical similarity within Sets 2 and 3, but it is not the end of the story.

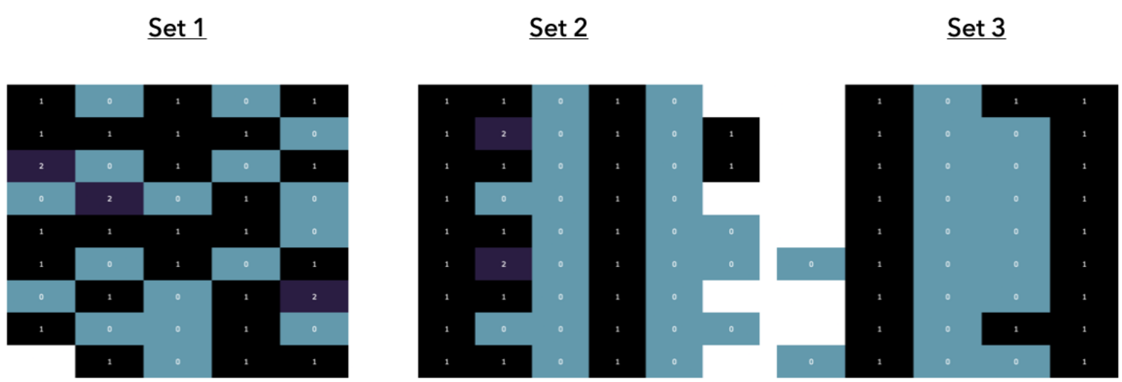


Figure 1.5: The underlying vowel patterns of phrases from word sets 1, 2, and 3 (above). Black represents a syllable with primary stress, teal is unstressed, and purple is secondary stress.

While Set 1 is composed of random phrases, Set 2 contains phrases from a recent viral meme which leverages perceptual confusability. For full context, its confusability arises while listening to a noisy British pub chant on repeat. Each time you hear the chant while reading any phrase from Set 2, it sounds like the recording has changed to whichever phrase you are currently reading (when really it is always the same - hear for yourself [88]). The amazing thing is that, despite their demonstrated confusability, the phrases from Set 2 still seem quite different on the surface. Can metrical stress alone really explain this much confusability?

Just like in the poem above (A Flea), one can observe just how similar the sounds of these phrases are by looking at their underlying vowels. Importantly, the vowels are color-coded such that more-similar vowels are represented by more-similar colors (more on this in Chapter 2). As might be expected, there is little visible structure in Set 1 and quite a lot of structure in the rhyming phrases of Set 3. The big surprise, however, is just how similar and well-aligned the vowel sounds of set 2 are. So, not only is Set 2 self-similar in terms of stress (as seen above), but also in its vowel sequence profile. This begins to demystify that the confusability of this viral meme's phrases are not some accident or magical effect, but rather, seem to be directly related to underlying phonological similarities.

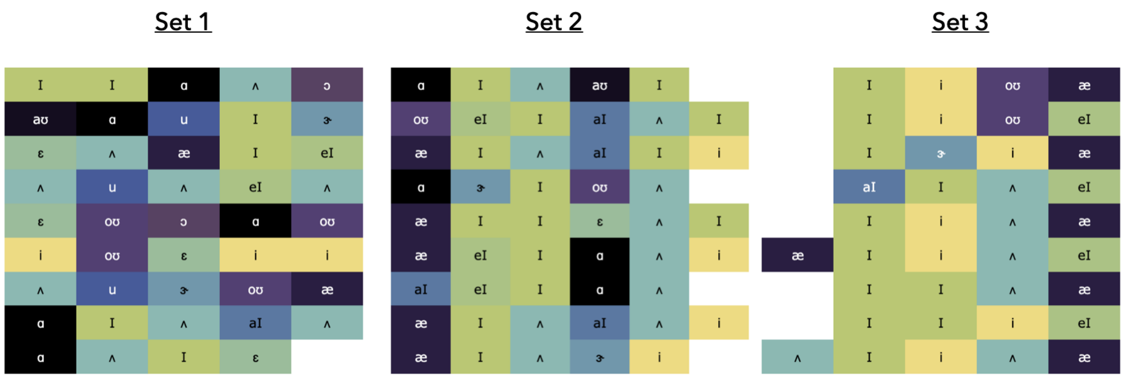


Figure 1.6: Note that the colors used to represent vowels are intended to provide useful intuitions about vowel similarity so that more similar vowels are represented by more similar colors (More on this in Chapter 2). For example, even though all the sounds in the leftmost column of set 2 are not identical, they do come from the same (darker) region in the color spectrum, which means the vowel sounds they represent also have more similarities with each other.

Making these patterns visible, as in Figure 6, reveals important intuitions about what kind of data are under consideration. One might expect that something about the vowels in Set 3 is bound to be similar; after all, those words are collected from a rhyming game. But visualizing the vowels is a different matter. Seeing them immediately makes clear that some vowel positions or columns are more stable, some can vary (Col 3 & 4), and some can alternate (right-most Col.). One of the issues I address is this: how can tools from computation, linguistics, and cognitive science be used to better describe the variety of structures found within sets of utterances like these?

I will discuss these examples in more detail in Chapter 6 & 7, but notice, for example, that none of these sequences (rows) of vowels in any of these sets repeat exactly. It would be difficult (perhaps impossible) to reasonably capture the structure and systematic variation present in the vowels of Sets 2 & 3 by using only the traditional static definitions of poetic devices. For instance, using a static way to define category membership (e.g. a condition where sequences must have identical vowels) one would find zero matches in any of these sets, which would not do justice to how plainly similar some of them are. Yes, their vowel sequences are unique, but still self-similar enough to be perceived as confusable (as in Set 2), or as rhymes (as in

Set 3). These visualizations make it clear that the vowels in Sets 2 & 3, respectively, form larger structures. Despite internal differences, their phrases seem to adhere to similar constraints, functionally becoming members of the same category.

So, if conventional poetic categories cannot capture these patterns well, how should these artifacts of lyrical language be treated? Ludwig Wittgenstein argued against Aristotle's suggestion that category membership is determined by static definitions. Wittgenstein suggested that some things, like games or music, do not have a static definition, or set of definitions, that can fully describe them. Instead, he claimed that category membership is captured by family resemblance [89]. He specifically discussed "language games", something that has been explored in terms of "semantical games" [90], but has yet to be explored in terms of "phonological games". Eleanor Rosch describes family resemblance as the ability for something to be more or less a category member, rather than the Aristotelian "all or none" options [91, 89]. In other words, category membership can be fuzzy (e.g. does this rhyme?), and members of the same category may have dramatic or subtle surface level differences.

Embracing fuzzy category boundaries in this space leads to many productive lines of inquiry that were not possible with only static and categorical definitions of poetic forms. Fuzzy category boundaries can be discovered by examining sets of phrases that humans have already grouped together due to (perceptual or phonological) similarity. Instead of being stymied because data from the wild does not match the current terminology, the rich and imperfect data that comes from rhyme sets can be used to reveal phonological structures, perceptual boundaries, and artistic style.

1.7.1 Rhyme Sets & More

In a seminal, but neglected paper from 1974 it was noted that some poets not only utilize rhyming couplets, but also use as many as 30 rhymes in a single pattern [92]. Shaw calls these structures *large rhyme sets* where each member (rhyme) is a rhyme partner with all other members. He presents linguistic methods for distinguishing nearby rhyme sets (in Russian) and argues for the validity of rhyme sets as a structure - even suggesting that "accepting these large rhyme groups as sets can aid in understanding and appreciating ... poems." Indeed, This generic structure allows rhyming data to be easily treated as systems with fuzzy categories rather than merely as collections of pairwise similarity relations, as is traditionally done. Pairwise similarity relations (e.g. identity, graded similarity) are important to observe, but they focus only on the local relations and do not consider broader category membership and changes over time. In contrast, a more systems level approach, enabled by rhyme sets, allows for describing the underlying structure and dynamics of these data.

In order to uncover sound structure in sets of utterances like rhymes, the basic idea of *large rhyme sets* introduced by Shaw [92] can be further developed. Whereas Shaw focused on large rhyme sets composed primarily of single syllable end-rhyme (mostly masculine and feminine), modern rhyme sets (especially in hip-hop) are not only repeated many times, but also span many syllables. Instead of representing rhymes in simple 1-dimensional space (ABABAB), which only captures *that* rhymes occurred, these data can be represented in 2 dimensions to examine *where* and *which* variations occur. For example, to make computation straight-forward, their underlying sounds can be vertically syllable aligned, much as was done above in Figure 1.5 and 1.6.

Rhyme sets, once identified, provide a powerful way to discover important positional and featural variation within perceived category members. That is, rhyme sets can be used to reveal reliable, shared, and changing sound structures. They allow us to assume a 'family resemblance' type of category membership (based on human attestations) and then systematically explore variation within these category members. Collecting and analyzing rhyme sets can serve an important role in understanding complex poetic structures themselves, as well as for investigations into phonology, perception, categorization, and language-based-music. Moreover, rhyme sets need not be narrowly 'rhyme' specific, but rather may be comprised of words that have alliterative, assonance, consonance, or confusability relations. These sets may be refereed to as alliterative sets, assonance sets, consonance sets or confusable sets – and they are similar to rhyme sets in that they provide a structure to investigate imperfection and family resemblance within poetic devices or other language.

Rhyme sets can be ordered or unordered. When ordered, rhyme sets are collected from naturalistic settings, such as lyrics. Their specific sequencing also permits them to be interpreted as time series data and represented dynamical systems where transitions between elements and clusters of elements, within and across rhyming words, can be used to describe their repeated sound patterns.

For example, Set 1 in Figure 1.4 represents random phrases, and since the ordering of these utterances is unimportant (though it might be important in some contexts), I will refer to this type as an unordered random set. Similarly, Set 2 in Figure 1.4 represents confusable phrases from a meme, and since order does not seem relevant to the effect (though it might be), it is treated as an unordered confusable set. Finally, Set 3 in Figure 1.4 represents ordered rhyming utterances from a turn-taking game. Since the ordering of these utterances is important, I will refer to it as an ordered rhyme set.

Within the rhyme scheme of any work, there many be more than 2 words associated with a letter representation (ABABAB). In Figure 1.7 Annotated Rhymes there are 3 As and 3 Bs. All the As together are considered a *rhyme set*, as are all the Bs.

Rhyme sets can be collected in various ways and may have additional properties discussed in more detail in section 3 of Chapter 4, as well as in Chapter 6 Section 1.

One can also generate unordered rhyme (or other) sets such that any given components can be artificially held constant (e.g. set of all words with vowel pattern /i/-/i/; set of all words with stress pattern 1010, all perfect masculine rhymes with

Figure 1.7: Example annotated rhymes (ABABAB) and related ordered and unordered rhyme sets.

'bins'). This enables exploring variation across all segments not held constant as well as properties of words in the set (part of speech, meaning, etc..). Unordered rhyme sets of this kind can be comprehensively generated using corpora and phonetic transcriptions. I also provide a simple related tool for automatically generating sets within a software toolkit called 'phonesse'. It is discussed further in Chapter 2.

1.8 Phonological Vocabulary

There is a foundational question about how to frame the variety of phonological structures and forms that I will highlight in this dissertation. Larger, long range sound patterns (rhyme, etc..) are composed of smaller patterns, and so it is important that any framing is coherent across scales (e.g. segments, syllables, multi-syllable rhyme). In addition, a focus on the cognitive underpinnings of these forms may be enriched by a framing that is easily interpretable in terms of learning. Here, I use the term "phonological vocabulary" to refer to the body of sound patterns (sound-sound relations) used in a particular sphere, either by a language or an individual. Below is a graphic that highlights the difference between a conventional (lexical) vocabulary and a phonological vocabulary, as defined. Figure 1.8 also includes applications of that definition within the domains of both language and language-based music.

But what is the scope of the sound patterns included in a phonological vocabulary?

Objects of the phonological vocabulary can be identified and documented at various scales, along with their frequency of use. These structures may correspond to lexical objects, but may also exist above or below that scale. Indeed, as shown in Figure 1.9, elements of the phonological vocabulary may include objects across at least three scales (e.g. segments, syllables, rhyming). Much as the orthographic elements of words (e.g. letters), or the combinations of words (e.g. collocations, phrases) can be

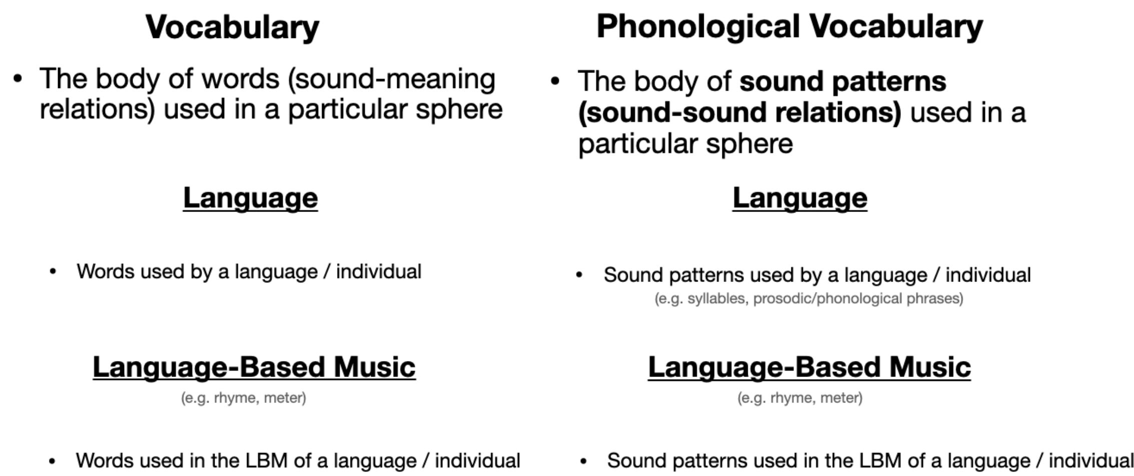


Figure 1.8: Definition of vocabulary and phonological vocabulary - applied to both language and language-based music contexts.

informative for the study of lexical vocabulary, in the same way segments, syllables, and higher-order poetic structures like rhyme are instructive for shedding light on the phonological organization within a language or individual. Framing these elements as a hierarchical system of vocabulary objects allows for productively documenting and grounding their interrelations within the phonological system.

The term ”phonological vocabulary” has been used in a previous 2006 paper [93] to refer to the phonological knowledge, compared to the orthographic knowledge, of second language learners. They probed for what they called ”phonological vocabulary” by implementing vocabulary tests where word stimuli were heard rather than seen. The hypothesis was that tests of orthographic knowledge may not well capture ’phonological’ (i.e. aural) knowledge of words. Furthermore, orthographic based tests may be differentially inaccurate for second-language learners of different backgrounds. For instance, orthographic tests of Arabic speakers learning English (two languages that use very different orthographic systems) may underestimate the learner’s phonological (aural) knowledge of English words. Since then, this literature has shifted its terminology from comparing ”phonological and orthographic vocabulary” to ”aural and written vocabulary”, terms which seem more appropriate given the methodological focus on heard verses seen lexical items.

Although the informing idea of ”phonological vocabulary” from Milton & Hopkins as representative of phonological knowledge is consistent what what I propose here, I broaden its scope significantly. The notion of phonological vocabulary I develop includes interaction with lexical items, but is by no means limited to that domain. Figure 1.9 highlights the objects and relations of phonological vocabulary. At the smaller level, phonemes and suprasegmentals interact and combine (at short range) to form larger elements such as syllables, consonant clusters and vowel sequences – phonotactic arrangements that are known to be influenced by speech production.

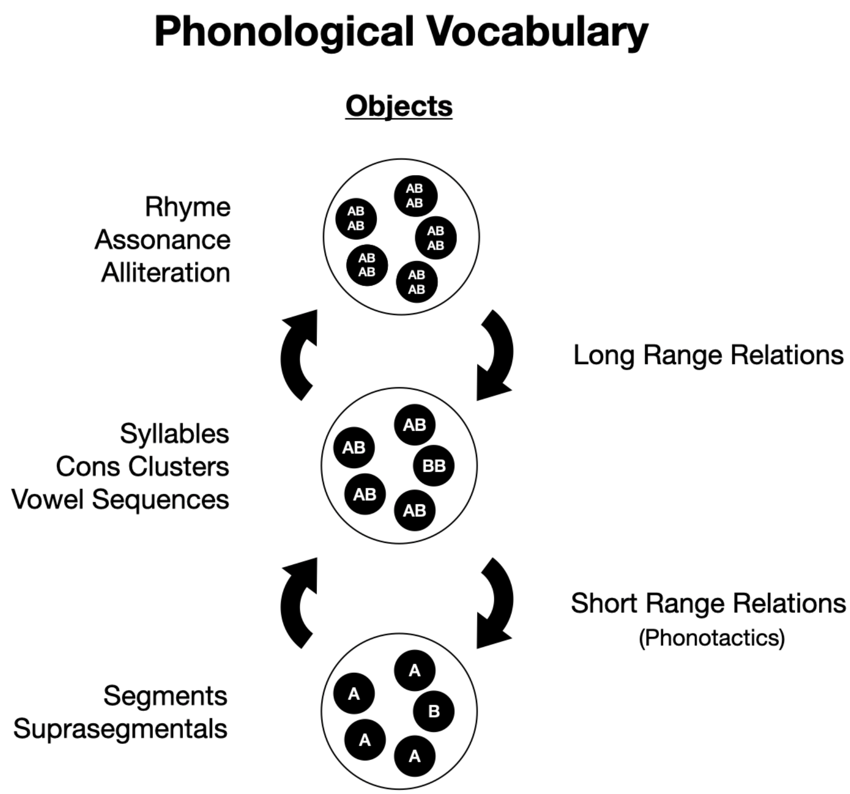


Figure 1.9: Examples of the Objects and Relations of Phonological Vocabulary.

In turn, syllables, consonant clusters, and vowel sequences interact and combine (at longer range) to form larger objects like rhyme, consonant, or assonance – forms that are influenced by the rules of language-based music.

While the phonotactics of language are well-studied, the dynamics of rhyme and language-based music, more generally, are much less well-understood. Some named forms of rhyme and meter have been the focus of various studies in poetics and linguistics, but have not been studied comprehensively or at scale. Other forms have scarcely been documented (e.g. imperfect multi-syllabic or covert).

An additional and relevant definition of the word 'vocabulary' also applies here, "a range of artistic or stylistic forms, techniques, or movements" [94], like those commonly found across creative domains, from music to dance to sports. In addition to treating various sound sequences as part of the vocabulary, emergent components of the phonological vocabulary (e.g. rhyme, alliteration) can themselves be seen as stylistic or artistic forms or techniques.

It is important to stress that the phonological vocabulary, as I have proposed here, is not meant to be a cognitive or linguistic explanation of any phenomenon. Rather, much like any vocabulary, it is meant to provide practical objects of study that can be used in order to develop future claims or hypotheses in linguistics, poetics, psychology, or cognitive science. The more formal identification of vocabulary items (language or individual) may ultimately be used to better understand the learnability of phonological sequences and distributions [95].

Intuitions for Exploring Phonological Vocabulary

In this dissertation, I investigate elements of the phonological vocabulary of language-based music in various ways, but largely focus on computational studies that explore a range of structures and trends from orthographic data. However, there are many other ways to approach and uncover the elements, features, and usage of phonological vocabulary.

As a toy example, imagine that all vowel sequences generated by an individual, up to 10 vowels (an arbitrary cutoff), are recorded as a (vowel) vocabulary. This limited data would seem to capture many important aspects of their phonological vowel vocabulary, though identifying the regular or salient patterns within these data (1-10 ngrams of vowels) is a different matter. If this individual's vowel n-grams were grouped by the register that was used to produce it (speech, writing, song lyrics, rap lyrics), one may expect to see differences in the distributions of vowel sequences across these domains. These observed differences in phonological (vowel) vocabulary can then be hypothesized about in terms of cognitive or linguistic explanations.

Of course, many phonological patterns are overt, and so a focus on the phonological vocabulary used within an individual's (or a language's) overt patterns (e.g. words, rhymes, syllables) in comparison to their covert pattern is also revealing of the underlying phonological vocabulary, and ultimately, the underlying phonological system.

One could also take a more behavioral approach to understanding the phonological system by targeting the perception or production of phonological vocabulary.

For instance, In a simple speech production task, subjects can be tasked with producing words consecutively for X minutes where the only condition is that each next word have at least one phoneme (sound) in common with the previous one. One would expect to see explore-exploit patterns of rhyme, alliteration, assonance, and consonance that differ across individuals. These differences would seem to be related to both the lexical and phonological networks of individuals (or languages). Documenting the phonological vocabulary or vocabulary properties from studies like this can also enable efforts at reconstructing the phonological network from the vocabulary (as is often done in lexical or semantic space).

On the other hand, perceptual tasks can also be used to uncover differences in the size or complexity of an individual's phonological vocabulary. For example, subjects can be given various language samples (e.g. poetry, rap, lyrics, etc..) and then be tasked to annotate all of the sound patterns (or poetic devices) they notice within a given duration of time (instructions may vary). It may be expected that differences in the count or complexity of identified patterns across individuals correlate strongly with their recorded phonological vocabulary at multiple scales. One might also expect that those with learned phonological skills (e.g. writers, poets, lyricists, rappers) and larger phonological vocabularies may attest to more or bigger or different phonological structures. Not only would this tell us whether phonological vocabulary is productive in predicting the perception of phonological patterns, but also, it could highlight the perceptual differences across certain kinds of phonological expertise, much like has been done in music with the study of musical expertise.

Throughout this dissertation, I frame the study of phonological vocabulary in language-based music at various levels:

- Landscape of possible perfect poetic devices (in dictionary)
- Components of overt perfect rhyme (in poetry)
- Predictability of overt multi-syllable rhyme sets (in rap)
- Predictability of phonemes sequences (in music)
- Usage of phonemes (in LBM and non-LBM genres)
- Usage of phonemes and rhyme sets (in rap)
- Changes in predictability (over time) of vowel and word sequences (in rap)

1.8.1 Chapter Summaries

The organization of my inquiry is as follows:

In Chapter 1, I introduced language-based music and gave intuitions for understanding this space. I began with a brief review of repeated verbal sound patterns in human culture and science. Then, I outlined an interdisciplinary approach to investigating this phenomenon, and why we should care about both computational and cognitive framings of it.

In Chapter 2, I introduce data and a suite of tools for manipulating and visualizing sounds drawn from orthographic data. Then. I develop an approach to sound visualization that pulls from sound symbolism and color research. These resources can be used to replicate the results found in this dissertation, or for other efforts in the language sciences.

Chapter 3 presents a literature review to motivate intuitions about the cognition and computation of language-based music. Here, I elaborate on the cognitive mechanisms that support verbal sound patterning, and why they are important to take into account when dealing with related data. Then, I review the computational and linguistic approaches used so far in characterizing rhyme and sound patterning in creative language.

In Chapter 4, I focus on data that are identified to belong to poetic devices. First, in order to better understand the landscape, I document all possible instances of perfect poetic devices across all dictionary words. Then, I transition to considering imperfect rhymes across 500 years of poetry in English. Finally, I turn my focus to more formal analysis, examining overt multi-syllable rhyme sets, something that has not yet been seriously considered in the literature.

In Chapter 5, I quantify phoneme frequencies across disparate corpora to explore cognitive and task effects on phonological production. Like much text data, these are untagged corpora where the location of poetic devices is not known. I characterize the distribution of phonemes across genres and compare samples from different creative genres and speakers against baseline samples drawn from conventional language. I begin by documenting phoneme n-gram frequencies across corpora. Then, I use these frequencies as features in supervised and unsupervised machine learning models to classify texts based on genre.

In Chapter 6, I develop case studies of improvised rhyming (rapping) samples using methods from both Chapters 4 and 5. First, I introduce and analyze data from a rhyming game tournament (Bar Pong). Then, I consider the phonological differences between written and improvised verses in a rap cipher. Finally, I look at samples of improvised rappers, using longitudinal data to uncover changes in phonological and lexical vocabulary, as well as in semantic similarity.

In Chapter 7, I summarize the main contributions of this dissertation and what they tell us about language-based music. I also discuss prospective applications for this line of research, including relations to literacy programs, pedagogy, and artificial intelligence.

Chapter 2

Tools & Data

In this chapter, I both introduce and review core elements of language sound. First, I briefly discuss how verbal sounds are associable with real world qualia such as color. Then, I discuss how sound symbolism can be operationalized intuitively to visualize sound patterns in language. Finally, I introduce relevant tools and data I have developed to investigate language-based music, and how they can be used to understand this phenomenon at scale.

Questions Covered:

- (Section 2.1) What are the elements of language-based music?
- (Section 2.2) How can language sounds be explored at scale?
- (Section 2.3) How can language sounds be visualized?

2.1 The Elements

Speech sounds are a continuous signal that can be represented by amplitude over time, where oscillating changes in amplitude are captured by frequency (or frequencies). Some particular portions of sound signals appear to be identifiable as discrete units, like musical notes, animal cries, or spoken sounds. When deciding how to represent many phenomena, including language, humans often represent complex concepts in simple and discrete terms. Representations of natural language utterances can be orthographic, phonological, acoustic, etc (Figure 2.1). Here, I will be focusing on phonemic representations, which can be analyzed in terms of natural classes and distinctive features.

This sort of simplification can be convenient, but it may sacrifice some important

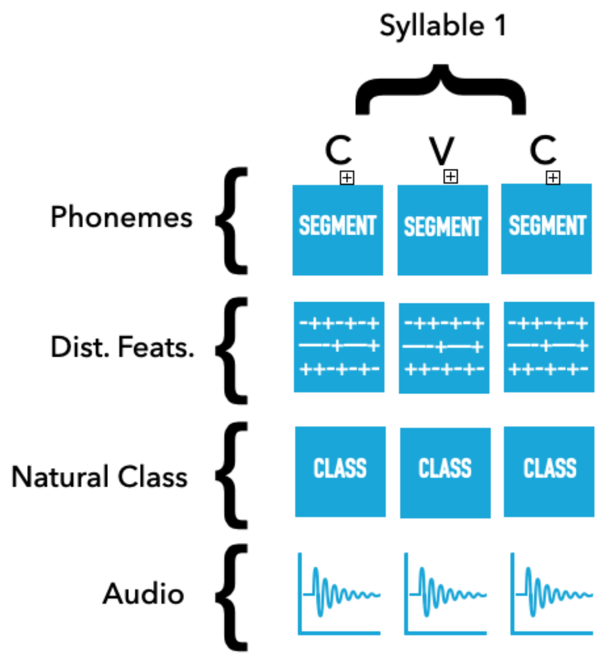


Figure 2.1: 4 Levels of Representation of a Syllable.

dimensions of resolution. For example, there is important variability in the continuous signals that make up spoken sound across languages, dialects, genders, and speakers [96, 97], even if the written words used to represent them are the same. However, humans typically sacrifice fine discriminations and the details of raw frequency and amplitude, in order to refer to sections of the signal in practical ways e.g. sentences, words, letters, phonemes, segments. These units and their differences have provided important insights, though some of the categorical distinctions in linguistic analysis may be further enriched by understanding more about continuous representations.

There is a long standing debate over the psychological reality of categorical sound distinctions like syllables, phonemes, and sound segments. Beginning in the early twentieth century it was suggested that phonemes and sound segments are not simply a unit of descriptive and theoretical analysis for linguistics, but that they may have some kind of psychological status [98, 99, 100]. In contrast, there have been researchers who are suspicious of imputing psychological status to linguistic constructs [101], asserting that categories like phonemes or segments are merely "convenient fictions" [102, 103]. Although discussion continues concerning the psychological grounding of categorical boundaries in language, there is some evidence that phonemes may exist in our minds, though we lack awareness of them without certain abilities like alphabetic literacy [99, 104, 105].

Nevertheless, categories of sound (syllables, segments, phonemes, natural classes, distinctive features) can be perceptually and analytically useful. The present investigation mainly utilizes the representations of sound patterns built on categorical compo-

Category	Definition	Example
Syllable	A unit of pronunciation having one nucleus (usually a vowel sound), with or without surrounding consonants (onset-before the nucleus; coda-after the nucleus), forming a whole word or part of a word.	Dogs - (Dogs) Table - (ta)(ble) Surprise - (sur)(prise)
Segment	A discrete unit, either physical or auditory, in the stream of speech. Can be phoneme, mora, syllable, prosodic unit, morpheme. Usually refers to a single phone in phonetics and phonology.	[t], [b], /t/, (sur), Dog
Phone	Any distinct speech sound or gesture, regardless of whether the exact sound is critical to the meanings of words. Represents the sounds themselves	Marked with [] e.g. [b], [p]
Phoneme	The smallest unit that distinguishes meaning between sounds in a given language. In contrast to a phone, changing any phoneme in a word can change the word. Represents the mental representation of sounds.	Marked with // e.g. /b/, /p/
Natural Class	A set of phonemes that share certain distinctive features. A natural class is determined by participation in shared phonological processes (articulatory and acoustic properties), described using the minimum number of features necessary for disambiguation.	Voiceless stops (/p/, /t/, /k/) Voiced stops (/b/, /d/, /g/), Voiceless fricatives (/f/, /θ/, /s/, /ʃ/, and /h/)
Distinctive Feature	Distinctive features are grouped into categories according to the natural classes of segments they describe: major class features, laryngeal features, manner features, and place features. Segments either have have the feature [+], don't have the feature [-], or are unmarked [] with respect to it.	Syllabic, consonantal, approximant, sonorant, nasal, lateral, labial, etc…

Figure 2.2: Chart of relevant linguistic categorical units related to sound.

nents of syllables, their vowel and consonant segments, and the distinctive features that comprise those segments. Figure 2.2 shows an example subset of the discretized segments of spoken language (and their representations).

The pronunciations are represented using the International Phonetic Alphabet (IPA). The IPA is a universal alphabet for representing phoneme segments. Sound segments fall into two large categories, vowels and consonants. In either case, the segments above can be decomposed into smaller components and uniquely differentiated from each other by means of a set of attributes called distinctive features. Figure 2.4 on the left, shows the canonical vowel chart (modeled after the shape of the human mouth), which plots vowels in terms of their highness and backness dimensions. On the right, Figure 2.4 displays the distinctive features of English vowels, providing a way to understand and operationalize the dimensions by which each vowel sound is categorically similar to the others.

While these are merely descriptive tools, there are attempts to provide theoretical accounts for how and why sound elements can get organized in this way. For example, studies of the Brownian dynamics of vowel change (random change due to environmental conditions) over generations suggest that dispersion in vowel systems leads to maximized contrasts between sounds. On the other hand, quantal theory suggests that there are non-linear relations between articulatory and acoustic components [106].

Categories such as phonemes in language are often explained as cultural attractors, systematic biases in a population that describe how instances can become part of

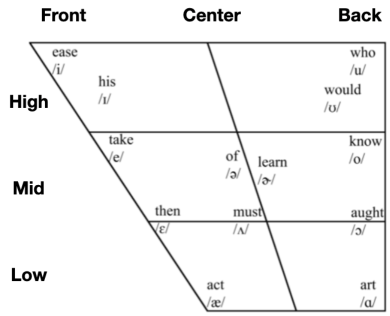
ARPABET	IPA	Example
IY	i	beat
IH	ɪ	bit
EY	eɪ	bait
EH	ɛ	bet
AH	ʌ	butt
ER	ɝ	bird
AY	aɪ	bite
UW	u	boot
UH	ʊ	book
OW	oʊ	boat
AO	ɔ	story
AE	æ	bat
OY	ɔɪ	boy
AW	aʊ	bout
AA	ɑ	balm, bot

Figure 2.3: A chart of International Phonetic Alphabet symbols for sounds, their ARPAbet encodings, words that use those sounds, and those words encoded in IPA or ARPABET representations. This graphic focuses on a subset of vowel sounds [1]

a group, and that influence how the learning environment. While there have been many efforts to describe the impact cultural attractors on language development (vowel categories in this case), attractors are usually asserted, rather than emergent. Recent work [107] presents a computational model that demonstrates how attractors themselves may have emerged (phoneme categories). Regardless of their complex and evolutionary trajectory, the simple categories shown above are the standard elements used in most analysis of language sounds.

As mentioned, discrete representations of sound items in language are useful, but are problematic for a number of reasons beyond their simplification of the underlying continuous signals. In 2001, George Miller noted that the ambiguity of the dictionary meanings of words creates a combinatorial semantic explosion which makes the human ability to interpret meaning in language quite remarkable. In particular, individual words have many meanings, and even within Standard American English, orthographic representations leave us with too many possibilities to be certain of the 'true' or intended representation. Miller's approach shown in Figure 2.5 is used to calculate how many distinct possible semantic interpretations there are of a given sentence when using simple (categorical) lookup dictionaries [108]. This ambiguity also applies to the representation of pronunciation.

I have conservatively conducted the same exercise with all possible pronunciations of these same words from the CMU pronouncing dictionary of Standard American English to identify 72 possible and distinct pronunciations of this sentence. This highlights the need to acknowledge uncertainty in the faithfulness of discrete rep-



Vowels and their Distinctive Features

	i	ɪ	u	ʊ	e	ɛ	o	ʌ	ɑ	a	ɔ	ɒ
[back]	-	-	+	+	-	-	+	+	+	-	+	+
[high]	+	+	+	+	-	-	-	-	-	-	-	-
[low]	-	-	-	-	-	-	-	-	+	+	+	+
[round]	-	-	+	+	-	-	+	-	-	-	+	+
[tense]	+	-	+	-	+	-	+	-	+	-	+	-

Figure 2.4: A vowel chart (left) [2, 3], and a chart of the distinctive features of vowel sounds (right). Although roundness and tense are not displayed in 6A (left image) they are critical to uniquely identifying vowel sounds, and so must be included in any complete distinctive feature chart (as on the right) [4]

Words	But	I	have	promises	to	keep	and	miles	to	go	before	I	sleep	**Total**
Meaning	11	3	16	7	21	17	5	5	21	29	10	3	6	**~3.6 Trillion**
Pron.	1	1	1	1	3	1	2	2	3	1	2	1	1	**72**

Figure 2.5: Miller multiplies the number of possible meanings of each word in a sentence. 11x3x16. . . x6 resulting in 3.6 trillion possible meanings of the above sentence.

resentations of sounds, even in simple or obvious-seeming utterances. The level of ambiguity in pronunciation is not as great as that of meaning, but it does still exist. This is not even taking into account differences in pronunciations of dialects or individual speakers. Here, uncertainty is an issue that is ever present, particularly when conducting analyses not based on (or grounded by) audio recordings of speakers. Most analyses of rhymes, the present study included, fall victim to this problem, as they are derived from orthographic representations alone. Despite this, as we will see, a great deal of progress can still be made by studying sound patterns extracted from the orthographic representations of words.

2.2 Tools

2.2.1 Phonesse

Given the approach of transcribing orthographic text into IPA representations, a convenient and reliable way to process and handle both individual samples as well as a large corpora of orthographic text is needed. Phonesse is a python package I developed for this purpose. It is built around the CMU Pronouncing Dictionary and offers some convenient features for encoding and manipulating phonemic information. Some of its basic utilities are discussed below.

Popular pronunciation lookup dictionaries like CMUdict or Celex offer a starting point for encoding the sounds of language from orthographic text. They provide simple lookup tables where individual words are paired with their IPA representation, often in ARPABET (computer friendly) format. A word like 'begin' can be represented in IPA as /bigın/ and in a more computer friendly format called ARPABET ('B IY0 G IH1 N'). These symbols serially encode the consonants, vowels, and stress pattern of the word 'begin'. They are written in a structured way, and so these elements can be reliably parsed and separated from each other, but that is usually done by each researcher individually.

It would be useful to be able to ask for different subsets of these serialized constituent segments, for example, only the vowel or only the stress elements, entire syllables, or just the rime elements. Accordingly, I present a brief introduction to the Phonesse package which provides a simple API for collecting, manipulating, and visualizing sound elements from strings of orthographic text. Some examples are shown below, while full documentation can be found at jordanmasters.github.io/phonomials/

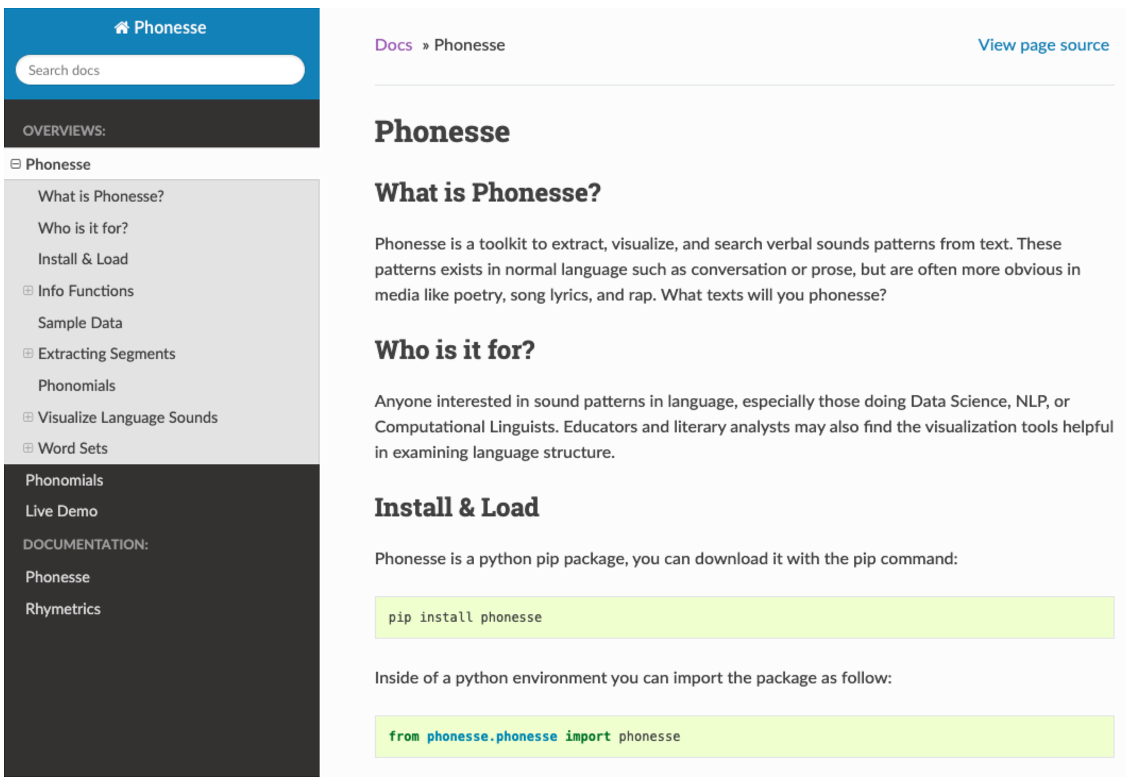


Figure 2.6: Phonesse User Documentation: jordanmasters.github.io/phonomials/

Pronunciations

In both speech and lookup dictionaries, words may have alternate pronunciations. Once a phonesse object is encoded it likewise provides possible pronunciations for each word.

Because of uncertainty about which pronunciation is actually uttered, the first (most common) pronunciation for each word is used, though this can introduce error. These encodings can be manually adjusted to more accurately reflect the specific spoken utterance.

There are also many words that are not included in a lookup dictionary like CMUdict, or that are misspelled, or have alternate spellings. In order to include words that orthographic lookup-tables fail to represent, I incorporate a grapheme-2-phoneme neural network model [109] into Phonesse. This way, when requesting the IPA encoding of some unusual or malformed English, a reasonable representation of the underlying sounds will be returned. This is particularly relevant for hip-hop where fast changing slang and abbreviations are often not present in static dictionaries.

As I have mentioned, transcribing from orthography introduces some error to the encoding. In order to circumvent this, the use of either human annotators or machine learning (to extract phonemes from audio) can be employed. Both of these approaches also can introduce some error, but are more closely grounded to the actual phone uttered. In many cases, these more stringent validations of the data encodings should ultimately be used to verify that the conclusions of work based on automatic orthography-¿text encodings are not spurious (present work included). However, for the current investigation I will proceed with automatic transcriptions as described.

Finally, Phonesse provides various other functions that facilitate processing related to natural class, distinctive features, n-grams, sound pattern search interface, and visualization.

Natural Classes

Show all phonemes that belong to a given natural class. Show all natural classes associated with a given phoneme.

N-grams

Extract all phoneme sequences & their frequencies, up to a given n-gram size.

Search Widget

Create a sound pattern and retrieve all words that contain that pattern. This is useful as a tool to explore sound patterns in the lexicon, and as a way to design lexical stimuli with particular phonological properties.

Specify a number of syllables for the phonesse 'search app' and it will generate an interface with appropriate vowel and consonant positions. Then populate each interactive position with a sound segment. Finally, click 'Search' to get words that match the created pattern.

2.3 Visualizations

Encoding relevant data is the first step, but it is also useful to sanity check and visualize what has just been encoded. Dan Levitin has suggested that "music can be thought of as a type of perceptual illusion in which our brain imposes structure and order on a sequence of sounds" [110]. Sometimes we can even see this larger structure in musical notation or in other popular visualization tools like MIDI. Verbal sound patterning may leverage a similar type of perceptual illusion, but built on top of, and often, smuggled into language. This is much like singing, but using the phonemes of language, rather than pitch, to create patterns. It can be hard to notice verbal sound patterns when phonology is an aspect of language that is often taken for granted (always heard, but rarely consciously attended to). Listeners only hear one sound at a time and don't get to see the bigger picture. Unlike in music, where notation systems and representations make appreciating, digesting, and engineering, in the language arts many patterns go unnoticed. Visualizing language sounds in perceptually valid colors and spatial configurations, as I have begun to demonstrate above, can help our brains impose more structure and order on a sequence of sounds, serving as a starting point for more rigorous investigation. As it turns out, the same representations (e.g. lists, grids, matrices) that are useful for visualizing sounds and sequences, are also amenable to computational and linguistic analysis.

Together, these tools enable more scalable and accessible processing of language sounds from orthographic samples.

2.3.1 Phonemes-2-Colors

Although the descriptive tools of linguistics are integral to this work, there are ways to visually represent sounds that are more psychologically and perceptually grounded. Language sounds can be associated with physical properties in the world (weight, shape, size, color, etc.). The study of this phenomenon is known as sound symbolism. In addition, the study of color associations and categorization is an independent and growing field in psychology. In order to enhance the interpretability of verbal sound visualization, these two domains can be joined to create a perceptually driven vowel-2-color map.

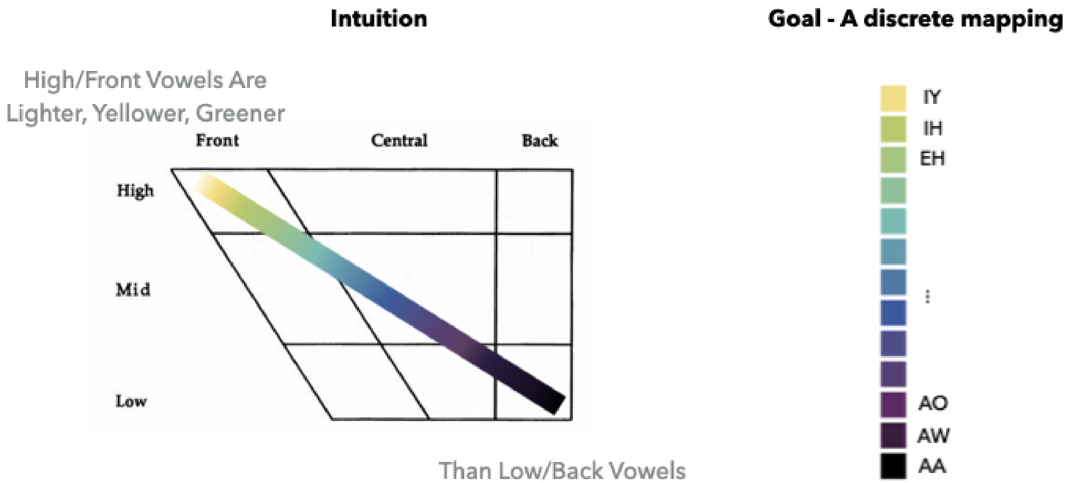


Figure 2.7: Left: Assuming the finding from Cuskley, a color scale can be superimposed on the 2-dimenional vowel grid to give a simplified instinct for the problem space. Right: Example mapping between colors and vowel sounds

Work by Cuskley et al. (2019) [111] shows that humans seem to reliably associate certain high-front vowels with lighter, yellower, greener colors than low-back vowels. Given this as a starting point, a representation of vowels, such as the vowel chart below, can be related to a popular color scale, like Viridis, that adheres to the theorized light-yellow to dark spectrum. This graphic gives an intuitive way to imagine how one might associate vowel and color information.

Vowels are not 1 or even 2-dimensional, so in order to obtain robust associations, more sophisticated methods should be used.

In collaboration with Karen B. Schloss at the University of Wisconsin, I have developed a vowel-to-color map with cognitively intuitive colors, where more similar colors represent more similar vowels. In this ongoing research, we conducted a pilot study among students from University of Wisconsin Michigan. All combinations of vowels pairs are presented to subjects who make judgements about their perceived position on a color scale. Specifically, for each each trial, a pair of audio clips is played and then the subject is tasked with deciding which vowel sound seemed 'higher' on the color scale. In this case, the color scale was presented vertically, higher was lighter yellow-green, lower was darker purple (e.g. 'Which sound is higher on this color scale?'). In the results below, the color scale is presented horizontally for convenience. Each subject (n=18) is exposed to all combinations of vowels. These pairwise judgments are then used to identify a discrete ordering of vowels in relation to the considered color scale. Additionally, subjects were checked for color-blindness with a vision test known as the Ishihara Test. 3 subjects were removed because they got 2 or more of the 11 Ishihara test questions wrong. The final number of subjects for this pilot study was 18.

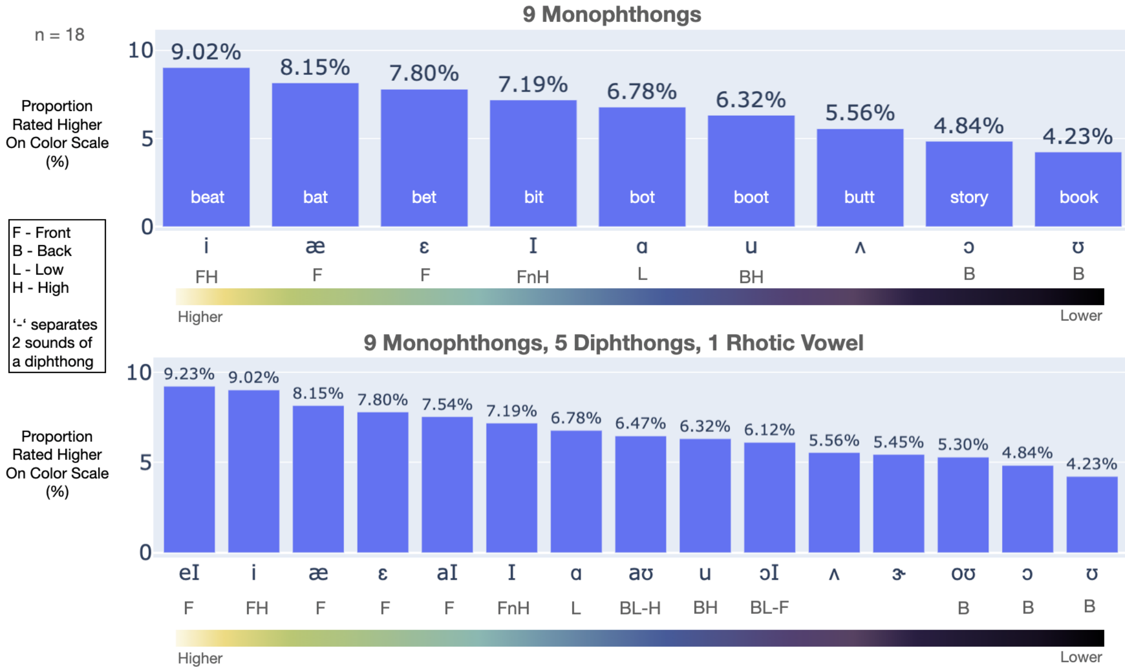


Figure 2.8: Proportion of times each vowel was judged as higher on the color scale. Top: Across all pairs of (9) monophthongs vowels. Bottom: Across all pairs of (15) vowels from CMU pronouncing dictionary .

For instance, the proportion of times each monophthong vowel is rated higher than all other vowels is shown in Figure 2.8 below. Each bar represents a single vowel phoneme, along with its corresponding front, high, low, or back relation.

This vowel set is similar to the one used in Cuskley et al. and it is clear that the Front and Front / High vowels are proportionally rated higher than Low or Back vowels. Generally this seems to correspond to the Cuskley et al. results. The congruence with previous literature gives more justification for extending this core set of vowels to a broader set for more practical use. For instance, the phoneme encoding used across much of this dissertation are based on the CMU pronouncing dictionary which contains 15 vowels, 9 monophthongs, 5 diphthongs, and a rhotic vowel (one with an 'r' sound in the nucleus). There are likely to be many viable mappings between sets of speech sounds and color, and these are only early steps in the process of developing more rich and practical cross-domain perceptual tools.

Phonesse uses the results from this work to provide vowel-2-color maps as a resource, and employs them when plotting various phoneme visualizations.

Robust vowel-2-color mappings may be useful for enhancing the interpretability of structure in language data. On the one hand, representations like the vowel grids used here can make some sound structures transparently obvious. On the other hand, vowel-color mappings may facilitate localized tracking of patterns or even a more global sense of structural complexity. Planned follow-up studies will compare entropic measures of sound complexity with user judgments of vowel grid complexity as the vowel-2-color mapping is varied. The related hypothesis is that psychologically based color-2-vowel maps will facilitate more accurate user judgements of structural complexity in sound grids. Experimentally driven color-2-vowel maps can be compared with inverse or random mappings in order to check whether these mappings offer benefits to holistic complexity judgements.

In addition, these maps may be utilized in color-coding vowel sounds from orthographic texts for language learning or literacy applications. This may be especially effective in languages where the mapping between orthography and pronunciation is not a simple one-2-one relation (e.g. Spelling-Pronunciation mappings in Spanish are largely one-2-one – this is not so for English). For instance the English letters 'oo' in 'boot' might be coded as one color, while the same 'oo' letters in 'book' might be coded as a darker color - reflecting the different underlying phoneme. This may aid in intuitively disambiguating pronunciations for learners by making clear that (and when) the same orthographic representations represent distinct pronunciations.

2.3.2 Phonemes-2-MIDI

Briefly expanding on the music analogy, MIDI outputs are common in digital music. Like the one below, they display notes over time for each pitch (row), resulting in a visualization of the chords, and melodies. These kinds of representations read like digital sheet music, where the x-axis is time, and each row (y-axis) is a specific musical pitch.

Figure 2.9: A traditional MIDI representation of pitch in music on the right paired with the pitches on the left (from a piano). [5]

We can also use MIDI representations to display the sounds of language. I again focus on vowels alone for representation as it is similar to the musical case (notes) for many reasons. First, there are a similar number of vowels in English and notes (pitches) in the western musical tradition ($\tilde{1}5$). Second, vowels are the most sonorant sounds and can exhibit rhythmic (stress) patterns much like those in music. Third, vowels are shown to be correlated with melody in musical lyrics, whereas consonants are not [112]. For these practical reasons, I only focus on encoding vowels in MIDI format, although other constituents can be encoded this way.

In the figure below, I combine these MIDI and color representations of vowels to visualize sounds from orthographic data. The chief difference from the musical case is that instead of rows representing unique notes or pitches, in a vowel MIDI, each row represents a unique vowel. Ordered from left to right, as I encode the vowels of any text, I print out a block for each next time-step (column) in the row that corresponds to the next vowel.

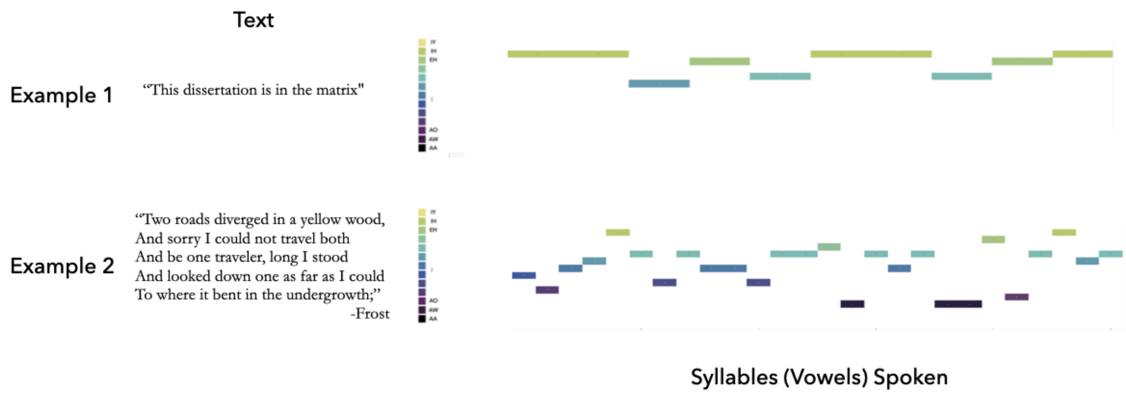


Figure 2.10: Colored MIDI representations of vowel phonemes from 2 short examples.

Figure 2.10 demonstrate this visualization concept on toy examples. Figure 2.11 show larger samples, one of which contains language-based music (Eminem's Lose Yourself).

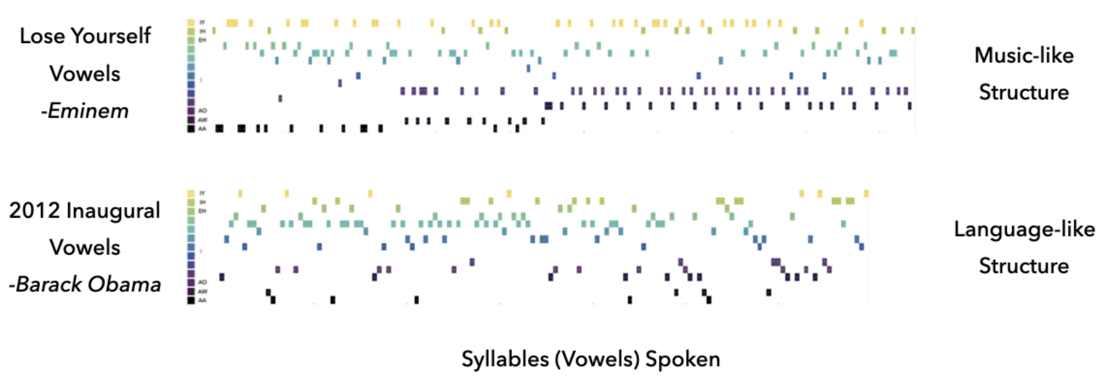


Figure 2.11: Colored MIDI representations of vowel phonemes from 2 longer examples.

As one might expect, the natural language of Obama's Inaugural does not look obviously patterned, though there is some structure. These data display clear trends, likely related to the frequency and phonotactics of sounds and their transitions in spoken English. This is largely meant as a baseline for comparison.

On the other hand, Eminem's Lose Yourself shows enormous amounts of periodicity and structure that are so regular they could easily be mistaken for beat or melody. Additionally, the patterns change, and even relations between patterns seem to have structure. This kind of complexity underlies a great deal of rhyming verse, however, humans do not ordinarily notice it in such a holistic way.

This approach can be used in many ways, but for current purposes, it is meant as an illustrative tool. Here, using MIDI visualization of vowels is a way to reveal the shape of the data under consideration, and to draw practical analogies to musical patterns.

But MIDI representations are not always the most revealing. This same rap verse can alternatively be plotted in a grid format using Phonesse (Figure 2.12).

Figure 2.12: Right-aligned lyrics and color-coded vowels from Eminem 'Lose Yourself' verse.

Notice how much easier it is to orient ourselves to large-scale vowel patterns when viewing Eminem's Lose Yourself verse in this way compared to MIDI (or orthography for that matter).

2.3.3 Phonemes-2-Grids

Even coded by color, some patterns can be quite difficult to see or interpret. As I just mentioned, another way to visualize the sounds of language is to assume a bit more structure and display our color-coded sounds in a grid. Text, IPA, and MIDI representations by default assume serialized data (1-dimensional). A grid, on the other hand, represents 2 spatial dimensions, a horizontal one, and a vertical one. Each row in the grid may represent their own meaningful unit of sound sequences (e.g.

sentences, lines, rhymes). These rows can then be right or left-aligned, according to the data in question.

Figure 2.13 displays examples of left-aligned and right-aligned vowels in color-coded grids.

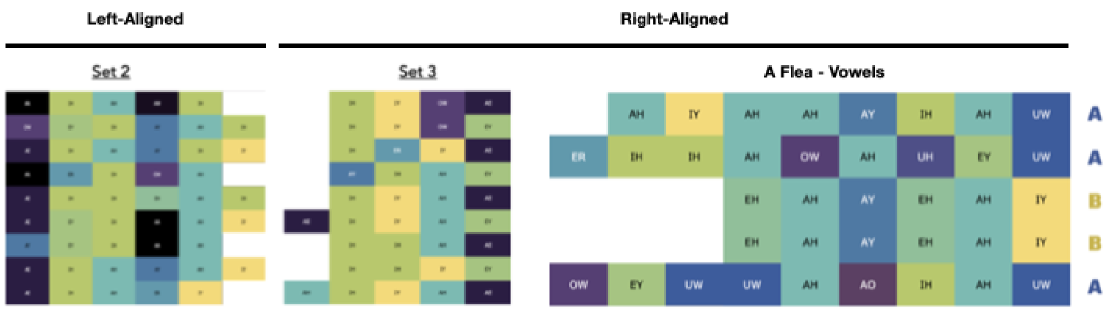


Figure 2.13: Vowel Grid Alignment, left and right

Use of such grounded and colored representations allows us to immediately observe certain qualitative differences between samples, qualitative differences that the presented tools also make much easier to quantify.

In Figures 2.14 and 2.15, I show the underlying vowels from a variety of contexts across the history. Many of the vowel structures you will see reflect imperfect rhymes. This is an enormous class that is not well understood. Here, I again focus on the vowel representation of syllables since they provide a reliable point of comparison over many types of rhyme and contexts. This is particularly relevant for imperfect rhymes, which tend to have a lot of variation in consonant positions, but are relatively more stable in their vowels (as will be seen in Chapter 4.3).

Seeing the patterns in comparison can make it clear how complex rhymes and rhyme schemes have become. In the next figure I again display only the vowels of four works, but break them down in a 2x2 display to give a sense of the different constraints and cognitive processes that can produce these forms.

Note that in 2.15 A and C, the sound patterns are predetermined, but the content (e.g. words, phrases, punchlines) are not. However, in 2.15 A, word matches for the predetermined sound pattern (ABAB...) are written (premeditated), while in C they are improvised in the moment during the context of a game. C represents the vowels of 42 4-syllable phrases that rhyme with the phrase, ‘Level 3 Vest’. In both B and D, the rhyme schemes are open-ended, but D was created improvisationally, while B was premeditated.

These structures demand attention as linguistic phenomena, but also rely on different cognitive processes. In Chapter 3, I synthesize the current literature on the aesthetics, learnability, and neuroscience of rhyme in order to build a foundation upon which the cognitive mechanisms of both simple, and eventually, more complex multi-syllabic

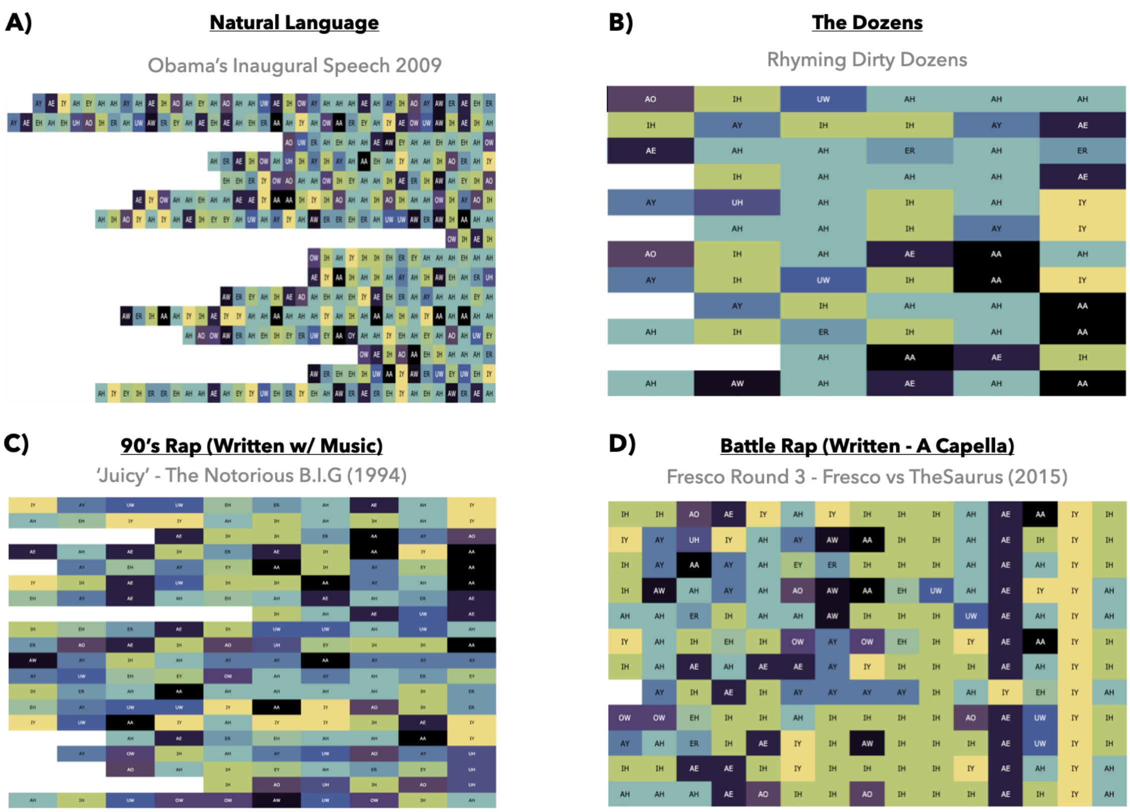


Figure 2.14: Examples of the underlying vowels (right-aligned) in four categories. Each row represents the vowels of a given line of text. Notice the expected lack of macroscopic patterning in the vowel sounds of Obama's inaugural address compared to the poetic forms. Finally, notice the 7-8 syllable rhyme pattern in D (the right-most columns)

patterns (like the ones shown above) can be better understood. Specifically, I highlight the cognitive processes that support the perception, production, and learnability of rhyme.

Although rhyme and poetic devices have been a subject of interest for a long time, some of these examples, Figures 2.14D, 2.15B, 2.15C, and 2.15D, illustrate a dramatic increase in complexity and size of these structures in recent years. One of the goals of this project is to begin closing the gap between the traditional single syllable perfect end-rhyme assumptions and these increasingly common and complex structures. Since many modern and complex forms of rhyme have not yet been thoroughly documented or studied, this effort constitutes an early step towards that goal.

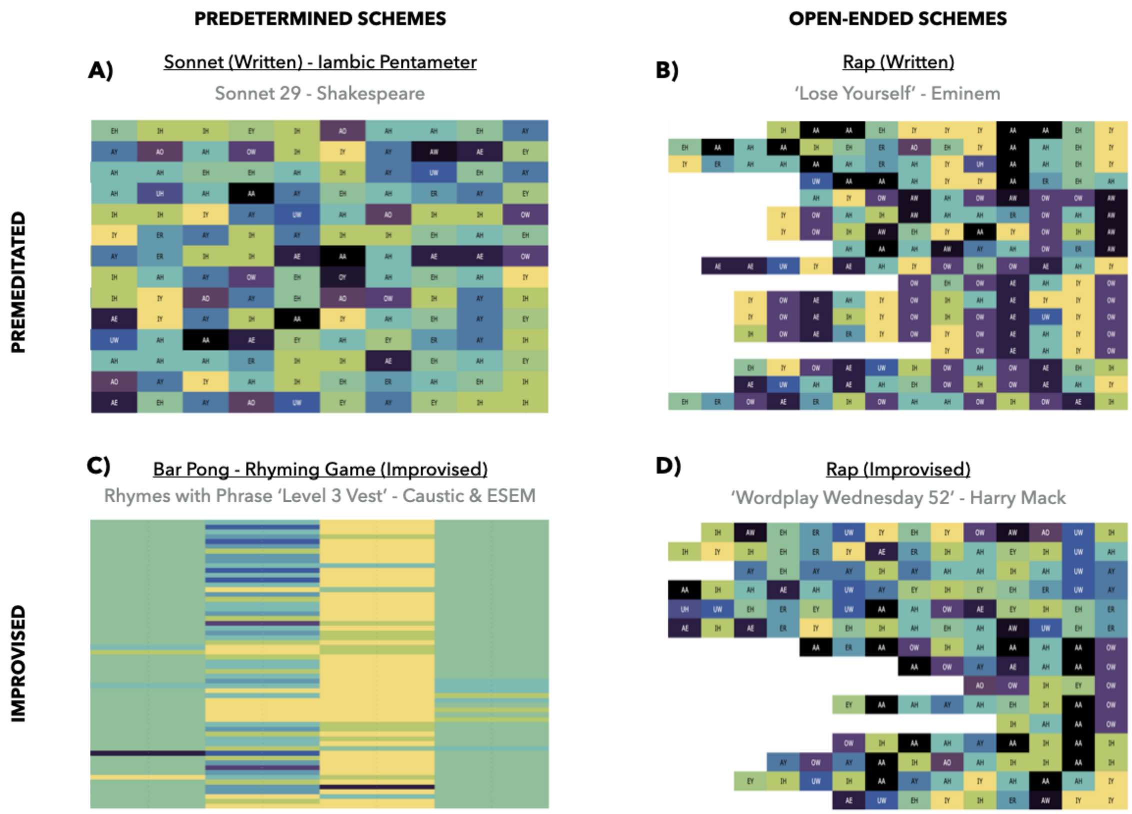


Figure 2.15: 2-by-2 display showing the vowels (in grids) of improvised vs open-ended lyrical patterns. In C, only the vowels from overt rhyme locations are selected and displayed (some 60 4-syllable rhymes with the phrase 'Level Three Vest')

2.3.4 Rhymable.com

Rhymable.com is an accompanying interface that incorporates some of these tools (Figure 2.16). It enables real-time interactive encoding and visualization of small samples of text (e.g. up to 1000 syllables).

This front-end tool is meant for analyzing individual samples and gaining intuition about their underlying sound structures. For corpus studies, I recommend using the python Phonesse package directly.

2.3.5 Data

In order to investigate questions about repeated sound patterns, I utilize data at two scales. At the larger scale, I employ data across the spectrum of human language production including fiction, non-fiction, speech, musical lyrics, and poetry, often with an eye towards zooming in and further exploring notable trends. At the smaller scale,

Figure 2.16: Screenshot of the Phonesse Demo User Interface

I target yet under-examined data from exemplary subcultures within hip-hop (Battle Rap, Improvised Rap, Bar Pong Rhyming Game). Below are short descriptions of these data.

5-Genre Corpus

In some domains (Lyrics, Poetry), sound patterns are employed explicitly and liberally (e.g. poetic devices). In other domains (Non-fiction, Fiction, Speech), phonological patterns are omnipresent, but may be subtle or simply reflective of the phonology of natural language.

This Dataset represents 60+ corpora across a diverse range of language data (non-fiction, fiction, speech, musical lyrics, poetry) and serves as a test-bed for exploring a variety of effects related to language, cognition, and creativity. An important distinction between Poetry and Lyric categories here is that they reflect non-musical and musical samples respectively. More details on the dataset can be found below.

Many of these corpora are stitched together from existing datasets. I also personally collected data including an A Capella Battle Rap corpus and an Improvised Rap corpus (Harry Mack Omegle Bars). A capella Battle Rap is included in the Poetry category and Improvised Rap is included in the Lyrics category (improvised over music).

Battle Rap

As I dicussed in an earlier section, although Battle Rap began as improvised lyricism over music, over the years, its transition to a cappella in format has coincided with common use of multi-syllable rhymes. For this reason it is one of the target domains for this project.

It should also be noted that improvised battles do exist, even ones where turn-taking is every 2 or 4 lines (measures). There are also those who improvise rhyme in a cappella format. Both battles and improvisation come in a variety of forms. I do not cover them all here, but rather select representative examples from their most common incarnations.

Improvised Rap

Improvised rap, sometimes called freestyle rap is another domain of particular interest. Freestyle is a term commonly used to generally refer to rap lyrics outside the context of song. Improvised rap is a difficult and complex skill to learn, and is often associated with simple one-syllable end-rhymes. But some individuals have broken through the simple "end-rhyming" structures of improvised rap, and have begun to utilize various and complex multi-syllable patterns in improvisation. In music, as in lyrics, cultural output can reflect the different cognitive mechanisms used to produce it, and can help us better understand the artistic form itself. In the same way that people study both the written symphonies of Beethoven and the improvised productions of Duke Elington, Miles Davis, here, I will review a selection of improvised rappers.

I have collected two small corpora of improvised rap lyrics. First, in order to compare improvised and written lyrics, I transcribed a rap cipher where 7 rappers take turns rapping in a circle – they each take three turns. Crucially, in the first two go-arounds the artists improvise, and in the third round they rap pre-written verses. This gives a relatively controlled dataset that allows for comparison of improvised and written sound patterns. These rappers are all prolific in the battle rapping scene, and participated in this cipher during the height of the transition from mainly improvised battles to written battles, and so are expert in both. Finally, in order to capture the arc of learning that improvised rapping represents, I also analyze longitudinal data from Harry Mack and one of his students, Ikaanic, who recently began the journey of learning to rap improvisationally. Both of these individuals document many of their training sessions as live streams on YouTube.

Finally, I collect 8 large rhyme sets from a rhyming game called Bar Pong (discussed more in Chapter 4.3). All rhyme sets are 3 or 4-syllable patterns and contains between 20 and 115 rhymes .

In some cases, I will also use data prepared by other researchers and will introduce it when relevant.

Catgegory	Corpus	Source	Word Count
Speech	Dialogue	UCSB Speech Corpus	130,649
Speech	WebChat	NPS Chat	45,010
Speech	Infant Directed Speech	CHILDES	14,468
Fiction	Adventure	Gutenberg	69,342
Fiction	Buster Brown	Gutenberg	18,963
Fiction	chesterton-brown	Gutenberg	86,063
Fiction	Alice In Wonderland	Gutenberg	34,110
Fiction	Edgeworth Parents	Gutenberg	210,663
Fiction	Sense and Sensibility	Gutenberg	141,576
Fiction	Hamlet - Shakespeare	Gutenberg	37,360
Fiction	Macbeth - Shakespeare	Gutenberg	23,140
Fiction	Caesar - Shakespeare	Gutenberg	25,833
Fiction	Science Fiction	BrownCorpus	14,470
Fiction	Romance	BrownCorpus	70,022
Fiction	Mystery	BrownCorpus	57,169
Fiction	Humor	BrownCorpus	21,695
Fiction	Fiction	BrownCorpus	68,488
Musical Lyrics	Pop	lyricsfreak.com	493,213
Musical Lyrics	Country	lyricsfreak.com	377,029
Musical Lyrics	Electronic	lyricsfreak.com	387,182
Musical Lyrics	Folk	lyricsfreak.com	363,884
Musical Lyrics	Indie	lyricsfreak.com	390,271
Musical Lyrics	Jazz	lyricsfreak.com	340,692
Musical Lyrics	Metal	lyricsfreak.com	348,383
Musical Lyrics	Rock	lyricsfreak.com	384,744
Musical Lyrics	Hip-Hop	lyricsfreak.com	991,923
Musical Lyrics	Improvised Rap (Harry Mack Omegle)	Annotated by Author	46,194
Non-Fiction	Editorials	BrownCorpus	61,604
Non-Fiction	Government Documents	BrownCorpus	70,117
Non-Fiction	Hobbies	BrownCorpus	82,345
Non-Fiction	Inaugural Addresses	inaugural Addresses	145,735
Non-Fiction	Belles Lettres	BrownCorpus	173,096
Non-Fiction	Learned - Academic	BrownCorpus	181,888
Non-Fiction	Lore	BrownCorpus	110,299
Non-Fiction	News 1	BrownCorpus	100,554
Non-Fiction	News 2	Reuters	1,253,696
Non-Fiction	Religion	BrownCorpus	39,399
Non-Fiction	Reviews	BrownCorpus	40,704
Non-Fiction	King James Bible	BrownCorpus	1,010,654
Poetry	Sonnets	poetryfoundation.org	61,140
Poetry	Allusion	poetryfoundation.org	14,091
Poetry	Ballads	poetryfoundation.org	34,973
Poetry	Refrain	poetryfoundation.org	18,533
Poetry	Blank Verse	poetryfoundation.org	123,597
Poetry	Confessional	poetryfoundation.org	17,797
Poetry	Couplet	poetryfoundation.org	146,604
Poetry	Dramatic Monologue	poetryfoundation.org	36,327
Poetry	Elegy	poetryfoundation.org	45,308
Poetry	Epic	poetryfoundation.org	131,910
Poetry	Rhymed Stanza	poetryfoundation.org	276,460
Poetry	Free Verse	poetryfoundation.org	640,017
Poetry	Imagery	poetryfoundation.org	17,936
Poetry	Metaphor	poetryfoundation.org	35,279
Poetry	Mixed	poetryfoundation.org	21,446
Poetry	Persona	poetryfoundation.org	32,030
Poetry	Prose Poem	poetryfoundation.org	14,140
Poetry	Series/Sequence	poetryfoundation.org	48,407
Poetry	Battle Rap	battlerap.com	19,116
Poetry	Paradise Lost	Gutenberg	96,825
Poetry	Leaves of Grass	Gutenberg	154,883

Figure 2.17: Brief Summary of Diverse Creative English Texts (DCET) data set.

Chapter 3

Cognition & Computation

The present chapter provides a literature review to motivate intuitions about the cognition and computation of language-based music. I elaborate on the cognitive mechanisms that support verbal sound patterning, rhyme in particular, and why they are important to take into account. Then, I review the computational and linguistic approaches conventionally used so far in characterizing rhyme and other sound patterns in creative language.

Questions Covered:

- (Section 3.1) What cognitive mechanisms support learning to rhyme?
  - Topics include, memory, language change, phonology, becoming literate, educational applications, and expertise
- (Section 3.2) What cognitive mechanisms underlie rhyme perception and production?
  - Topics include rhythm, similarity, rhyme acceptability, orthography, priming, words, and improvisation
- (Section 3.3) How has rhyme been formally investigated?
  - Topics include, stress, sound sequences, poetry, rap, and lyric generation

3.1 Learning

- What cognitive mechanisms support learning to rhyme?

The cognitive mechanisms that underlie rhyme (and other forms) are related to both music and language. Rhyming is an ability English speaking children learn at a young age, and as adults they reliably know how to produce and perceive simple rhymes. Because the learnability of multisyllabic rhyme has not generally been explored, I do not focus on it in this review, although that is the ultimate target of this course of investigation. Instead, I break down rhyme into its relevant components and contexts in order to begin building a more cohesive picture of rhyme and its elements: from its basic cognitive processes and mechanisms, to the complex and expert manifestations of rhyme as they are found in the wild.

3.1.1 Memory

Memory is crucial for learning and can be broadly construed as the process of acquiring, storing, and retrieving information. Repetition and similarity, taken together, gives the functional elements required to begin addressing the well known effect and benefit of rhyme on memory. Some have noted, both in personal reflection and scientific study, that if sentences are encoded in a rhyming fashion, humans are more likely to remember them [113, 28]. But how does this translate to the cognitive mechanisms that are theorized to exist in the minds of individuals? And how do these memories get passed from one person to another or one generation to the next?

The Phonological Loop

Memory is critical in guiding behavior as humans process and apply temporal information from the world [114]. At this point I should distinguish between more short term processing and long term memory. The preeminent theory of short term memory, also referred to as working memory, breaks down various aspects of mental processing into a few specialized components. This theory, for the most part, stands up to scrutiny [115, 116]. Here, I briefly describe the Baddeley model of working memory in order to motivate the context for more theoretical interpretations of various rhyme related studies.

The Baddeley Working Memory model (Figure **??**) consists of 4 components, a central executive module and 3 'slave' modules: the phonological loop, the visuo-spacial sketchpad, and the episodic buffer. The basic idea is that the central executive module allows coordination of the other 3 systems, as well as various other 'executive control' mechanisms like task switching, attention, updating, and encoding information.

The episodic buffer is the most recent addition to this theory of working memory [117], which is believed to participate in the integration of verbal, visual, and spatial information and their organization in terms of chronological sequencing [118]. It is also suspected that this buffer interacts with long-term memory and seman-

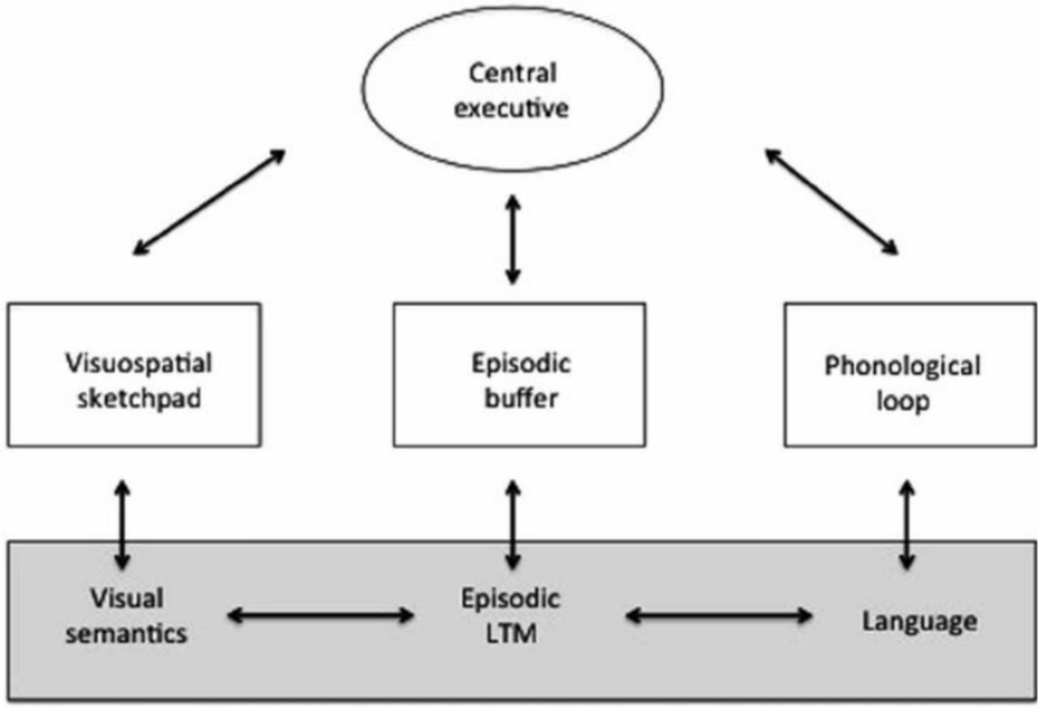


Figure 3.1: Baddeley's Working Memory Model [6]

tic stores [117]. The visuo-spatial sketchpad, as one might suspect, is a short term memory buffer that supports visual (color, form, etc.) or spatial (position, orientation, movement) information. These visual and spatial components are considered to be somewhat independent of each other [119, 120], and interestingly, are almost entirely independent from the last 'slave' system, the phonological loop [121]. There is evidence suggesting that not only do the phonological loop and the visual-spatial sketchpad operate independently, but also they can be utilized in parallel [119]. This may help explain the cognitive mechanisms underlying how it is that Harry Mack is able to fluently read, make semantic associations, and produce rhyming lyrics at the same time.

The phonological loop is the main target of interest here because it is thought to process sound and rhythmic (stress) information, both of which are fundamental to the phenomenon of rhyme. Once perceived, spoken language is presumed to be automatically encoded into the phonological loop's memory store, whereas written (and read) language makes its way into the phonological store by 'silent' or 'mental' articulation. There are a number of lines of research that attempt to isolate and explain the workings of the phonological loop. Here, I will discuss a few research projects that are relevant for understanding the lower level workings of rhyme processing in the brain.

As previously discussed, from at least the age of 3 [122], humans seem to be better at remembering semantically related lists of words than phonetically related ones

[123]. Indeed, one of the main effects associated with the phonological loop is that phonologically similar lists of words are harder to recall than semantically similar lists of words. The assumption is that in both cases, words are being encoded into the phonological store, but that in the case of similar sounding words, the memory traces of each next word are so similar that they overwrite previous word traces [123, 124]. Indeed, this effect of overwriting memory traces in the phonological store can also be seen when subjects are asked to produce additional distractor words out loud (words in between the words of interest). These distractor words are believed to introduce interference into articulatory rehearsal, encouraging the decay of memory traces in the phonological store [125].

The mode of presentation of stimuli is also impactful (e.g. word lists). A related recency effect in memory has been observed when hearing versus reading a list of words, where more recent auditory signals are remembered relatively better than recent visual signals, implying that memory interacts with auditory and visual signals in different ways which are consistent with the phonological loop proposition [126].

The phonological loop, and related phonological interference effects associated with the cognitive processes of working memory, could play an important role in the perception of complex rhymes. Together with working memory capacity, they could potentially affect the complexity of the rhymes that can be discerned or produced. Furthermore, for people producing improvised rhymes from stimuli, their alternating between perception and production, or intentional and automatic action, would seem to depend heavily on the central executive module. Specifically, shifting focus from reading stimuli, to thinking of semantically (and phonologically) associated words and concepts, and then producing coherent and rhythmic language includes involvement from task switching, attention, and updating or encoding information processes.

Crystallization

Despite a natural tendency of language to mutate, change, and evolve over time, rhyme seem to facilitate crystallization (fixing or cementing) of language content against these normal forces of language change, in some contexts.

Although accounts of the phonological loop help illuminate this phenomenon in terms of short term memory, it largely leaves out long-term memory and crystallization. As noted above, there are suspected connections between short term and long term memory stores (episodic buffer, etc.), but the issue of memory crystallization, both at the individual and the societal level, are a different phenomenon altogether.

One of the most significant contributions to the cognitive understanding of rhyme and societal memory is "Memory in Oral Traditions: The Cognitive Psychology of Epic, Ballads, and Counting-out Rhymes" [28]. It investigates observable changes in various components of oral traditions as they are maintained and passed down through generations. Over time, *short* poems like ”Eenie meenie minie moe”, which belong to the the most common genre explored in the book (counting-out rhymes), remain so stable that only two words have changed in over 100 years, and neither were part of the rhyming structure. This may feel intuitive, given how compelling, stable, and memorable works like ”Eenie, meenie, minie, moe” seem to us, but it points at an important difference between this kind of language art and natural language: a strong tendency for crystallization.

Of course not all language arts crystallize in this way. For example, in epic poetry the length and content demands are so great that specific words, and associated rhymes, are extremely prone to change as they are passed down orally through the generations. However, other components are crystallized, largely preserving the plot, the rhythm, the characters, and many other genre/style features. In the case of ballads, which are considered medium length, the crystallization looks much more like those of epic poetry, though the words change more slowly [28].

There are a variety of other memory effects that could be relevant for rhyme. For example, the integration effect shows that “recognition of the melody (or text) of a song is better in the presence of the text (or melody) with which it had been heard originally than in the presence of a different text (or melody)”. It is still unclear if this effect may be due to the physical interaction hypothesis (song components exert subtle but memorable physical changes on the other components) or the association-by-contiguity hypothesis (any two events in proximity act as cues for each other). Both factors seem to be at play to some degree [127].

3.1.2 Language Change

In addition to the forces that help codify or cement the patterns of language in our minds, there are also forces that push language towards change over time. In some cases it is due to neutral drift (and other large scale environmental factors), as has been shown to impact biology and language [128, 129, 130, 131]. But there are mechanisms in communication that seem to operate at even shorter time scales. This isn’t surprising in light of the well-known “telephone game” effect [132], where messages tend to dramatically change as they are passed from agent to agent, often changing in directions that reflect cultural common ground or that facilitate communication. Moreover, change over time that is observed within these behavioral and simulated communication experiments, happens in both linguistic and non-linguistic contexts [133, 134, 135, 136].

Iterated learning (derived from the telephone game) is often touted as a paradigm that demonstrates the emergence of compositionality and systemiticity in communication, meaning that certain elements come to have stability or meaning, and then are used in combination with other elements to create novel combinations or messages. However, in experiments with children and adults, the emergence of compositionality was only found in adults, whereas systematicity was found in both groups [137]. This indicates that certain cognitive processes or linguistic abilities are not available to children (or must be developed or learned). This, among other factors, points to age as an important factor for the cognitive mechanisms of language acquisition.

How is this relevant to the world of rhyme? On the one hand, rhymes and rhyming patterns are likely subject to iterated change across individuals and generations. In fact, when the messages being passed are sounds, iterated learning experiments have recently been shown to produce "verse templates" of syllabic patterns [38]. On the other hand, the larger rhyming patterns become, the more relevant it is to ask questions about not only the systematicity of multi-syllable sound patterns, but also their internal structures and combinations. For instance, is a given 5-syllable rhyming pattern just that? Or is it composed of two smaller 2-syllable and 3-syllable structures? Or perhaps a 4-syllable and a 1-syllable pattern? Direct investigation of the structure of multi-syllabic rhymes, in individuals and across genres, help help determine how multi-syllabic rhyming patterns emerge. Much like previous iterated learning studies, artificial language learning paradigms can be used to test how computational and behavioral iterated learning studies compare to the phonological structures of multi-syllabic rhymes found in the wild.

3.1.3 Phonology

How are phonological items and patterns learned in the first place? The most notable line of research on this topic stems from the statistical learning literature, which broadly suggests that learning happens as agents extract statistical regularities from their environment. Here, I explore the phenomena of segmentation and statistical learning.

Learning to Segment

In order to notice the frequencies of items and co-occurrence patterns one must be able to distinguish between elements in a speech stream in the first place. This skill is relevant for identifying the rhyming/matching components of pairs of words, but also for distinguishing between words (or sounds) themselves. This is a difficult task as there are not clear markers of boundaries between sounds or words in spoken language. It should be noted that there is evidence to suggest that segmentation between syllables may not operate specifically on a mechanism for identifying boundaries, but rather, based on recognition of certain statistical regularities and properties within syllables (onset & coda), including sonority [138]. Furthermore, speakers seem to use sequential probabilities of sound items as cues for segmentation. This is the case, particularly for sound sequences within words, and notably, these regularities are most reliable for sequences in the onset position of the syllable [139].

A number of both computational [140, 141] and experimental models [142, 143] attempt to account for the task of word segmentation in auditory streams of speech. It seems that segmentation cues, which include lexical, segmental (phonemes), and prosodic (stress) features, influence segmentation learning in a hierarchical fashion (lexical = strongest, prosodic = weakest). A complex of these features contribute to successful segmentation, and understanding how these levels are integrated is critical to painting a full picture of the task of segmentation [144]. Although the suite of features contributing to segmentation ability is often considered within native languages, their relative weights can change in second language learners, particularly in relation to the statistical distributions of the speaker's first language [145]. Furthermore, a hypothesis that word segmentation is based on the statistical regularities of language, irrespective of perceive attention, is not well supported. Indeed, some evidence suggests that segmentation performance is dramatically impaired when attention is interfered with, suggesting attention is critical for segmentation [146].

In terms of rhythm specifically, by 9 months of age, infants are sensitive to rhythmic structure in 'head-turning' based segmentation tasks [147]. Not only that, 7.5 month olds seem to have the ability to segment two syllable words with particular rhythmic structures, specifically the common stress form (in English) of strong-weak, or stressed-unstressed. Infants of the same age are also able to segment 3-syllable words from fluent speech, as long as the word begins with a stressed syllable [148].

As adult English speakers, there is a general "heuristic ... that strong syllables (containing full vowels) are most likely to be the initial syllables of lexical words, whereas weak syllables (containing central, or reduced, vowels) are nonword-initial, or, if word-initial, are grammatical words." This heuristic is supported by experimental segmentation evidence [149]. Although this is specific to English, one would suspect the particular phonological distributions of each given language will have an impact on the statistical regularities leveraged in learning and segmentation. Finally, it has been demonstrated in longitudinal studies that groups with 2 years of musical training show marked improvement in word segmentation over groups without musical training. This suggests that the cognitive processes underpinning rhythm and rhythmic training, even in a musical context, can facilitate verbal word segmentation [150].

Statistical Learning

As I just discussed, humans are able to differentiate and segment components of continuous streams of speech, explainable through generalized statistical learning mechanisms. Can extracting statistical regularities from the environment also shed light on the cognitive processes behind the ability to perceive and produce rhyming patterns?

Form-meaning relations are fundamental to the grounding of various phenomena. As mentioned above, word segmentation is critical to the processing of perceiving and interpreting language. Some of the earliest evidence for statistical learning comes from infants who demonstrate a sensitivity to transitional probabilities of sounds in language that facilitate language acquisition [151]. Again, this is not wholly surprising given that the ability to mentally map the distribution of sound structures in language facilitates the mapping of sounds to meaning in the learning of words [152].

Effects of sensitivity to the distribution of sound structures in language learning are evident across age groups. Children of 7, 9, and 11, as well as adults, were better able to recognize high-frequency words that had low neighborhood density (i.e. few words that sound similar) than high densities. The effect is reversed for low frequency words. This suggests not only that is there awareness of the sound distribution of vocabulary of a lexicon, but also, that there is competition in the heads of subjects that leverages statistically learned neighborhood density information to facilitate (or delay) the recognition of words [153]. This cognitive mechanism of competition would also seem to impact the ability to recognize rhyming words (or sound patterns) from high versus low neighborhood environments.

3.1.4 Becoming Literate

Rhyming, and manipulating other poetic devices, may represent special cases of phonological awareness, i.e. the ability to remember, discern, and manipulate sounds at different linguistic levels (words, phrases, syllables, phonemes, etc.). Early theories of literacy suggested that "phonological awareness was a result of the experience of learning to read", however, some have contested that it is, in fact, the reverse, and that phonological awareness facilitates literacy [154]. A follow-up to this has shown that, in educational settings, it is effective to take a balanced approach using correspondences between letters, phonemes, rimes, and their sequences, to best facilitate learning [155]. Cognitive neuroscience is also beginning to be applied to the field of educational psychology and literacy [156, 157], but much more progress is required before definitive recommendations can be made [158]. As mentioned earlier, a growing body of recent work has shown that there is not only a positive correlation between rhythm and literacy [155, 159], but also that phonological awareness is causally linked to reading and spelling skills, especially in children [160, 161].

Additionally, approaches to learning can be either implicit or explicit. Explicit learning produces knowledge that is able to be recalled consciously, whereas implicit learning produces knowledge that has some impact on cognition, yet the learner is unaware of [162]. Much of language learning is done without explicit instruction, especially for children (grammar, morphology, etc.), and yet, native speakers become fluent. Pertinent to the discussion of rhyme, sound patterns can also be learned implicitly [163]. Implicit learning can even provide some advantages over explicit learning [164], and this implicit learning can happen after only one exposure to a sound pattern [143].

However, components of certain processes, like reading, do rely on some amount of explicit learning, spelling, phonological awareness, etc.. It is important to be aware of this dynamic between implicit and explicit learning. Rhyme is a kind of phonological pattern, and humans clearly have an aptitude for absorbing knowledge in this domain implicitly. Simply focusing on delivering explicit knowledge related to rhyme may not necessarily be the most effective teaching approach. Effective learning of rhyme patterns will likely follow paths of learning in other areas of both language and music that rely on a balance between implicit and explicit acquisition of knowledge.

3.1.5 Educational Applications

It is well known that rhyming language [165] can be easier to remember than non-rhyming verse [28, 166]. Moreover, phonological awareness [160] and rhythm [161] are strongly linked to the ability to read and spell. For these core reasons, rhyme has been used to promote literacy and language learning from preschool [167] through higher education [168, 169]. It has also been used in the classroom in order to promote the learning of specific content areas [2] and to create more engaging and memorable learning experiences [170, 171]. More recently, rhyme has been utilized as a scaffolding to promote English learning in second language learning environments [172, 173].

While rhyme can be applied in a number of ways for learning both content and literacy, it is also possible to consider rhyme as a skill and an art form, in and of itself. This is particularly so in the context of producing content with high densities of rhymes or in improvisation. In these contexts, rhyme becomes more than just a delivery mechanism for other purposes (content-areas, literacy development). Rhyme matching, search, expertise, and the ability to improvise or express oneself in high density rhyme, while still maintaining grammaticality and semantic themes or narrative structure, highlights rich domain of learning. In the previous subsection, I reviewed some of the basic guidelines and insights educators currently employ with regard to rhyme and learning, but here, I note the absence of insights/methods/curricula for learning and progressing in the realm of fluent and dense rhyme environments.

The current approach frames learning to rap improvisationally as a form of language skill acquisition, a productive and creative kind of expertise that seems to be learnable in any language. Learning technology has even been shown to "generate new linguistic habits" in rappers [174], so both beginner and veteran rappers are likely to benefit from more formal pedagogical resources.

In order to facilitate the learning of more complex rhyming abilities, one might integrate a number of representations, skills, and tools, combined with insights from those few who have achieved expert rhyme fluency. First, understanding the space of phonological patterns and poetic devices, particularly multi-syllabic ones, is a fundamental step in allowing for the development of this discipline. Second, search and visualization tools (RhymeGenie; Rhymezone.com; WordSurge.com, Rhymable etc.)

are necessary for identifying and navigating these patterns, and their connection to the more familiar aspects of language. They can be used for both individual investigation as well as pedagogical curricula design and development. Applications like RhymeGenie and RhymeZone allow people to find words that rhyme given some initial word.

Rhyme (or sound pattern) search is an obvious application, but only the first step for computational tools in this arena. Search interfaces like WordSurge.com [175] extend this approach by providing networked connections between rhyme, consonance, and other semantic categories from WordNet (definition, synonyms, antonyms, hyper/hyponyms, etc.). Finally, the description of rhyme, both within overt schemes, and within their larger poetic contexts (poems, lyrics, speech), is critical to visualize and document the complexity of sound patterns in language.

Many other resources would be similarly useful. For example, a resource containing all multi-syllabic stress and vowel patterns (e.g., up to 8 syllables), their frequencies, and word/phrases that match them could be employed to develop teachable knowledge of the elements and networks of phonological patterns that underlie multi-syllabic rhyme. These could then be used as elemental building blocks for a number of other applications. The development of programs for learning the art and science of rhyme could facilitate not only the techniques and styles of rhyme, but also the pedagogical paradigms that leverage rhyme as a delivery mechanism for content, literacy, memory, or expression.

All of these tools for explicit instruction and learning raise a question. When is explicit instruction the best technique, and when might educators utilize implicit learning approaches? Explicit learning produces knowledge that can be recalled consciously, whereas implicit learning produces knowledge that has an impact on cognition, yet the learner is unaware of it [162]. A lot of language learning is done without explicit instruction, especially for children (grammar, morphology, etc.), and yet, native speakers become fluent. More specifically, sound patterns can be learned implicitly [163]. On the other hand, components of certain processes, like reading, rely on some amount of explicit learning, spelling, phonological awareness, etc.. When considering the appropriate tools and applications for educational contexts, it is important to be aware of and balance the dynamic between implicit and explicit learning. Rhyme is a kind of phonological pattern, and since humans clearly have aptitude for absorbing knowledge in this domain implicitly, simply force feeding learners vast amounts of explicit knowledge related to rhyme is not necessarily the most effective approach.

In the space of music and language more generally, tools for explicit learning have had a long time to develop, and a balance between implicit and explicit components of learning has matured to be sufficient for facilitating advanced learning while reducing learning times. These frameworks are even more needed when learning improvisational rhyme, where exploring, identifying, remembering, and utilizing underlying phonological patterns can be dramatically more inaccessible. At this time, the use of tools, technology, and curricula that target the learning of phonological patterns are almost non-existent. However, it has been shown that technology can "generate new linguistic habits" in rappers [174]. It seems reasonable that technology and pedagogy may be useful across the development in language-based music skill.

3.1.6 Expertise (Perception)

Discussion of the *generation* of improv performance will be addressed in the next section. Here, I review work that focuses on the *perception* of improvised performance.

As I noted earlier, musicians can identify details in music that the untrained ear cannot. So it would not be surprising if both experience and a range of features underlying language, like stress (rhythm) and sequence size or complexity (syllables), impact the discernment of rhyming in the relevant perceptual, atriculatory, and phonological domains, although no work has been done yet to investigate this question.

There are a variety of contexts and forms in which rhyme can be found. Although the eventual targets of interest are multisyllabic and improvised rhyme, contexts with large multi-syllabic rhyming schemes have only just become prevalent, and as such, there are no studies of this phenomenon. Similarly, improvisational rhyme has only recently become the subject of a few academic works which I review below in relation to similar studies on jazz improvisation.

To begin with, as seen in Figure 3.2, the neural patterns measured in musicians vs non-musicians who were listening to music are quite different. Whereas non-musicians primarily demonstrate activity in the delta bands (attention tasks), musicians primarily showed activation in the gamma bands (memory matching). In other words, the brain activity, and likely the associated perception of music, is different when the listener is also an expert in the form in question [7]. More specifically, musicians are associating what they hear with past experience (memory matching) and relying less heavily on attention mechanisms than non-musicians.

Unlike pre-meditated performances, improvised performances have a character of spontaneity which distinguishes them, but can humans perceive this difference? When 22 jazz musicians were asked to judge piano melodies as improvised or imitated, they only performed slightly above chance (mean 55%; range 44–65%). "Amygdala activation was stronger for improvisation than imitations", implicating it in detection of the relative uncertainty associated with improvisation [176]. In a related effort, another group tried to computationally classify improv vs not-improv by examining the EEG signals of 14 improvising jazz guitar players as they listen to alternating improvisations and musical scales. Machine learning techniques were able to attain an accuracy of 75% in differentiating between them, higher than human performance [177], suggesting that there are additional cues available in the signal that human subjects may not be integrating in their judgements.

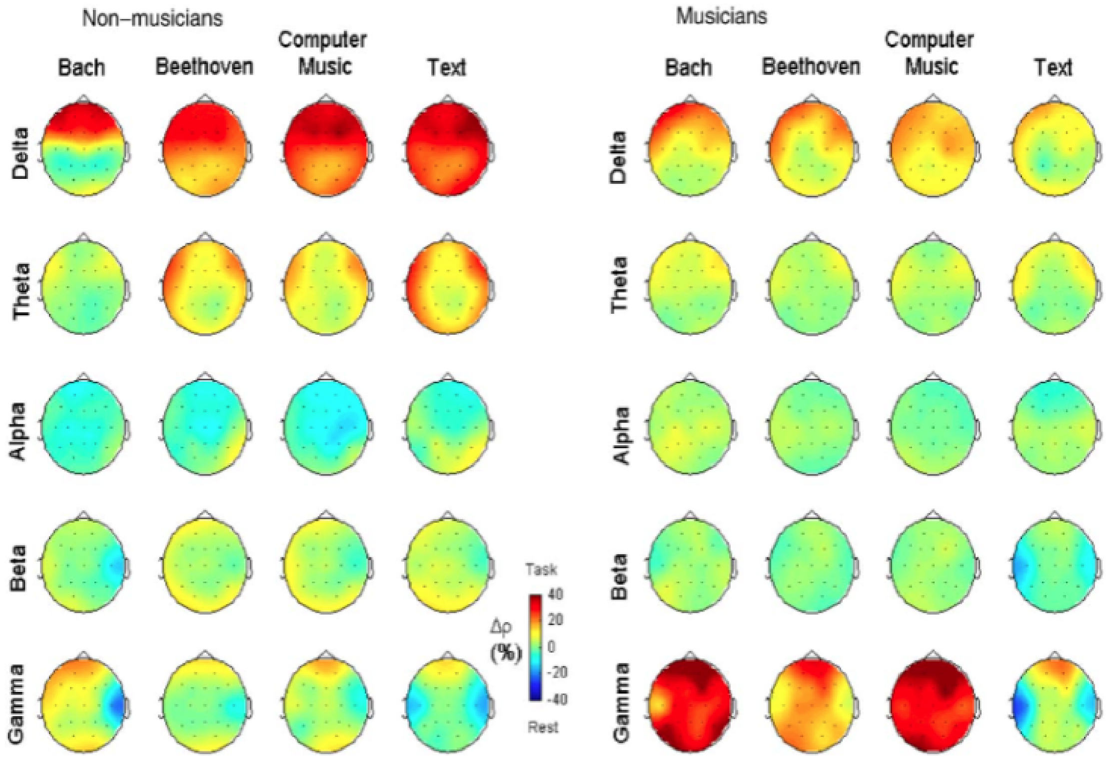


Figure 3.2: Demonstration of the different brain regions activated across musicians and non-musicians when presented with various kinds of stimuli. EEG of mean differences of phase synchrony in frequency bands [7].

Anecdotally, many individuals who have been introduced to the growing world of rhyme have commented that there is a noticeable learning curve in their ability to recognize and appreciate complex phonological patterns when listening to lyrics (i.e. multisyllabic rhyme, alliteration, or assonance). At first, accounts often reflect hearing just one-to-two syllable end-rhymes, and only many listens or many months later do they realize these same rhymes are in fact nested within much larger structures (e.g 3-10 syllables). Empirical studies could be conducted to identify if, and where, possible learning effects are related to perception of multi-syllable phonological patterns, rhyme expertise, or musical experience.

It should also be noted that there is constant confusion in the rhyming world over the classification of rhyming performances as improvised or not. Part of this is due to the term "freestyle" which, at its base, simply means 'rhymes outside the context of a song', but is often confused for (and can sometimes be cover for those wanting their pre-written lyrics to be confused for) improvised or "off the top" performances. Not only is improv rap genuinely difficult, but also, it is a rare form, and therefore few possess the relevant expertise to distinguish it from written performance. The specifics of improvised rhyming and its relevant linguistic cues may provide different opportunities for the perceptual differentiation of improv vs non-improv rhyming;

these features include rhythm, melody, rhyme, semantic content, filler phrases, jokes, observations, vocabulary. Much like in improvised jazz performance, machine learning classification (using a subset of those features) and human judgements could also be employed to test differentiation between written and improvised performance of expert rappers.

3.1.7 Expertise (Production)

Expertise can take many forms. In the literature, two main kinds of fluency tasks are common, semantic fluency and phonological fluency. They usually consist of a subject being asked to produce as many words as possible in a given semantic category (e.g., farm animals) or phonological category (e.g., words starting with F or P) within a given duration. That said, expert rhyme fluency, especially in improvisation, is about more than just the canonical verbal fluency tasks. A long term goal of this review is to begin asking the question, how is it possible to produce improvised poetry or rap with the level of complexity and fluency seen in premeditated poetic works or improvised musical performance? What cognitive networks are involved in this kind of fluency, and why does it seem to be so rare? To investigate this, one could contrast improvised rhyming, premeditated rhyming, improvised conversation, and improvised musical performance (jazz), with an eye towards identifying differentially active brain regions in subjects who are measured to produce more fluent and complex content. There are also a number of traditional verbal fluency tasks that can also shed light on the phenomenon of expert rhyme fluency by revealing the dynamics of relevant semantic and phonological networks across ages and levels of skill.

But how complex are the phonological structures that human generate improvisationally? Can anyone become fluent in rhyme at high levels of complexity? Arguably, a handful of rappers have gotten close (e.g., Harry Mack, Juice, Juice WRLD, Blind Fury, King Los, Franco, Thesaurus, Charron). Although many individuals in the world of rap try, very few succeed at becoming anywhere near fluent improvisational rhymers. Moreover, the varying styles of elite improvisational rhymers seem to lead them to develop dramatically different skill sets that play on different distributions of linguistic constraints (vocabulary, grammar, complexity, multisyllabic rhyme, puns, punchlines), as well as musical constraints (stress or rhythmic patterns, a capella vs to-music, cadences, melody, etc.).

Why is this? Is it just their individual styles (like in conversation), or is it a lack of pedagogical tools or instructional materials? It is surely possible that complex improv rhyming may impose particularly demanding or unique constraints on the normal production of language that make it difficult to learn. Perhaps there is some vocabulary or grammar that must be acquired? Often improvised rhymers (who are learning) stop mid-utterance not knowing how to complete a thought and still rhyme. (Try to rhyme a few thoughts together yourself to see what I mean.)

No matter the domain of expertise, the underlying cause of peak performance is hotly debated. One theoretical camp, the habitualists, focus on the realm of automatic execution [178], while another camp, the intellectualists, focus on higher-level cognition and intention as the main drivers of peak performance. A recent paper takes a pluralistic stance on this problem, suggesting that "skilled behavior weaves together automaticity and higher-level cognition... [as] both [are] normal features of skilled behavior that benefit skilled behavior" [179]. This is consistent with the unique presence of both intention and automated action in an EEG study of lyrical improvisation, discussed more in Section 3.2.7 [180, 181].

In order to understand the cognitive dynamics of rhyming expertise, it would be extremely productive to present canonical learning, perception, and production tasks to both improvising, and non-improvising rappers (as well as non-rappers) in order to understand how they perform, both neurologically speaking, and in terms of various language/cognitive tasks. For example, administering a suite of tests to subjects including working memory capacity tasks, phonological and semantic fluency tasks, the flanker task (inhibition), stroop task (verbal/visual mismatch), attentional control/attention capture, and sound similarity judgment tasks may enable differentiating a particular set of individual abilities that facilitate or are reinforced by rhyming skill. It could also allow for better backing-out of some of the cognitive mechanisms related to expertise in multi-syllabic and improvisational rhyming. Finally, it would also be informative to compare the neural correlates of the various styles, stimuli, and abilities of rhyming in humans, with special attention paid to the phonological structures and words they produce.

3.2 Perception & Production

- What cognitive mechanisms underlie rhyme perception and production?

In order to consider the full scope of the phenomenon of language-based music and its forms and cognitive processes, I take a closer look at the cognitive mechanisms that support phonological processes and rhyme. I will also cover questions related to the source of rhyme processing in the brain and how it is related to other processes.

The male proclivity for rhetorical phonological displays of skill has not been well studied [37], but modern subcultures still exist that continue to demonstrate this trend. It can be difficult, however, to differentiate where this tendency is due to gender preferences and where it arises due to gender inequality. Although this paper is not focused on gender differences, a simple review of relevant data (Figure 3.3) shows a dramatic disparity in participation of heavily rhyming language arts, and is certainly cause for future investigation.

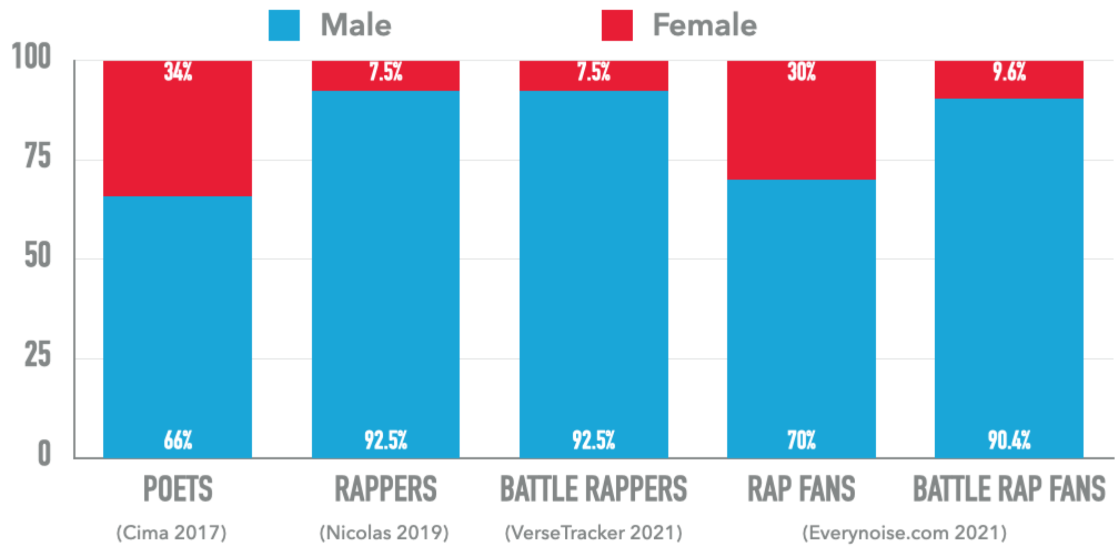


Figure 3.3: Gender Differences Across Rhyming Domains

3.2.1 Rhythm

Rhythms, or the repeated patterns of intervals in time, like most other aspects of language and music, unfold over some duration. They are a macroscopically noticeable aspect of both domains, and serve as a foundation for much of the experienced time course of events. The idea that rhythm has aesthetic quality and provokes a sense of movement is not particular to music, language, or dance [182, 183], but it is in those domains that the variations and complexities of rhythmic patterns have attracted much study. Some phenomena, like groove, for example, are often understood as a sense of movement [184, 185] or a "wanting to move", brought on by certain rhythms in music. Sometimes groove is even interpreted as patterns of performance or a state of being. Nonetheless, the rhythmic patterns that tend to underlie groove can be described in terms of five classes, pulse or regular beat; subdivision of the beat; syncopation; counter-rhythm; and microrhythm [186].

Other empirical measures of rhythm have been developed to compare the rhythmic structures of language (syllables, duration, stress) and music (tone, beat, organization, multiple time-scales) [187]. Early findings of this work indicate that the rhythmic structures of a language or culture can, in fact, impact the rhythmic structures of that culture's musical forms [188].

Together, these observations about rhythm bring up two fundamental issues, how and where does temporal processing, and more specifically, rhythmic processing, occur in the brain? One might expect that there is an identifiable region responsible for coordinating time based processing, but instead, meta-analyses on a variety of time-based processing studies reinforce the consensus that perception of time in the brain spans a vast distributed network. However, depending on the task design and demands, various regions are shown to be associated with temporal processing, including the supplementary motor area, premotor cortex, inferior frontal gyrus, anterior insula, inferior parietal cortex, posterior superior temporal gyrus/sulcus, and subcortical areas (basal ganglia, thalamus, cerebellum) [114].

Music

Rhythms are central to musical perception and production. However, it has been demonstrated that the auditory system is coupled with a variety of motor regions during rhythm and beat perception, indicating that the neural correlates of perception and production of rhythm may utilize similar mechanisms [189]. Building on top of a distributed network of temporal processing in the brain, rhythm and beat perception seem to depend on the premotor cortex, supplementary motor area, and basal ganglia [190]. More specifically, listening to musical rhythms seems to recruit motor areas of the brain, including supplementary motor area (SMA), mid-premotor cortex (PMC), and cerebellum [191].

In addition, these "[m]usical perception-action coupling[s], set up repeated cycles of prediction, expectation violation, and resolution" [192]. And due to the similar rhythmic underpinnings of music and rhyming, it would not be surprising if a similar repeated cycle of prediction, expectation violation, and resolution of rhyme and rhythm creates an analogous perception-action coupling in language, or even recruits similar cognitive mechanisms used in musical processing (perhaps additionally recruiting language centers).

Language

All words and phrases have a stress (or rhythmic) pattern. As was discussed earlier, rhythm patterns can be an important cue for infants [193, 148] and adults [194, 149] as they learn to segment words. Stress patterns in language also tend to define a stereotypical rhythm of alternating stressed and unstressed syllables within a given language [195, 196], and have even been argued to be an underlying force behind the internal change and evolution of language structure [197]. Furthermore, rhythm in language can be examined from a number of perspectives including symbolic representation, production, perception, and communicative function [198].

Rhythm and rhyme patterns are coupled at many scales. Not only do all words or phrases have a rhythmic (stress) pattern, but rhyming words often agree in their underlying stress pattern. Additionally, forms like sonnets have both end-rhyme patterns and stress constraints (iambic pentameter), that define their poetic structures. Although sonnets often contain metrical and rhyme constraints, both of which imply stress requirements, they are independent, but overlapping, sets of constraints.

3.2.2 Similarity

So far, much of the discussion has invoked notions of similarity, but how is similarity perceived? And how does this relate specifically to speech sounds? First, it should be acknowledge that there are no universal measures of similarity. In the case of speech sound similarity, for example, reliable measures of similarity can be discussed in terms of distinct dimensions like phonological, articulatory, or perceptual similarity [199].

Faithfully representing sounds and their relations is much more complex than just discretizing them into phonemes and placing them into a grid. As mentioned above, speech sounds in natural language are continuous signals which are practically discretized into differentiable phonemes for perceptual, linguistic, or computational purposes. The mode and level of description chosen (e.g. discrete or continuous) has deep implications on the measures of similarity that are appropriate for comparing sounds.

Indeed, much like similarity in other realms, the specific measure chosen will change based on the mode, context, and purpose. In the realm of verbal sounds there seem to be at least 5 relevant types of similarity: Articulatory (vocal tract), Acoustic (sound waves), Perceptual (behavioral & cognitive), Phonological (distinctive feature based), and Phonological Patterning [200]. As it turns out, perceptual similarity has the strongest impact on rhyme similarity judgments, followed by articulatory similarity [199]. Although some measures are simpler and more completely defined (e.g phonological feature based, articulatory), and one may expect them to account for the "ground truth" of the perceived features of sound, there are counter-intuitive relations between these elements that have important influences on 'perceptual similarity and confusability' [201]. Furthermore, while distinctive feature or identity based similarity measures may be assumed to be symmetric, when perceptual similarity is taken into account, this is not true. Instead, in some cases, similarity judgements depend on the order that sounds are presented, for example "the difference between [y] and [u] is more detectable when [y] is presented before [u] than in the opposite order", a phenomenon called vowel perception asymmetry [202, 203, 204].

Because most work on rhyme uses mainly IPA representations of phonemes, which are based on discrete segments, two simple approaches to similarity should be noted. First, perfect similarity or identity can be easily established when two segments are identical as in the rime components of 'CAT' and 'BAT'. Second, some amount of imperfect similarity can be established by comparing the number of similar/different distinctive features between two segments (/p/ vs /b/, or /p/ vs /m/, or /p/ vs /p/, etc...) [205]. For example, one might determine how similar the words CAT and CAP are based on comparing how close /t/ and /p/ are in distinctive feature space.

Since rhyme is so understudied, most approaches reasonably focus on perfect identity. Anything less than perfect similarity often gets lumped into the broad category of "imperfect rhyme". This is currently a catch-all for a variety of possibly different levels of similarity that have yet to be comprehensively described. Because of the sparsity of research in this space, it is reasonable to expect the study of rhyme, and similar language sound patterns, to begin with perfect rhymes in order to identify larger and more robust macroscopic structural and phonological components. As research in rhyme and language-based music becomes more developed, researcher can graduate to a more thorough inclusion of segmental similarity, distinctive features, perceptual similarity, and finally, more subtle and continuous measures of similarity (including signal processing based measures).

One might wonder about the degree to which humans are sensitive to sound similarity. For example, are bilinguals sensitive to sound similarity in both their L1 and L2 languages? To some degree, yes. This effect was demonstrated in terms of sensitivity to neighborhood density (number of similar sounding word neighbors) in picture naming response times and eye-movement data, as well as in auditory word choice tasks where presented words have an overlap in the sounds of the subject's L1 and L2 words [206]. Also, when tested on their ability to detect phonological mismatches in words, children seem to be able to detect up to 3 featural differences, more than the 1 to 2 that were implicated by prior work [207]. Furthermore, some words that are classified as homophones, and that many believe are pronounced the same (e.g. freeze vs frees; paws vs pause) are actually spoken with reproducible differences in duration (4-7%), in cases where the words originate from different morphemes and one of the homophones ends with an inflected fricative, e.g. [f, s, v, z] [208]. Not only can these differences be detected with laboratory measurements, but also, humans have been shown to perceive these subtle differences in the duration of Dutch homophones [209]. This highlights that the human cognitive mechanisms for detecting and producing slight deviations in otherwise similar words are actually quite fine grained.

Zooming out a bit, one of the more surprising relations between language and music is that in certain contexts, they can be confused for each other. In fact, the mere act of repeating the same words or phrases seems to change the way they are perceived, leading to repeated spoken words sounding sung to listeners [210, 211]. This is referred to as the speech-to-song effect, and it suggests that it is not just the acoustic properties of speech sounds that impact the perception of them, but also, higher lever properties such as repetition and sound-similarity.

The speech-to-song effect is also found to be stronger in non-tonal languages (e.g. English, German, Italian) as well as when listeners do not understand the language stimuli. The authors suggest that perceived prosodic structures may be negatively correlated with the speech-to-song illusion. This makes intuitive sense because prosodic features are highly associated with speech, but not with music [212].

It may even be that certain phonological patterns (not just prosodic features) are positively associated with this effect. if this is the case it would imply that particular phoneme sequences may seem more "musical" than others. It also motivates questions about the degree to which rhyme is perceived as language versus music. Following from that, might the speech-to-song illusion be evoked using multi-syllable rhyming sets? Future studies can be conducted to address these questions.

These issues of similarity perception are particularly relevant to the phenomenon of rhyme where specific sounds from different words or phrases are in agreement and held constant while the words and sounds around them shift (often many times in rhyme sets). I will discuss this more in the following rhyme acceptability section. Given that the 'matched' components (or positions) of rhyming words are repeated, yet can have a great deal of imperfection, it raises many questions; How do similarity judgments change as the length of the sound sequences increase? In what ways might rhymes have to be similar in order to sound sung when said in succession? If rhyming pairs or lists can sound sung, it would provide important insight into the holistic understanding of rhyme perception, as well as the relationship between language and music perception.

3.2.3 Rhyme Acceptability

Rhyme is often thought of as a matching between the final vowel, stress, and coda of two words. This particular configuration of matching elements is called a perfect masculine rhyme, but there are at least 228 [39] other strict and named forms of rhyme, including feminine, dactylic, holorime, etc.. Historically, most rhyme analysis has focused on a few of these specific forms.

It has been suggested that both perception and production of rhyme involve "similar cognitive processes, including orthographic coding, graphic-phonetic transformation, phonological representation and phonological segmentation" [213]. Below I briefly review what has been found from the limited number of studies in this field.

There is a long established paradigm of detecting neural responses to ungrammatical elements of speech through the appearance of N450s [214], a particular component of event-locked EEG signals. Although there is debate concerning whether it truly represents ungrammaticality, conflict resolution and detection, or something else [215], if there are underlying grammar-like or conflict resolution properties to the process of rhyme, then one might expect that similar neural activations would be detected when rhyming expectations are violated.

Indeed, when comparing rhyming vs non-rhyming pairs, non-rhyming pairs elicit a larger N450 in subjects, indicating that non-rhyming pairs might require more conflict resolution or detection. This effect may be better described by a combination of expectation and consonance rather than grammar specifically, but it is interesting to note that there may be a grammar-like error awareness in stimuli for non-rhyming pairs. Furthermore, exploration of this N450 effect (using orthographic letters as the rhyming elements) suggests that the effect is more likely related to phonological properties than simply to orthographic (spelling) differences [216].

At what level of perception does the classification of rhyme patterns actually happen? Is rhyme perception global or local, or some combination? In other words, what are the cues used to perceive or detect a rhyme?; Is it similarity in distinctive features? Segments? Segment Sequences? Syllables? Rhythm (stress)? Perception? Higher order poetic structure? How much does the size (number of syllables) of the rhyme impact perception? I will explore some of these questions here, but many still remain unanswered. A focus on formalizing imperfect multi-syllable rhyme sequences and using the discovered patterns in behavioral and cognitive studies can begin to uncover these mysteries.

Perfect rhymes, especially of the conventional masculine and feminine forms, are widely accepted as matches, but imperfection in rhyme matches, and the variety of forms imperfection can take, make relevant the discussion of rhyme acceptability, and more broadly, sound similarity and pronunciation flexibility. Building on the various notions of similarity discussed in the last section, it seems that many kinds of similarity play a role in the classification of rhyme. There is also a modest body of literature that suggest that perception of half (or imperfect) rhyme may be impacted by degree of similarity [84, 46, 47, 48, 44]. Imperfect rhymes have even been used as a measure of phonological similarity [49, 50, 51]. Moreover, the speed of rhyme and non-rhyme judgements are much faster than imperfect rhyme judgements, which points to an increased level of uncertainty in making judgements about imperfect rhymes. That said, imperfect rhyme acceptability judgements do significantly increase when presented in the context of poems [51]. Given that most rhymes in the wild are not perceived in isolation (e.g. poems or songs), the increased acceptability of imperfect rhyme within poetic contexts should be taken into account in future studies on imperfect or multi-syllable rhyme similarity.

So, how much detail are humans attending to when determining the acceptability of a rhyme? One study indicates that "rhyme detection does not involve attention to segments... [n]or does it necessarily involve precise identification of the segment shared by rhyming words" [217]. Regardless, many poets seem to prefer certain half rhymes due in part to their perceptual similarity instead of their articulatory similarity [48]. Imperfection in rhyme also provides more practical options for word choice by loosening the matching constraints. However, "It [has been] shown that Japanese speakers do take acoustic details into account when they compose rap rhymes" [48]. Following on this line, rappers have been shown to exhibit detailed implicit knowledge of speech sounds that are not limited to their native language, but transcend it [45]. All of this seems to support the hypothesis that speakers have well-developed cognitive processes for and knowledge of the psychoacoustic similarity of words, which may be particularly relevant to rhyme judgements.

Much like the perceptual effects found in the elements of sound similarity (segments phonemes) [202], there is a preference for perceptual over articulatory similarity when people are making judgments of rhyme similarity [199]. Additionally, it has been shown that lexical decisions are faster when pairs of words are rhyming (perfectly), however, when pairs only partially rhyme, there is a much more limited amount of facilitation. It should also be noted that there seems to be a general ”yes” response bias in subjects when presented with a target word that rhymes with a priming word [218]. This can serve as a warning to experiment designers to use care when creating studies which use rhyming pairs and solicit yes/no responses. This bias is possibly also reflective of the amount of perceived agreement or resolution associated with high degrees of perceived similarity.

The cognitive mechanisms related to rhyme acceptability are associated with both articulatory and perceptual similarity. In addition, learning and phonological awareness may be related to the ability to detect or be aware of certain complex sound patterns and structures. This may make certain individuals more likely to find a rhyme acceptable (discussed in the following section on Expertise). Collecting multi-syllable rhyming judgments while holding various stress or phoneme patterns constant can help determine which aspects of words or sounds are used to make determinations that one sequence of sounds is similar to another. Furthermore, conducting these studies across expertise (rappers, lyricists, and neither), age, and language will allow us to consider the degree to which learning might impact the acceptability of rhyme judgements.

Although there have not yet been studies on the acceptability of multi-syllabic rhyme structures, they would be of particular interest given how large multi-syllable rhyme schemes have become in modern usage and how much imperfection (or flexibility) they often display.

3.2.4 Orthography

The orthography, or spelling, of words is also important to consider. Lindell & Lum show that the left hemisphere is more active when phonological similarity was higher, but that orthographic similarity was associated with both left and right hemisphere activity. Moreover, Lindell's experimental setup was able to determine that “both hemispheres are capable of orthographic analysis, [however,] phonological processing [was] discretely lateralized to the left hemisphere in males”, but not in females, for whom it is available in both hemispheres [219]. It is not clear if or how this finding is related to the historically male propensity for rhyming displays, but it highlights both the complexity of the cognitive processes used in phonology, and a potential neuro-physiological difference in the way males and females process rhyme, which could be a launching point for further investigation.

Finally, orthography also has an influence on perception that can interact with phonological information. Some studies have even suggested that orthographic information can interfere with phonological information or reaction-time tasks [220]. It may not be surprising, then, that processing for orthographic and phonological information are suggested to be separate neural mechanisms. If humans have to “recheck for phonological similarity when word pairs are visually but not phonologically similar," it is hypothesized that articulatory encoding, instead of orthographic ones, may avoid the slow down in reaction time that orthographic encoding introduces to rhyme judgements [221].

3.2.5 Priming

It is also useful to review some priming effects related to rhyme generation. In general, a priming effect refers to the notion that the presentation of a stimulus can influence responses to subsequent stimuli. A 1980 study measured semantic activation based on spoken production of synonyms, antonyms, and rhymes, finding no significant effects that involved cross-category rhyme priming. That said, in all cases, rhyme responses were associated with the fastest response times. Although the study explored lexical access, the authors note that pre-lexical access may be related to phonological access or sound production [222].

Extending from this, both rhyme and semantic priming effects are observed in phonological and semantic categorization tasks, suggesting that either kind of task activates both networks. Two further experiments were conducted within this study that used color-categorization tasks which required no phonological or semantic processing, yet priming effects for both were still found. This implies an automatic (and early) activation of semantic and phonological networks in a variety of contexts [223, 224]. These networks also seem to work together to facilitate lexical access, as shown by subjects when generating a missing word in sentence completion tasks, Their produced words show influence from interacting semantic and phonological features.

Both orthographic and phonological priming effects are also found in spoken word recognition tasks (even when orthography was not shown). However, these effects seem to emanate from different neural topological distributions, with phonological priming localized over the centro-posterior regions, and orthographic priming in more anterior regions. This indicates that phonological priming operates separately, and at a different level of representation, than orthographic priming [225]. Phonological priming is also shown to interact in a complex way with morphology, facilitating lexical decision speeds when the prime word only rhymes with the root of the target word (dough – snow-ed) [226]. This indicates that priming and similarity effects are relevant outside the context of strict end-rhyme paradigms.

Interpreting these findings through the lens of the stimuli-driven improvised rap can provide some helpful intuitions (e.g. Harry Mack Improvisation, discussed in Chapter 1 & 6). If "phonological facilitation ... operates prior to lexical access" [222], this may support the ability to rhyme improvisationally while reading orthographic stimuli. Moreover, the ability for the reading of stimuli and the production of rhyme to happen either concurrently, or at least in a way that does not cause destructive interference to fluency, is likely aided by the noted separation in neural networks that power phonological and orthographic processing. Finally, the early and automatic activation of both semantic, morphological, and phonological networks across a variety of tasks, provides evidence for a kind of priming that would be beneficial to the task of constructing semantically related and phonologically patterned lyrics on the fly.

3.2.6 Words

What are the neural processes underlying the production of non-rhyming words and how is it distinct from rhyming word generation?

There is evidence that word recognition depends asynchronously on phonetic cues (voice & manner) that are used in lexical access as they are perceived [227]. In normal speech, it is also assumed that word production is initialized from a "semantic base", a process involving morpho-phonological representation, syllabification, and articulatory gestures (vocal tract, etc.). Although disentangling effects in larger samples of language is difficult, there are theories of lexical access that offer methods for investigating both individual words and multi-word utterances. This is critical for understanding the rhyme phenomena under investigation at scales from individual words to more naturalistic (and longer) utterances [228].

Although silent rhyming and spoken word generation tasks activate perisylvian language regions (inferior frontal gyrus, posterior superior temporal lobe, and fusiform gyrus), rhyming generation activates more left hemispheric regions than simple word generation [229]. These findings can act as baselines for future studies that investigate both more targeted neural activity associated with various types of sound patterns, particularly in relation to different lengths and combinations of words (e.g. mosaic rhyme - rhymes across phrases - multiple words).

Speakers also modulate their pronunciation of sounds in words based on the context, hyper-articulating (over specifying) or hypo-articulating (under specifying) based on context [230, 231, 232, 233, 234, 235]. In general, speakers tend to hyper-articulate when there is competition between words or a potential for confusability. These kinds of online adjustments to the way humans naturally produce language sounds, allow multiple semantically, and even lexically equivalent utterances, to be expressed in a variety of ways by speakers [236]. This introduces additional degrees of variation in sound production that builds on top of the perceptual flexibility discussed in the previous sections on rhyme acceptability and similarity.

Focusing on sound processing generally, simple noise bursts cause increased activation in the primary auditory cortex of listeners, whereas speech syllables increase activation in secondary auditory cortices across hemispheres. Again, a functional lateralization was found, where pitch discrimination was associated with the right prefrontal cortex, and phonetic discrimination with the left hemisphere's Broca's area (generally implicated in speech production) [237]. It is suspected that the hemispheres may have become specialized over time for processing different kinds of acoustic information, with the left auditory cortical regions more optimized for processing temporal cues in speech and the right auditory cortical regions are more optimized for spectral, or frequency information [238]. But the story of speech and sound processing is hardly as simple as lateralization across hemispheres of the brain. Other studies have shown that both hemispheres are implicated in decisions on non-rhyming words (or semantically related words), but that rhyming judgements transfer information through the corpus callosum to the left hemisphere [239].

One fMRI study explored the neural activity associated with phonological versus semantic word generation in terms of rhyme, synonym, and translation tasks (bilinguals). Much like the Rouibah [223] study, the neural processes implicated in tasks in two languages (English & French) were similar and reliably overlapped in their activation of the 'left inferior frontal region' [240]. Neural activity during word and rhyme generation can be broadly described in terms of hemispheric activations. Rhyme generation has in fact been shown to predict "hemispheric language dominance" better than other neuropsychological paradigms. Not only is this academically interesting, but also, can provide a non-invasive diagnostic screening phase for a variety of neurosurgical interventions [241]. For example, the Wada test is an invasive, but commonly used pre-surgery test to identify hemispheric lateralization of various language and motor functions. This invasive test could be replaced by a more robust classification of hemispheric language dominance based on a simple word rhyming diagnostic test [242].

3.2.7 Improvisation

Music

Although improvisation is colloquially considered a spontaneous behavior, it is built on the foundation of prior knowledge. Specifically, it requires a balancing of short term ongoing attention to sound sequences and the automatic retrieval of musical patterns from long term memory [243].

Musical improvisation is associated with activity in both the default mode network (DMN) and the executive control network (ECN). These networks are implicated in a variety of cognitive processes, but relevant to the tasks at hand, DMN is associated with mental simulation and self-referencing [244, 245], while ECN seems to be related to goal oriented tasks and the evaluation of ideas. These networks are generally negatively correlated [246, 181, 247], and may be at the heart of creative ability [181, 248]. When these networks are deployed together, associations with increased creative performance have been observed [249].

Some have also suggested that creativity, which is highly correlated with improvised behavior, can be construed as an identifiable mental state. This was determined using EEG to test for "distinct patterns of neural activity" during Alternative Use Tasks [250] and Consensual Assessment Techniques (ways to measure the divergence of use cases or creativity of artifacts, respectively) [251]. It has even been shown that in less familiar settings, improvisation becomes more predictable, as measured by entropy and conditional entropy, pointing to the utilization of more predictable prior knowledge that underlies this complex skill. [252].

In terms of neuro-anatomical regions, musical improvisation is "consistently characterized by a dissociated pattern of activity in the prefrontal cortex" [253]. Furthermore, increased activation of the perisylvian language regions seems to be associated with collaborative improvisation involving two musicians. Activation of this language region in music may point to leveraging the cognitive processes used in musical 'syntax' [254]. In comparison to classically trained musicians, improvisations performed by trained Jazz musicians reliably activated certain regions (Broadman's Area 7) thought to be related to altered states of consciousness (hypnagogia/sleep). Similarly, brain regions active in the production of jazz music were less efficient, less clustered, and less synchronized in the right-hemisphere than during production of classical music [255], again indicating differences in the cognitive mechanisms involved across disciplines.

It should be noted that the neural correlates of both musical composition and improvisation vary across individuals and are not particularly robust or widely accepted (and subject pools for experiments have been small). Furthermore, the study of improvisation has largely been limited to jazz and more specifically, piano [256]. Lu et al. take this into account and include composition tasks for instruments that the composers do not know, finding similar results as above (previous paragraph). Many more studies using various kinds of composition, performance, instruments, and improvisation (e.g. improvised rap, beat-boxing etc.) will be needed to tease apart more reliable form-function relations in the brain.

Rhyme

As limited as the understanding of improvised musical production is, improvised lyrical production is even less well studied. A recent characterisation of improvisational rhyming strategies focused on the practical use of end-rhyming strategies [257] where target words are chosen and rhymed with. This is in line with characterisations of what freestylers are doing; "always looking slightly ahead, like when reading music" - Harry Mack.

Improvisation is, in general, associated with contexts where stimulus-independent action occurs without being monitored consciously or actively. However, it appears that in lyrical improvisation, as opposed to musical, there are distinct associations between regions that are thought to couple intention and action (whereas musical improvisation is dominated by action/motor). This suggests that, for certain aspects of production in lyrical improvisation, normal "executive control may be bypassed, [allowing] motor control [to be] directed by... motor mechanisms" [180]. These findings are consistent with a recent study that incorporated various kinds of improvised performance, including improvised rap, noting that both cognitive control (executive function) and motor planning can facilitate "seemingly unconstrained behavior" [181].

These findings suggest that two intertwined processes are involved in improvised lyrical production. On the one hand, there is an automatic, reactive process, where complex language might produced, but without the patterns becoming the focus of the speaker's attention. On the other hand, there is a more intentional process, where a target word is selected, becoming the temporary focus of attention. The target word is often silently recited in order to introduce its sounds to the phonological store. Target words are often not uttered immediately, but rather, are rhymed with first, only revealing (speaking) the target word as new target word is chosen. And then the process repeats. This intentional process is consistent with the end-rhyming strategies mentioned above (target words) [257] and with the pluralist notion of expertise introduced earlier – the "weav[ing] together [of] automaticity and higher-level cognition..." [179].

Comparing the elements of language and music can be a revealing exercise. On a piano, for instance, each octave has a patterns of 7 white keys and 5 black keys, this pattern repeats across the keyboard. All possible notes are apparent (visually and physically) and can be pressed in succession or in parallel to produce musical patterns. In language, however, the sound elements of interest (phonemes) are not apparent all at once (as on a piano). In addition, these sound elements are tied to lexical items (words) which are also not apparent all at once. Without direct access to the elements and patterns of language sound, in lyrical improvisation, it seems that lexical items become targets of focus. These target words are intrinsically associated with sound patterns, which can then be repeated or patterned to create various language-based music structures.

While the study of musical improvisation has been increasingly utilized in therapeutic applications [regulating mood, etc.] [258, 259], there is also work suggesting that some linguistic skills are complementary to musical improvisation and can provide therapeutic benefit, skills such as "(1) attending to sound and music; (2) using descriptive language about music; and (3) facilitating verbal processing of improvisation" [260]. These benefits are not necessarily specific to improvised musical production, and may indeed overlap with the space of lyrical improvisation.

3.3 Language

- How has rhyme been formally investigated?

3.3.1 Stress

Prosody is the study of intonation, tone, stress, and rhythm. These elements are often related to the syllable or individual segments, but exist as what are called suprasegmental elements. Stress, and combinations of stress patterns, encode rhythm and melody and are particularly relevant for rhyme constructs. Some modern approaches to English stress in linguistics are built on grid-based theories of meter and phonological relations [261] (French [262]). More wide-ranging computational approaches have shown that English words, for example, tend to follow 9 different syllable/primary-stress patterns, and that these words are not evenly distributed across the lexicon [263].

Figure 3.4 shows stress pattern permutations from the lexicon, named patterns in the English. Note that yellow-highlighted dashes denote stressed syllables, while the u-shaped symbol denotes an unstressed position.

Because of the limited number of stress types (stress, no stress, and secondary stress - not depicted here), the set is limited enough in size that each of these patterns can have a name and be the subject of academic study. Although the study of individual vowel patterns has not followed in this same enumerative tradition, it would be possible to perform a similar enumeration, though the sets would be much larger. For example, using the 15 English vowels in CMU pronouncing dictionary, there would be up to 15x15=225 two-syllable vowel patterns, rather than the 4 2-syllablle stress patterns seen in Figure 3.4. Exploring the distribution and use of multi-syllabic rhyme at a deeper level may require a more explicit and comprehensive investigation of all possible vowel sequences, following in the tradition of stress and prosody.

Describing patterns of stress in language, and more specifically, how they apply in the context of poetics, is critical for understanding the patterns and variation in instances of rhyme. Various defined forms of rhyme incorporate stress pattern requirements, both in terms of their quality (stressed or unstressed) and their position, often focusing on the importance of matching sounds with particular stressed positions. For example, forms like iambic pentameter define large alternating stress patterns, where the terminal (rhyming) positions of each line are both stressed and expected to rhyme.

Probabilistic regularities are important for learning. And although these regularities may be necessary for learning, they are not sufficient, as evidenced by the paper "Learning English Metrical Phonology, [...] Probability Distributions Are Not Enough" [264]. Theoretical, experimental, and behavioral approaches are all critical for a comprehensive understanding of stress and phonology. There are complex and subtle components of the learnability of language, but few paradigms exist where stress and repetition are so macroscopically observable as in the case of multi-syllable rhyme. Yet, it is a realm where academic study is almost absent. This phenomenon, in the context of rhyme, demands more exploration.

3.3.2 Sequences

The distribution of individual sounds in a language is important for understanding the patterning present in a phenomenon like rhyme. But as seen from Figure 3.4, it is not just individual phonemes that are important. Sounds do not exist in a vacuum, but rather, in the context of other sounds, making their transition probabilities and natural clustering important aspects of linguistic inquiry. Much like the unique stress patterns shown in the previous section, unique vowel sequences can be used to investigate the underlying vowel sequence vocabulary of multi-syllable rhyme. In contrast to stress sequences where only 2 states exist (stress, unstressed), for vowel items, many more possible states (vowels) exist. For instance, for 4-syllable sequences of stress 16 (2x2x2x2) unique possibilities exist, while for 4-syllable sequences of vowels 50,625 (15x15x15x15) unique possibilities exist. Of course, even more unique sequences are possible if both stress and vowel pattern are considered. It should be noted that many of these sequences will be covert in practice.

Furthermore, the internal transition probabilities of syllables and sounds used in multi-syllable rhyme (phonotactics), as well as the possibility space of matches (neighborhood density), both demand further exploration. For example, it is known that there is a significant correlation between the vowel and coda of a syllable in English (vowel-coda -¿ the rime), but not between the onset and the vowel [265].

Different individual sounds in language also occur at different rates. For example, in Figure 3.5 below, the blue bars represent the average probability of occurrence of each of 15 vowel phonemes from the Brown corpus [266], and the red bars represent the distribution of those same vowel sounds from Shakespeare's Sonnet 29, pictured in Figure 3.5.

Of course this is a small sample, and it will be expected to vary from the population mean, but it can reveal the shape of sample distributions (languages, dialects, subcultures) and help identify outliers in individual samples. Simple comparisons of unigram distributions like this can also tell us about systematic deviations of different samples from the population mean in terms of sound usage and how they change over time. Analysis of bigram and trigram sound sequences, etc., can provide even more fine-grained descriptions of the variation across sound sequences in different forms of rhyming language. One could begin asking questions like, do rappers have some learned vocabulary of sound sequences? Does that vocabulary have a grammar? Do they employ more consistent use of particular kinds of sequences?

Phonotactics

Phonotactics is the study of the possible legals transitions or combinations in the context of sound elements within language, particularly phonemes. Phonotactics is an important domain of inquiry, both for descriptive linguistics efforts and for better understanding the learnability/acquisition of language. Here I briefly review some of the roots of phonotactics that will provide a foundation for understanding larger rhyme patterns.

There are conditional probabilities associated with the likelihood of transitioning from one sound to another. Although there have been shown to be long distance relationships between vowel sounds [267, 268], the most reliable and noteworthy transition probabilities exist in adjacent sounds, as conditioning of sound in language tends to be a short range phenomenon. Rhyme patterns are a rare case of reliable and long range (often periodic) vowel, stress, and consonant relationships in language.

Although there are a handful of studies on the phonotactics of rime [269, 270, 271, 272], no studies on the phonotactics of rhyme matching have been conducted. Also, this might be simplified even further by considering vowel patterning alone to represent a significant portion of the macroscopic structure of rhyme. Describing vowel transition probabilities of natural and rhyming language would be an important starting point for motivating a more granular understanding of the phonotactics of complex rhyme. The more that multisyllabic rhyme has become a prominent form, the more that these transitions between and within sequences of sounds demand investigation.

Considering that many kinds of phonological patterns occur across genres with different traditions of patterning, it is also of interest to explore how much variability and structure there is in sound sequences across genres and sound sequence types.

As I will discuss in Chapter 5.1, it seems that some lyrical genres have longer and more predictable sound sequences than others. In a recent study I compared the conditional entropy of sequences of phonological patterns in lyrics and found that, in general, Battle Rap and Sonnets maintain noticeably lower entropy [higher predictability] than other genres across vowel and stress sequence sizes, respectively [273].

Neighborhood Density

Neighborhood density measures the number of “sound neighbors” that a given word has. For example, CAT-BAT live in the same sound neighborhood, as they only differ by the first sound. Because they only differ by one sound, they are also called minimal pairs. Some words have a high neighborhood density (e.g. CAT), meaning that there are many other words that are distinct but are still very similar on the whole (one or two sound segments differ e.g. HAT, SAT, CAP). Words with low neighborhood density exist in a less populous ‘sound space’ (e.g. ”echolocation”, ”hippopotamus”).

Neighborhood density has impacts on learning, perception [274], and production [275] of words. In general, humans perceive and produce high probability phonotactic sequences faster than low probability sequences, regardless of their neighborhood density [274, 276, 277], even when those sequences comprise only partial words [278]. This demonstrates the separation between phonotactic and neighborhood density effects, and that their cognitive effects are about more than just word elements, drawing from knowledge of sound distributions rather than just lexical distributions.

In spoken language, vowels are sometimes ‘reduced’ or ‘weakened’, indicating some (predictable) kind of change in the acoustics of vowel production. This can include changes in stress, sonority, duration, loudness, articulation, or position in the word. While this is a broad phenomenon, vowel reduction has been shown to be positively associated with words that have high neighborhood density [279]. In both English and Dutch there is only a weak correlation between high frequency words (written and spoken) and high neighborhood density. Furthermore, higher neighborhood density has been shown to be more strongly associated with word bi-grams than with unigrams [280]. This notion of high neighborhood density and reduction in relation to multi-syllable rhymes is especially relevant given that word/phrases in low-neighborhood density areas would seem to have more degrees of freedom for sound change (mutation, insertion, deletion, substitution, etc), given that it is less likely for changes in low neighborhood density words to result in collisions with other distinct word forms. For example, most sound changes in the high neighborhood density word ‘cat’ would change it into another distinct word (kit, cut, kate, cap, cab) whereas similar changes (minimal pairs) to the word ‘catatonic’ (kitatonic, cutatonic, capatonic, canatonic) are all still potentially perceivable as ‘catatonic’, even with limited contextual clues.

Learning does not happen in a vacuum, but it is, in fact, related to the complexity of the signal being learned, as in the case of phonological complexity and the learnability of language [281]. Interestingly, a recent study has shown that statistical learning is actually facilitated when the signals being learned are in a particular range of complexity (as measured by entropy efficiency: observed entropy / maximum entropy). Sets of stimuli with entropy efficiency similar to that of language (Zipfian distributions) are learned better than sets of stimuli from other distributions, even better than distributions that are *more* predictable than Zipf, and intuitively, should be easier to learn [282, 283]. This suggests that many of the previous studies that have explored learnability and language complexity by using uniformly distributed stimuli sets may be investigating a more artificial or laboratory invented phenomenon that does not well reflect the distribution of stimuli in more naturalistic learning environments.

Age can also be a factor in how humans pick up cues from the environment. For example, a developmental study found that high vocabulary five year olds show neighborhood density effects in a rhyme task, whereas children at low vocabulary-ages (but still 5 years of age) did not [284]. More generally, neighborhood density effects are observed in spoken word recognition tasks across age groups (preschool, elementary school, adults). Although word recognition is facilitated in early-acquired words from sparse neighborhoods in children, adult word recognition was facilitated by later-acquired words [285]. In addition, the recognition of words by children was less dependent on the position of sounds in words (initial, medial, final) than it was for adults [286]. That said, it is not clear how dependent the recognition of words is on the specific order of sounds. For example, a recent study showed that recognition of words (lexical access) in adults was not as strictly dependent on the order of sounds in words as some models suggest. They used phonemic anadromes, (sub–bus or pat–tap) to show that words with their letters rearranged (e.g. tap) actually do facilitate activation of the related word (e.g. pat) [287]. Finally, reaction times and accuracy in reading tasks improved with age, indicating that certain aspects of phonological processing may depend on different neural systems or developmental time courses [288].

In relation to variation or imperfectness in rhyme, these studies suggest two distinct arenas of flexibility. First, in high density environments, vowels are more prone to reduction, resulting in common and predictable alternate pronunciations that may be leveraged in the matching of rhyming words. Second, in low density neighborhoods, arbitrary sounds may be changed or removed without creating confusion with other words, since so few similar sounding words exist. This provides additional flexibility for imperfect or slant rhyming for low-density words that can be much less constrained than words from high-density neighborhoods. These two scenarios indicate that facilitation of slant or imperfect rhyme patterning may come from both high and low density environments.

Questions concerning exactly what does impact rhyme cognition and learning are largely outstanding as most forms of rhyme have yet to be investigated in any theoretical context, such as statistical learning or learning theory. However, vocabulary acquisition and phonological neighborhood density networks are implicated in the processing of rhymes. Together, these studies also suggest that complexity, statistical regularities, age-of-acquisition, order of exposure to items, and sound position (within words), all impact the learning process and should be taken into account when considering underlying cognitive mechanisms.

3.3.3 Poetic Analysis

Most analysis of poetry takes a relatively holistic approach, analyzing particular works or authors. However, there have been a couple of efforts that computationally explore the space of possible poetic devices. For example, despite the common belief that there is not much possible variety in the end-rhyming patterns of sonnets (e.g., ABABCDCD), Höft demonstrates otherwise. He explores the space of patterns used by individual authors, and the space of all possible patterns, finding more than 4.3 million possible variations on sonnet end-rhyme schemes (w 14 lines) [289].

In Chapter 4.2, I outline the space of all possible matches for various common poetic devices. This study used simple templates that operationalize the positions of syllabic components which must agree across any two words in order to be considered a match in terms of assonance, consonance, masculine rhyme, feminine rhyme, etc.. [290].

These approaches are informative descriptively, but they also provide a lens through which the availability of patterns in language can be explored compared to their perceived availability or their frequency of usage in the wild (e.g. poems, lyrics).

In addition, rhyme style and language features have been explored in the context of text (genre) classification. Features include rhyme, part-of-speech, bag-of-words, and various combinations of those features are shown to be productive in improving text classification and or model simplification (dimensionality reduction) [291].

Stylometrics

Exploring the elements of phonemes, stress, and their distributions is surely important for a comprehensive understanding of rhyme, but it is also important to review the larger poetic and linguistic context that rhymes most often exist within. This is often done using stylometrics, the quantitative study of linguistic style. There have been a number of efforts to describe the features of poetry, from manual analysis to automated rule-based approaches, to various forms of visualization [53, 54, 55, 56]. Often the aim is to identify a broad enough base of linguistic features to successfully and automatically classify or compare works or authors; relevant features include vocabulary, poetic devices, meter [292, 52], semantic density, prosody, concepts, and rhyme schemes [53]. This line of investigation has much in common with traditional authorship identification techniques where a large swath of linguistic features are used to distinguish between authors. Famously, Shakespeare and the The Bible have been the target of much of this investigation as it is suspected that these techniques can help discriminate the author(s) behind their creation [293].

Visualizations

Although direct feature visualization can be difficult in texts like The Bible, which involve largely non-rhyming prose, more poetic works with added phonological constraints tend to be more amenable to effective visualization. The most notable work on visualizing rhymes uses vertical vectors of color, as in Figure 3.6, to denote the ABAB rhyme scheme patterns across sonnets [289, 294, 295].

This approach is wonderful for illuminating the variety/distribution of possible 1-dimensional ABAB rhyme patterns, but, like most research on rhyme until recently, it assumes that rhyme is limited to (or well represented by) single end-rhyming syllables. Additionally, the colors here represent the pattern of sounds relative to a single poem, rather than the sounds themselves. In other words, A is always blue and B is always green, but the words/sounds that A or B represent are relative to each poem. Therefore, the analysis by Höft is about the abstracted end-rhyme patterns, and in many ways ignores their sound content and related variation.

Although visualizing patterns of end rhyme in sonnets can be informative, even in the realm of phonological patterns there are many kinds of representations and information in natural language that should be considered independently. Recently, a multi-level visualization scheme for poetry was described that allows for the display of phonological elements of poetry [8], shown in Figure 3.7.

The approach highlighted in Figure 3.7 begins to visually identify how complex and multi-layered the patterns and representations of sound can be. Going from words to phonemes (1), then fitting them into a metrical grid (2), and representing them as syllables (3) or as orthographic patterns (4).

Finally, in an attempt to focus on both rhyme identification and imperfectness, graph theoretic approaches to rhyme similarity allow for a network based separation of rhyming words into clusters based on feature similarity and occurrences in rhyming corpora. It works based on automated pruning of these networks, resulting in relations between words (nodes) that represent various linguistic relationships. This approach has many benefits, including computational tractability, discovery of partial rhymes, as well as historical pronunciation detection. Below, I highlight the resulting graphs of rhyming clusters to demonstrate this approach to both classifying (clustering) and visualizing rhyme relations [9]. This is not specifically an approach to rhyme visualization, but rather, a graph theoretic one that results in visually digestible networks of rhyming clusters.

As shown in Figure 3.8, these graphical approaches can build networks of words, and then segment them into clusters along rhyming relations. Based on the particular pruning approach applied, they can also identify half (partial) rhymes, as well as sight (spelling) rhymes. Automated clustering of words based on sound similarity and co-occurrence in written samples is important, and could be applied to a variety of language data to identify stable word relationships. Although it was developed with the idea of helping to identify historical pronunciations of words (where historians are not sure how words were truly pronounced), it can also be used to identify the pronunciation of words in more modern contexts where slant or imperfect rhyme have been utilized. Co-occurrence measures allow us to overcome the limitation that many rhyming pairs drawn from text (partial rhymes, uncommon pronunciations) cannot be identified based on conventional pronunciation lookup tables alone.

3.3.4 Rap Analysis

Given the abundance of rhyme in general, and multisyllabic rhyme in rap, it is surprising that so little computational work has been done in this realm. There have been a handful of efforts to identify the notoriously complex sound patterns in rap lyrics, however. Unique spelling conventions have been explored [57], as well as relative vocabulary size, showing that one artist (Aesop Rock) uses many more words than other well known verbose rappers (Eminem, Andre 3000, Nas, Twista) [58]. But this level of spelling or lexical diversity says little about the most defining features of rap lyrics: flow and phonological patterns.

Much like groove in music, which involves rhythm, meter, and musical structure [184], flow is an aspect of rap lyric delivery that encompasses both metrical and articulatory techniques. Metrical techniques include the number and placement of rhyming and accented syllables as well as the relationship between syntactic units and metrical elements. Articulatory techniques include articulation of consonants and alignment of syllable articulation with the beat (earlier or later or simultaneous) [25]. This approach frames the analysis of rap in terms of the musicality of flow rather than directly as poetic or linguistic content. Two different approaches to analyzing flow in corpora found a great deal of evidence of end rhymes on or near fourth beats of each measure [68, 296, 297]. Much more analysis is needed to uncover the musical and phonological features of rap flow.

There have also been a limited number of attempts to identify the complex rhymes contained in rap lyrics [11]. In part, this dearth of research is due to the lack of tools for noticing or encoding these patterns in the first place (which I discuss later in this section). But even once patterns are identified, there are no robust frameworks for linguistically or computationally analyzing/characterizing rhymes, especially when they are imperfect. Here, I review the approaches that have been taken to identify phonological patterns in rap lyrics.

Complex rhyme is composed of repeated stress, vowel, and consonant elements. Because vowels occupy the sonorant (i.e. singable sound) nucleus position of the syllable, they are often the most prominent element of analysis and the most highly weighted feature of rhyme. An algorithm that counts rhymes on the basis of vowels alone (assonance) was developed in 2015 which characterized artists by the average length of vowel patterns observed in their lyrics (Rhyme Factor or average length of vowel patterns). The approach is simple: it encodes lyrics as just vowels and then iterates, word by word, to find the longest sequence of vowels that ends with vowels from one of the previous 15 words. This rule based approach seems to overlap with many rhyming patterns and provide some amount of intuitive multi-syllable rhyme identification [298]. Recognizing just how underexplored poetic devices are in poetic works, more recent work attempted to identify assonant clusters by localization across features of vowel closeness and backness [299].

Complex rhymes are often characterized not only by their multi-syllabic structure, but also by their imperfectness and their location in poetic works (internal overlapping as opposed to merely end-rhyme). There have been some relatively successful efforts to automatically identify rhyme schemes in the noisy data of rap lyrics, including applications of frequency scoring [300], expectation maximization [59], and Hidden Markov models [301]. The available data is messy and unlabeled, however, making automatic identification difficult. Yet, Addanki Wu, for example, were able to achieve rhyme classification scores of 35.81% for precision and 57.25% for recall. There is certainly room for improvement in all of these computational approaches.

3.3.5 Lyric Generation

While there are not yet comprehensive linguistic or computational descriptions of complex embedded rhyme schemes, a few attempts have emerged to reproduce these patterns in machine generated lyrics [62].

There are 3 broad approaches to generating lyrics, including templatic, compositional, and end-to-end machine learning. In the templatic approach, some underlying template of stress or sound patterns can be identified, and then some form of lyric generation is superimposed on top, with the requirement that the resulting output meets the phonological (or other) constraints set in the template [63].

Some compositional approaches integrate knowledge and semantic components [64], while others employ intelligent mixing and matching of newly or previously generated lines of lyrics, which have been shown to select next lines at a rate 50x better than a random baseline, and 21% better than human experts [65]. Furthermore, support vector machines [302], recurrent neural networks (RNNs) [66, 303, 304], and Long Short-Term Memory (LSTM) models [305], have all been used with limited success in generating poetry or rap lyrics, usually with compositional or end-to-end approaches. There has even been work that combines an RNN with a finite-state machine that deals specifically with the rhyming words and meter [67]. Finally, some poetry generation has been based on evolutionary algorithms using many agents to search in parallel across the vast space of possible continuations (stochastic hill climbing), which then evolve over many generations to optimize constraints like meaningfulness, poeticness, and grammaticality [306, 307].

Lyric generation efforts could be improved with a better understanding of rhyme matching options (and other phonological patterns), both in terms of multi-syllable and imperfect patterns. For example, knowing which kinds of imperfect rhymes are more perceptually convincing, or which mosaic (across word) rhymes are available, would greatly benefit lyric generation efforts by carefully expanding the space of possible rhyme matching, while adding more vocabulary diversity to the semantic and grammatical constraints of poetry writing.

Example Templatic Generation

Language is filled with long-range relationships, a notoriously difficult-to-capture dimension of language models. LSTM models and transformer models have become popular for their abilities to better capture long-range relationships in language. There have been various implementations of models like these which are trained to produce lyrics in some way. This usually involves retraining the model on relevant lyrics (to implicitly capture rhyming patterns), or building in some knowledge of rhyme or sound similarity. Here I use a state-of-the-art transformer model, GPT-2, to automatically generate language that matches a customizable sound template, a mixed templatic approach to lyric generation.

Below is sample output from a GPT-2-VowelTemplate model I created. It enables one to define a vowel template, and then uses GPT-2 to generate text while also matching the template. This approach allows for generation of text with arbitrary and controllable vowel sounds.

The approach here uses GPT-2 to generate multiple possible next words. The underlying vowels of each potential next word are extracted and checked for agreement with the next required sound in the template. If a viable word is found, it is added to the output chain. Sometimes all the vowel constraints of a given output cannot be met using this approach, these are considered 'failed' attempts.

An example context is shown below. These contexts are used to guide the topic or theme of each generated sample is an integral. The context will compel GPT-2 to generate text that is somehow related to practicing to become an expert. The vowel wrapper around GPT-2 then forces language output to adhere to a phonological template.

In Figure 3.9 below, I define a minimal vowel template that is used to constrain the output of GPT-2. If no vowel phoneme for a given syllable is specified, any vowel can be matched (asterisk *). 4 example outputs from a single template and prompt are shown in Figure 3.9.

3.3.6 Summary

In Chapter 3, I summarized the approaches of poetic analysis, both in terms of stylometrics and visualization techniques. Then, I focused on the specific case of linguistic analysis of rap lyrics, a domain where complex multisyllabic rhyme patterns dominate and, to a large degree, guide the phonological structuring of the genre. Finally, I discussed approaches that have been used to generate lyrics that include rhyming components. All of these techniques would certainly benefit from a better understanding of the dynamics of complex rhyme.

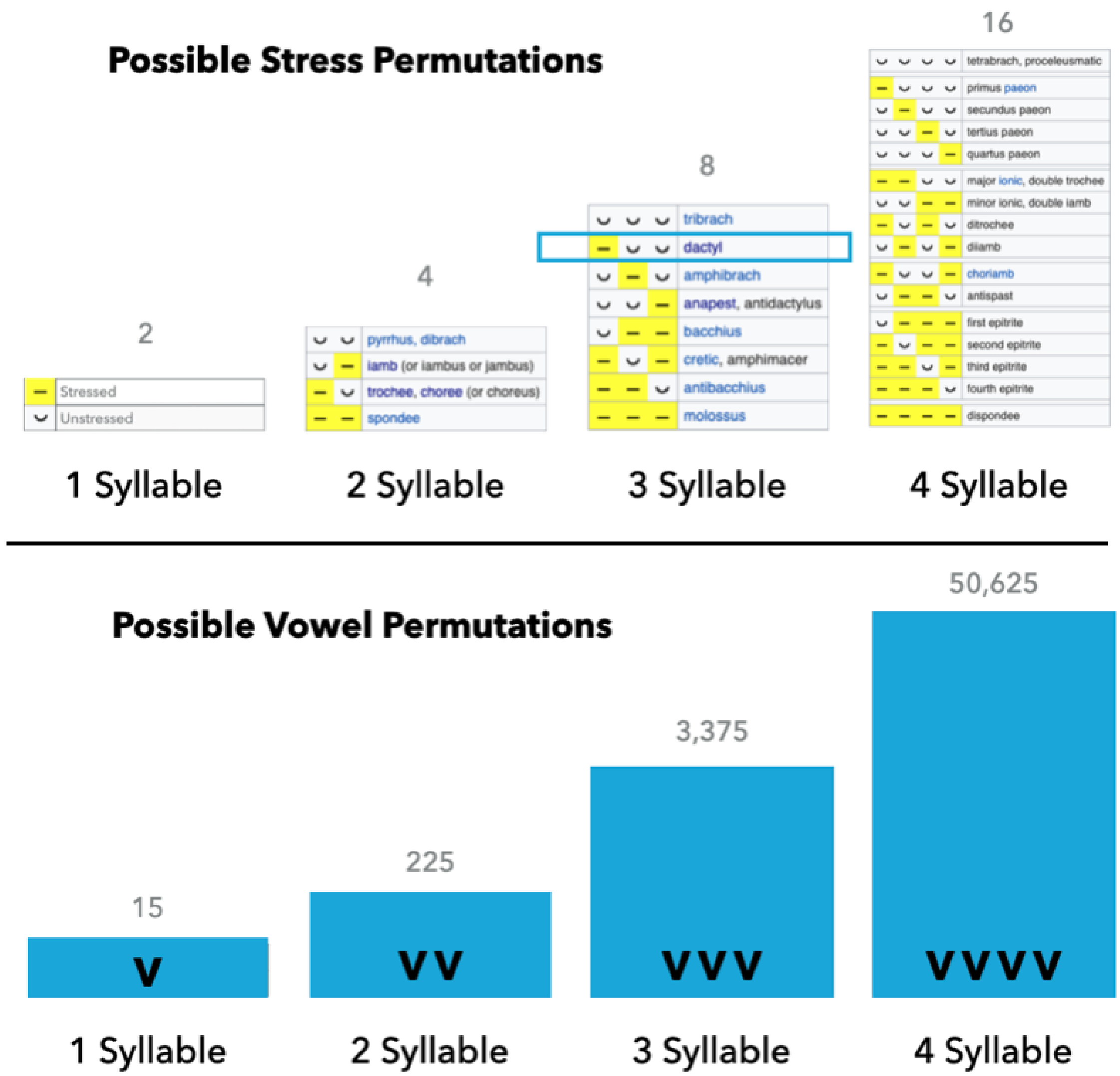


Figure 3.4: Charts showing all possible stress pattern combinations, also known as metrical feet up to four syllable sequences. Stress Images from Wikipedia (Foot (Prosody), 2021).

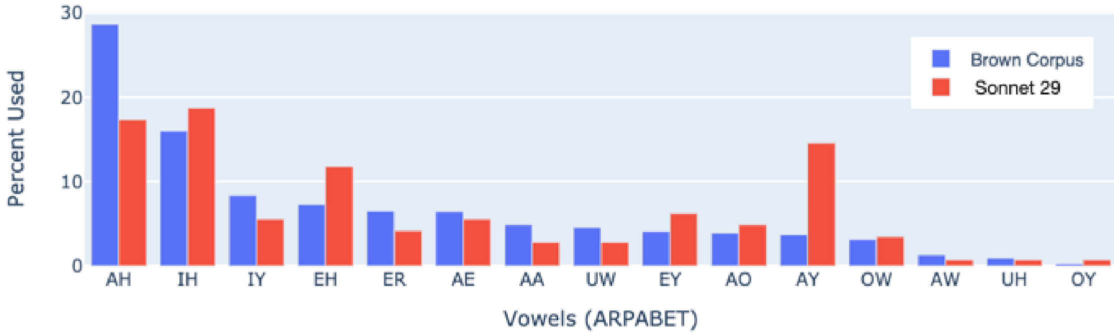


Figure 3.5: Rates (in percent) of vowel occurrence across the widely used Brown corpus are shown in blue and are compared to rates for a sample (Sonnet 29) in red. Note the disproportionate use of the /aɪ/ vowel in this poem (eyes, cries, my, etc.), which is prevalent within its rhyme schemes.

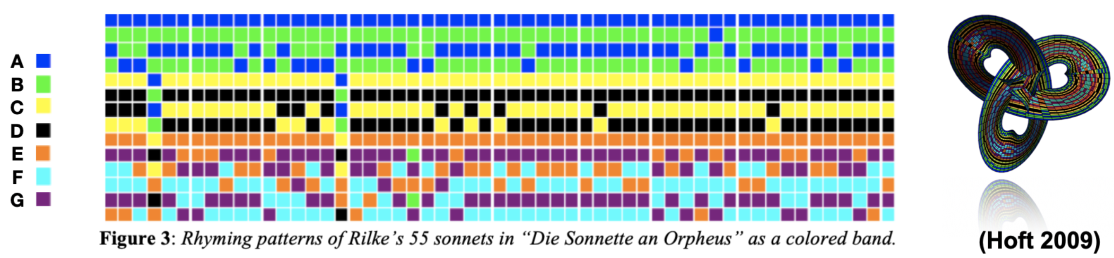


Figure 3.6: Each column represents the pattern of end-rhyme from one sonnet (each with 14 rows). The first column (sonnet) begins with ABAB, the second and third (columns/sonnets) begin with ABBA, and so on. The image on the right is an artistic projection of this same data.

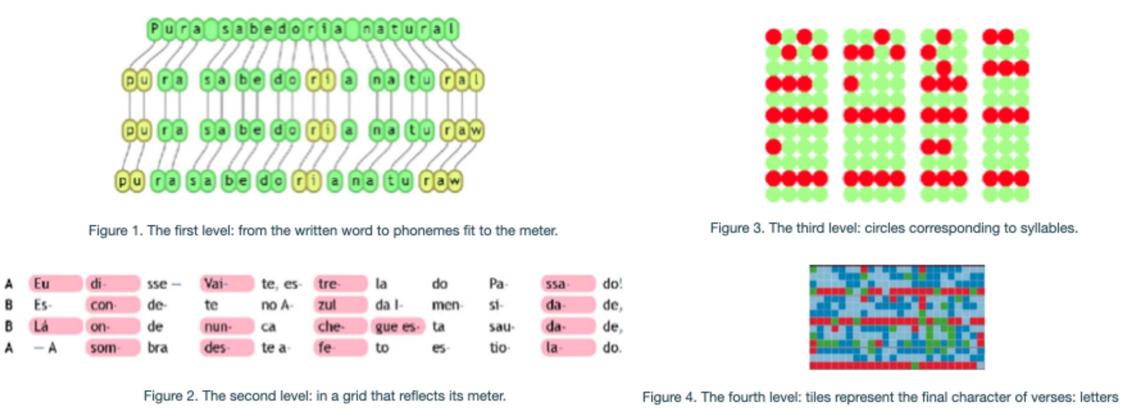
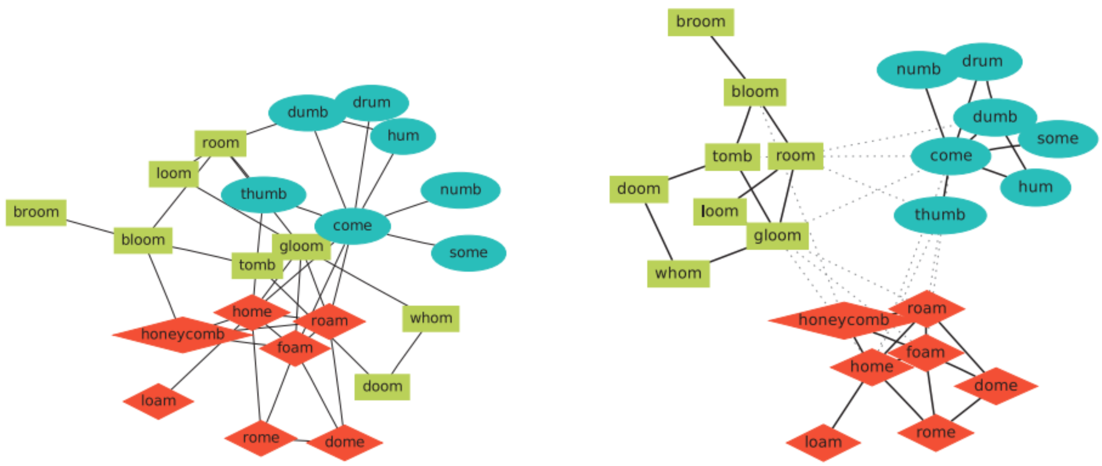


Figure 3.7: 4 different types of visual representations for poetic sounds from [8].



(a) Unpartitioned (b) Partitioned

Figure 3.8: In the above graphs, nodes are words, and observed rhymes are edges. The degree of similarity is implicitly gathered from poetry with annotated rhyme schemes and represented here as edges. The color and shape of nodes are visual conveniences for distinguishing between tight clusters of nodes. [9]

Context (Prompt)

"The truth is, you have to practice to become an expert at anything."

Vowel Template

,,*,*,*, AE, AE, IY
,,*,*,*, AE, AE, IY
,,*,*,*, AE, AE, IY
,,*,*,*, AE, AE, IY

* = Any Vowel

Poem 1

You can't quit. You put **an axe deep**
in an old tree, which **has an ee-**
rily deep root system that **has an ee-**
rily slow progress and no **chance at speed**

Poem 2

What gets lost in the **math class we**
read and the math club **that hangs each**
year? Annie Wilke taught **math, math she**
didn't think she could **grasp at least**

Poem 3

One of the first things **that has me**
impressed, when I read **an ad he**
wrote for a job, one **that had me**
as an agent, was **that that he**

Poem 4

But I would start with **that fact here**
on my own for my **fans that need**
to know to learn the **craft as we**
go Q: Do you come out **as black these**

Figure 3.9: Example output from GPT-2 with a vowel template wrapper. 345 million feature GPT-2 takes a few parameters, including temperature (relatedness), batch size, and a context or prompt.

Chapter 4

Overt Patterns

In Chapter 4, I focus on data that are identified to belong to poetic devices. First, in order to better understand the landscape, I document all possible instances of perfect poetic devices across all dictionary words. Then, I transition to considering imperfect rhymes across 500 years of poetry in English. Finally, I turn my focus to more formal analysis, examining sets of multi-syllable rhymes, something that has not yet been seriously considered in the literature.

Questions Covered:

- (Section 4.1) How many perfect rhymes are possible?
- (Section 4.2) How perfect is perfect rhyme?
- (Section 4.3) How can imperfect multi-syllable rhyme be quantified?

4.1 Perfect Poetic Devices

How many perfect rhymes are possible?

How many perfect rhymes are possible in English? In other words, what is the shape of the entire phonological vocabulary at the level of perfect poetic devices. This study explores the availability of common, perfect, phonologically driven poetic devices like rhyme, alliteration, and assonance. In addition to providing a way to enumerate the space of possible poetic devices, this effort offers related metrics such as entropy and network size. I show that certain devices, such as alliteration and stress, have many fewer possible unique patterns, while providing dramatically more possible word matches than other devices like rhyme, assonance, and consonance. These results are discussed in terms of their cognitive and poetic implications.

Traditionally, the study of poetic devices and artistic patterns in language has focused on the analysis of creative works themselves. In this section, I instead focus on the language resources that scaffold and constrain their creation. Exploring the possible space of common poetic devices across the English dictionary, I examine the baseline frequencies of patterns like assonance, consonance, masculine, and feminine rhyme across unique words. How big are each of these networks? In other words, how available are these poetic devices? And how many matches are there for each of the incarnations within a poetic device? This investigation is motivated by both a desire for more data driven descriptions of creative linguistic phenomena, and a desire to better profile the landscape of possibilities that bound the production and perception of these patterns.

Understanding the underlying networks of poetic devices, and their availability, can also provide a statistical learning based lens through which these devices can be interpreted. Phonotactic explorations such as this, which study the frequency, probability, and constraints of transitions in the context of sequences of sound elements, are an important domain of inquiry, both for descriptive linguistics efforts and for better understanding the learnability and acquisition of language.

4.1.1 Data

For this analysis I use the CMUdict [308], a phonetic dictionary, as the corpus of unique English words which provides ARPABET representations of each word's pronunciation. Below is an example of how words are encoded. There are a total of 120,413 words in the current dictionary.

As I will discuss further in the methods section, for each word, I extract the components of words that represent their potential use in poetic devices, highlighted in green in Figure 4.1 to create a new, more convenient dataset catered to checking for poetic device agreement.

Approach

The sound components evaluated here, vowels, consonants, and stress, are elements of the syllable. Agreement in various combinations of these syllable components are characteristic of different poetic devices. In concept, I test all possible pairs of words in the dictionary to check if they satisfy the conditions for various poetic relations. There are at least 229 [39] named forms of rhyme alone, including feminine, dactylic, holorime, etc.... Figure 4.1 shows the set of poetic devices selected for this study. For both words in any given pairwise comparison, the components of their syllables which are shown in green must match exactly to satisfy the constraints of a device. For example, both words in the pair "cat" and "hat" have the same vowel, stress,

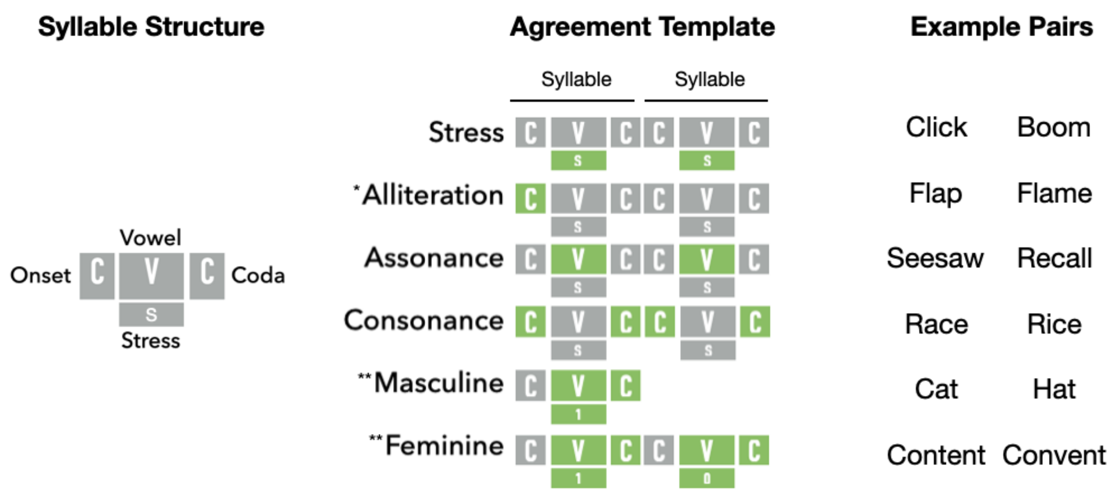


Figure 4.1: Template for components of syllables that must agree (green) in order to conform to the constraints of a given poetic device. C=Consonant, V=Vowel, S=Stress (0=Unstressed, 1=Primary Stress, 2=Secondary Stress). If not specified, words may be any number of syllables where every syllable must match the specified condition. * pertains to the first syllable of a word of arbitrary length. ** indicates that matches must be exactly the syllable length specified in the template.

and ending (coda) consonants, so satisfy the requirements of both masculine rhyme and assonance (Figure 4.1), but they have different initial consonants (onsets) so do not satisfy the conditions of alliteration.

Again, the idea is to check every possible pair of words from the dictionary against every template of poetic device similarity conditions from in Figure 4.1. The goal is to arrive at a list of "match" or "don't match" answers for every possible pair of words across each condition. So for the cat-hat pair, or any other pair for that matter, a vector of items would be produced (e.g. [0, 1, 0, 0, 1, 1, 1]). A 1 means the pair satisfied the requirements of one of the tested poetic device (condition), and 0 means it did not.

However, using binary equivalence checks over all unique word-pairs can be time consuming as it is O(n*n) time complexity. In order to examine all pairs of words, each word (120,413 from CMUdict) is compared to every other and across all poetic forms. That is 120,413 * 120,413 or about 14.5 billions pairs of words to check. In total, 7 checks (one for each poetic device) must be done on each pair, which is just over 100 Billion checks.

In order to simplify the problem and allow for future scaling to larger datasets, I use pre-indexing to simplify the required computation. For each word I pre-extract the sound components, shown in green in Figure 4.1, that could potentially participate in the chosen poetic devices. These skeletal sound representations of words can then easily be compared and computed over to extract sound pattern frequencies. For example, for the word "begin", '01' would be extracted for stress, 'B' for Alliteration, 'IH0 N' for Masculine Rhyme, and so on. The result is a table with 120,413 rows, and 7 columns, one for each poetic device (but just the extracted green sound segments of the words). This gives a representation of how many words represent each of the unique instances of poetic device patterns. One can then simply count the total number of words, for example, with exactly the stress pattern of '01' (stress components), or with first consonant onsets of 'B (alliteration components), or ending with 'IH0 N' (masculine rhyme components). If there are 8868 words with just the sound B in the alliteration position, this represents the number of options available in the 'B' sub-network of alliteration. Of course, this would need to be combined with counts of all unique alliterative patterns, not just B, to account for all of alliterative patterns. I report the Top 5 most common sequences for 4 categories in Figure 4.2 and average frequencies in Figure 4.4.

Rank	Stress	Assonance	Masculine	Alliteration
1	'10' - 41774	'AE AH' - 2206	'IY1' - 375	'K' - 8868
2	'1' - 15990	'IH AH' - 2049	'OW1' - 355	'M' - 8486
3	'010' - 11686	'EH AH' - 2020	'EH1 T' - 332	'R' - 6538
4	'100' - 11637	'AA AH' - 1886	'EH1 L' - 324	'B' - 6506
5	'12' - 7120	'IH' - 1713	'EY1' - 314	'D' - 6314

Figure 4.2: Top 5 most common poetic device patterns and their frequencies across words in the dictionary. A given device, like assonance, can be thought of as a network, while patterns within a device, like /æ-ʌ/ (AE AH) can be thought of as a sub-network of the larger assonance network.

Figure 4.2 shows the Top 5 most frequent pre-extracted patterns, and how many words from the dictionary matched each pattern exactly. For example, in the stress column, it can be seen that the most common stress pattern is "10" (stressed-unstressed), and that 41,774 words from the dictionary match this pattern exactly. This provides the ability to examine how many word options are available for each instance of a given device, as well as the distributions of these instances across devices.

As an example, Figure 4.3 shows the frequency vs rank plot of all vowel sequences (assonance) observed in words in the dictionary. This represents the complete continuation of data from Figure 4.2 in the Assonance column.

One might also conceive of these sound relationships in terms of networks or graphs, construing each poetic device as a network comprised of many fully connected smaller networks which represent incarnations of the poetic device in question. For example, all (6506) words that start with /b/ or all (8868) words that start with /k/ make up two of the fully connected networks in the larger alliteration network.

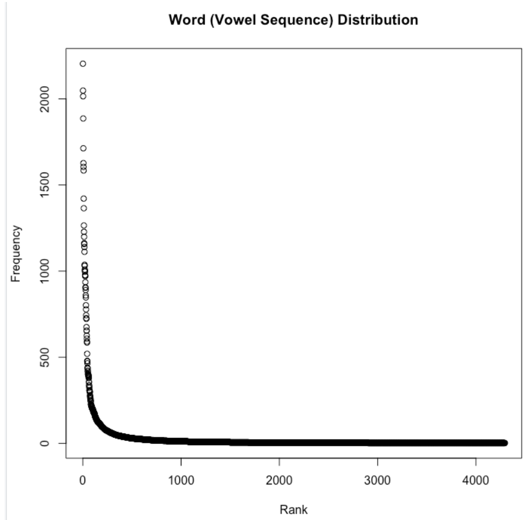


Figure 4.3: Frequency vs Rank of assonance (vowel) sequences extracted from unique words in CMUdict

Results

	All Patterns			Only Repeated Patterns		
	Patterns	Avg. Words/Patt	Entropy	Patterns	Avg. Words/Patt	Avg. Transitions/Patt
Alliteration	147	696.3	3.41	113	905.5	2,116,575.4
Stress	260	458.8	2.44	178	669.7	6,707,507.5
Masculine	1,281	27.7	6.19	954	24.8	1,108.7
Feminine	2,579	23.3	5.9	1,566	37.8	6,994.4
Assonance	7,756	15.4	6.60	4,295	27	7,931.8
Consonance	47,587	2.5	9.9	14,330	6	82.4
All	102,471	0.28	11.47	11,480	2.5	2.2

Figure 4.4: Results about poetic devices and their patterns. Patterns represent the unique phonological sequences possible for each type of device.

Unique Poetic Device Patterns

Here, I count the number of unique sound sequences related to each poetic device. Figure 4.4 column 2 (patterns) shows how many unique instances of each type of poetic device were identified. I am specifically interested in repeated patterns, so I also report in terms of the number of unique patterns for each device with a word match count greater than 1 (column 5). For example, there are 260 stress patterns from all the words in the dictionary, but only 178 of them characterize more than one word, meaning 82 of these stress patterns appear only once. 'Only Repeated Patterns', on the right, is a subset of 'All Patterns', on the left, which excludes patterns that only occur once.

Avg. Words Per Poetic Device Pattern

Taking this a step further, one might start to ask questions about how productive these poetic devices could be. For example, on average, how many word options are available per pattern in a poetic device? I calculate the average number of words that match patterns within each sub-network of a poetic device. This approach can be usefully descriptive is some cases, but simple averages do not necessarily well capture the dramatic differences in distributions and frequencies.

For example, I sum the frequencies of all patterns in Figure 4.2 Stress column (not just the top 5) and divide by the number of unique patterns, in this case 260. This gives us 458.8 words per unique stress pattern. This is useful for getting a sense of scale, but is an oversimplification.

Entropy of Poetic Device Patterns

There are many unique patterns for each type of poetic device (147 alliterative patterns, 260 stress patterns, 7,756 vowel patterns, etc), and different numbers of words match each of these patterns (e.g. 01, 11, 10). In order to better capture information about variation in the frequency distributions while taking into account the changing vocabulary size (number of patterns), I use Shannon entropy. It provides a lens through which we can compare the predictability of the frequency distributions of each poetic device archetype. Describing this complexity as information allows for a measure of poetic device category in terms of 'bits per sequence'. Low entropy is indicative of a more predictable distribution. Figure 4.5 gives us an intuitive information theoretic way to measure predictability or uncertainty across the landscape of poetic devices. H (entropy) is the average amount of information or uncertainty in the possible outcomes of a given variable X. The variable X has an alphabet x. The entropy H of X is given by the summation (p(x)) times the log-base 2 of (1/p(x)) for each x in the alphabet - here giving results in thee form of bits.

$$\mathrm{H}(X) = \sum_{x \in A_x} p(x) \log_2 \frac{1}{p(x)}$$

Figure 4.5: Shannon Entropy

I use the number of unique patterns in Figure 4.4 column 2 as the alphabet size for each poetic device (A), and the counts of their various instantiations, exemplified in Figure 4.2 as frequencies in the Shannon entropy equation (x).

Poetic Devices as Networks

Another approach is to measure availability based on the assumption that the size of the sub-networks within a given poetic device might not be scaled in relation to the number of unique word possibilities (as above), but rather, may be scaled exponentially as a function of the number of transitions (edges) between members in a group (the density of a fully connected sub-network). There would be undirected edges between all words in a sub-network as in Figure 4.6. These sub-networks are functionally equivalent to the notion of rhyme sets or alliteration sets, where all members are matches with all other members on some phonological dimension.

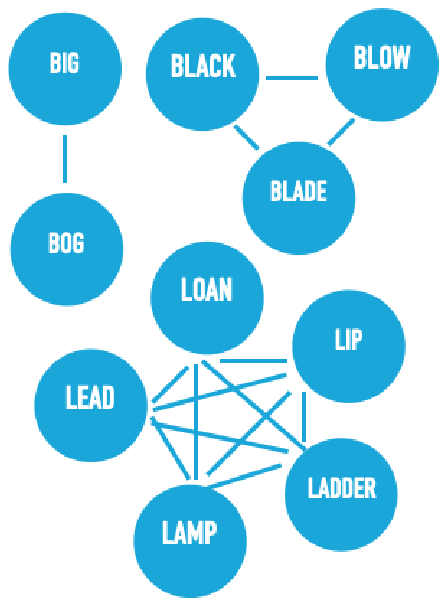


Figure 4.6: 3 Examples of toy alliteration sub-networks - Each sub-network is fully connected within itself.

Imagine a toy case where there are 3 words that start with 'BL', and 5 words that start with 'L' as shown in Figure 4.6. Comparing these cases only in terms of the number of words that match, one could say that the ratio of resources between these two sub-networks is 3:5. However, if considered in terms of how many transitions (matching pairs) are possible within each sub-network rather than the number of nodes, the ratio is different (3:10). Calculating the degree of a fully connected network can be done with the following expression (number of edges = n(n-1)/2)). The number of nodes in each sub-network (2, 3, and 5) provide a different scale than the number of corresponding edges (1, 3, 10).

What this accomplishes is to more heavily weight the larger sub-networks in order to accommodate the intuition that the resources available in these networks may scale with the transition counts between words (degree), not simply with unique nodes (possible word options).

Applied to the data, the most common stress pattern, '10' occurs in 41774 words, and since each of these words can act as a transition (a match) to every other one with the same stress, the total possible transitions in this sub-network is 41774(41774-1)/2, or 872,512,651. If the degrees (edge counts) of all sub-networks of stress patterns are summed, the total edges are 1,193,936,341 (possible pairs or transitions). Then, dividing this by the number of unique repeated stress patterns, 178, gives us 6,707,507.5, the average number of transitions for each sub-network of stress patterns. So in this case, unique words that match the '10' stress pattern comprise about 35% (41,774/120,413) of all stress matches, but 73% (872,512,651/1,193,936,341) of all transitions. Again, this has the effect of more heavily weighting poetic device sub-networks that are more densely connected.

4.1.2 Discussion

In general, the number of unique patterns for poetic devices is inversely related to their availability as a resource in terms of average words per pattern, and their predictability in term of entropy. This inverse relationship also holds true for average transitions (or density) of poetic devices, but scaled dramatically differently due to counting the edges rather than nodes of the fully connected graphs within each poetic device.

Stress & Alliteration

There are a relatively small number of unique stress and alliterations patterns, but for different reasons. In the case of stress patterns, there are only 3 possible types of stress, 0 - Unstressed, 1 - Primary Stress, 2 - Secondary stress.

On the other hand, the first (alliterative) sounds of a word can be comprised of both vowel and consonant segments, constrained by phonotactics.

This results in a similarly limited number of possible unique patterns that can comprise stress or alliterative sequences within words. Due to these smaller number of patterns, the average number of words that match each of their patterns is quite high, 458.8 and 696.3 respectively. This means that for any given instance of a stress or alliterative sequence, there are hundreds of individual words that will match or agree.

It should not be surprising then, that the information theoretic measures demonstrate that stress is the most predictable (least surprising) type of sequence, with an entropy of 2.44 bits per sequence, followed by alliteration, assonance, and consonance, at 9.9 bits per sequence. These entropy measures take into account the frequency distribution of these sequences, and so give a more standardized comparative perspective on the resource availability of poetic devices than simple counts or averages

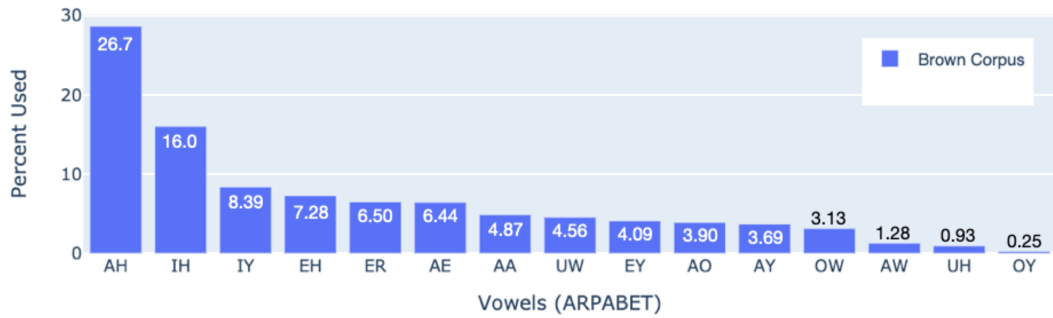


Figure 4.7: Rates of usage for each of 15 vowels in spoken Standard American English - 52,000 sentences from Brown Corpus

communicate.

Vowels

On the other hand, there are many thousands of unique assonance (vowel) and consonance patterns (7,756 and 47,587), due in part to how large their alphabets are (15 vowels & 25 consonants) and the fact that these sequences can be quite long. So when selecting any given assonance or consonance pattern (e.g. /ɛ-ɪ/ or /bl-k/) there will be, on average, 15.4 words that match the vowel pattern, and 2.5 that match the consonant pattern (also differing at different sizes).

Following from this, I inspect the more popular structured poetic devices of masculine and feminine rhyme. In this domain, there are a total of 1,281 and 2,579 possible unique masculine and feminine sound patterns respectively, where each pattern has, on average a set of 27.7 or 23.3 words that match it. Their predictability in terms of entropy is also relatively in the middle (for the data explored here), at 6.19 and 5.9 bits per sequence respectively.

The frequencies of these poetic devices are also often related to the underlying frequencies of the phonemes involved in their construction. I use the vowels specifically to exemplify this notion. Figure 4.7 shows the relative frequencies of individual vowel sounds from the Brown Corpus of sentences. Comparing these rates to the Top 5 sequences from Figure 4.2 reveals a few things. First of all, 4 of the the Top 5 assonance (vowel) sequences from Figure 4.2 contain the number one most frequent vowel sound AH, which here represents both /ʌ/ and /ə/ (26.7%). The other vowels in the top 5 assonance sequences are /æ/, /ɪ/, /ɛ/, and /ɑ/, ranked 6th, 2nd, 4th, and 7th respectively in the Brown Corpus of sentences (Figure 4.7). It seems natural that many relatively high frequency vowels populate the most common vowel sequences. However, in the masculine column of Figure 4.2 it can be seen that the second most common masculine rhyme pattern in the lexicon is one of the least common individual vowel sounds in natural speech (12th of 15), which occurs only 3.13% of the time in the Brown Corpus. The remaining vowels in top masculine rhyme patterns contain the 3rd, 4th, and 9th most common vowel segments. Further exploration of this kind may provide insight into the relationship between sound patterns as they appear in thee lexicon compared with use in natural language.

Homophones

The "All" category can be understood as a representation of the faithful phonetic transcription of each word in the dictionary. This means there are 102,471 unique word pronunciations in the dictionary, and some 11,480 of them are spelling homophones, sound sequences that are associated with more than one unique spelling of a word (e.g. brake/break, sell/cell). For each spelling homophone there are on average 2.5 words with alternate spellings (Figure 4.4). The vast majority of words in the dictionary are not homophones, and their sound sequences fully disambiguate their spelling. Again, it is worth pointing out that because of the huge vocabulary and relatively flat (close to uniform) distribution of unique full word pronunciations, Shannon entropy is extremely high, 11.47 bits per sequence, indicating high unpredictability in this poetic device's in the dictionary. This is not surprising given that there seem to be strong pressures towards discriminability of the lexicon [309]. If the English lexicon were dominated by homophones, as in Chinese for example, other features of the language like phonological, lexical, and contextual information may adjust to facilitate their better recognition and interpretability [310, 311].

Poetics

Conventionalized poetic structures leverage a number of these, and other named and unnamed sound patterns. Casting poetic disciplines and the choices of individual authors in terms of the space of all possible patterns can give context to the forms that crystallize and become popular, as well as their frequency of use.

Forms like sonnets, often require many instances of masculine or feminine rhyme in aligned end-rhyme positions, which also nested in the context of even larger constraints on stress like iambic pentameter. How is it possible to satisfy both constraints at once? Perhaps these differentials in poetic resource availability plays a role.

It is also interesting to note that most end-rhymes in sonnets seem to use masculine rhyme, despite the fact that feminine rhyme is shown here to be more resource rich. This may indicate an overriding preference for shorter words in these contexts, or an increased difficulty in the search for longer sequences with more constraints.

On the surface, phonaesthetics is about sounds in language that are euphonious (pleasing) or cacophonous (displeasing). This might include investigations of structures that humans report as attractive as well as structures that human brains respond to more (higher activation).

In a broader sense, aesthetic worth has been suggested to positively correlate with "unity, complexity, and intense human quality" [16]. Describing the relative complexity of the constraints involved in each of the poetic devices may also contribute to better understanding their aesthetic utility. For example, in the case of /ʊ/ (OW) being a relatively uncommon vowel sound, but one of the most resource rich masculine rhyme patterns, it is worth exploring if features related to its sound aesthetics may be implicated.

Different sounds in languages occur at different rates, that is, they have particular probabilities of occurring. But how much do these unigram rates of occurrence relate to their prevalence in larger more complex poetic structures? This work can ultimately be used in conjunction with studies on observed usage rates of poetic devices in order to bridge the gap between real world usage and potential resource availability (measured various ways). Is it the case that usage rates broadly map onto relative resource availability rates for more complex forms? Which other factors (perceptual enjoyment/intelligibility/cultural influence) impose additional constraints on the utility of poetic devices networks in practice?

Cognition

Although I have outlined the space of possible poetic sound correspondences, I have said little about their cognitive implications. It seems unlikely, for instance, that humans use or produce poetic devices with distributions equivalent to those discussed here. First of all, certain individual sounds, and combinations of sounds are more pleasing to the ear than others. Second, it may be the case that particular poetic devices are perceived as more euphonic (pleasing) than others. If the more complex phonological structure of repeated rhyme resonates in a way similar to resolution in music, it could help explain the more common explicit use of rhyme than assonance patterning.

It has also been suggested that both perception and production of rhyme, and other sound patterns, involve "cognitive processes including, orthographic coding, graphic-phonetic transformation, phonological representation and phonological segmentation" [213]. That said, certain patterns provide more available options for word matches, and so may be deferentially easier or harder utilize. Said another way, particular poetic devices, or sub-networks within those devices, may act as attractors that make us more likely to or more familiar with navigating them.

Future Work

The scope of phonological pattern templates included could be expanded in order better represent the availability of more possible phonological patterns. It is also important to understand the relationship between rates of usage of these patterns in the wild and their availability in the lexicon. To what degree does the availability of various poetic devices align with their use in the wild? This question can be investigated at the level of poetic devices broadly, or at the level of distributions of sub-networks within poetic devices. Furthermore, related behavioral experiments could be conducted to investigate the structure of human phonological networks. This may illuminate the level of correspondence between the poetic resources strictly available in language in principal, and those adopted by their users in practice.

Conclusion

In sum, I have operationalized the underlying patterns of common poetic devices, and used them to identify and count the occurrences of these pattern in English words. This represents an exploration into phonological vocabulary at the level of legal perfect device archetypes - further analysis may also be done on the specific instances and distributions of patterns within of any given device (masculine, feminine, etc..). I have shown that the varying constraints which characterize poetic devices manifest in the frequency of poetic patterns and their availability and predictability. This study is a first step towards a more comprehensive description of the landscape of phonologically driven poetic tools and phonological vocabulary.

4.2 500 Years of Imperfection

How perfect is perfect rhyme?

In the last study, I assumed that poetic devices are composed of perfectly matching elements, and they often are. But even in traditions where perfect rhyme is standard, imperfect agreement is also common. In this study, I quantify different forms of segmental matching within overt rhyming word pairs from five centuries of poetry (1450-1950). It constitutes an effort in documenting the components of phonological vocabulary at the level of perfect rhyme (in poetry). Specifically, I investigate the relative rates of matching in the vowel, stress, and consonant (onset and coda) components individually, as well as in combination. I demonstrate that rhymes most commonly exhibit perfect matching (agreement) in the coda, followed by stress, and then vowels. I also explore differences in these patterns as a function of the number of lines apart the rhymes are, and the historical time period the pair occurs in. Finally, I discuss possible reasons for the differential in the rate of matching components of rhyme including ambiguous pronunciations (transcription errors), and the great vowel shift.

Rhyme is a prominent feature of verbal art practices around the world. It is not only a named component of syllable structure, but also a type of similarity that conventionally requires the 'rhyme' (Nucleus + following consonants in the syllable) of corresponding syllables across two words to pass some similarity threshold. Remember that cases where a nucleus and coda of separate words share identity completely, such as CAN-PLAN, are sometimes called perfect (masculine) rhymes. Nucleus-coda pairs that fall short of perfect identity, such as CAN-CAM or CAN-CANS, are often called imperfect rhymes. In previous studies, the perceptual similarity of imperfect rhymes have been explored, often using similarity judgements of rhyming pairs. Here, I use the rhyme schemes of the Sonderegger dataset [9] as a way to explore imperfections in their pairs. Although many of these rhymes are perfect, some are not. These rhymes are marked with conventional ABBA notation, which obscures the degree of matching across their sound segments. I examine the variation within these rhymes, describing rates of feature matching across perfect and imperfect pairs.

Analyzing rhyme across time periods is difficult because the pronunciation of words in a given language changes over time, and we lack the relevant phonetic transcriptions [312]. This is made worse by the fact that these phonetic transcriptions are static, and do not reflect the variation of spoken language (as discussed in Chapter 2). The approach is to break rhymes down into binary comparisons along various component parts, which will account for both conventional perfect rhyme, and cases of imperfect rhyme containing at least some perfect correspondence in component parts.

Validating segment equality is a type of similarity measurement that can be easily applied to entire syllables, or individual components of syllables. In this way, the rhyme and associated features can be treated as binary features. In other words, when comparing CAN and PLANS, a series of identity matching tests are run such as, do the vowels of this pair match? do the codas match? do the stresses match? do both vowel and coda match? etc... I compare only the last syllable of each line (in rhyming and non-rhyming pairs).

This approach will pull apart the component features of rhyme to understand their relative weighting, and ultimately, the role they play in annotated instances of rhyme and non-rhyme pairs.

4.2.1 Data

Data Collection

The data used here is from a collection hand-coded by Morgan Sonderegger [9] which has been the basis of previous rhyming studies. The Sonderegger dataset consists of hand annotated poems from 32 authors across five one-hundred year periods from 1450 to 1950 from which I extract 22,219 rhyming pairs and 152,857 non-rhyming pairs. These pairs are identified by the ABAB, AABB, etc. coding that accompanies each work.

Within the rhyme scheme of each work, there many be more than 2 words associated with a given letter (e.g. A,B,C). For the current purposes, I use rhyme sets, extracted from the annotated rhymes schemes of Sonderegger [9] to generate all possible rhyming pairs from within its works. I employ ordered rhyme sets, instead of unordered ones, in order to track the relative location of rhyming utterances. This allows us to ask questions about the degree to which segment similarity in rhyme pairs changes when rhymes are farther away from each other (line distance).

Phoneticization

Phonetic transcription is done using the CMU Phonetic Dictionary [313]. It should be noted that while the Sonderegger dataset represents English as pronounced over five centuries, the CMU database is intended to capture the pronunciation of modern Standard American English only. This is a significant limitation of the current approach.

4.2.2 Results

Rhyme and Non-Rhyme Pairs

In order to provide important context, Figure 4.8 displays rates of feature matching not only for rhyming pairs (a-a b-b c-c etc) but also for non-rhyming pairs (a-b, b-c etc). This provides a chance to directly compare differences in feature matching rates in the subject of interest (rhyme pairs) and how they are systematically differ from non-rhyming pairs.

Although agreement on these stress and word length features appear common in pairs of non-rhyming line final words, agreement is greater in rhyming pairs. In both cases, rates of matching are higher for rhyming than non-rhyming pairs, 91.2% vs 79.7% for stress and 70.7% vs 61.5% for word length.

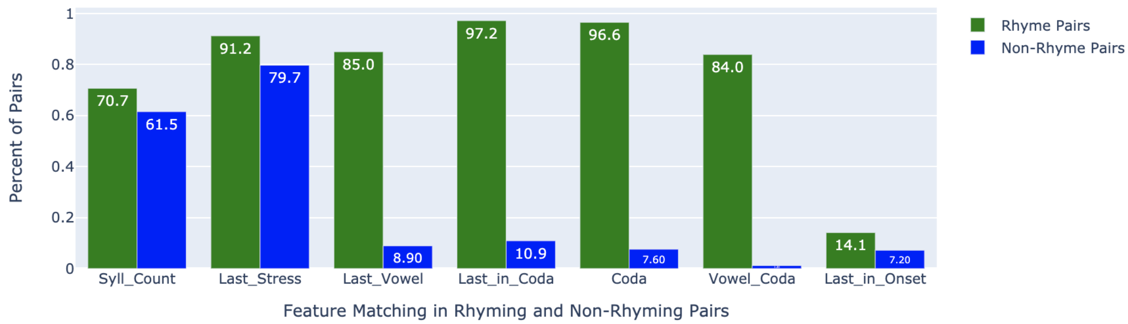


Figure 4.8: Featural Agreement in Rhyming vs. Non-Rhyming Pairs

Three decompositions of the terminal syllable, vowel, coda, and final consonant, illustrate the expected and dramatic differences across rhyming and non-rhyming pairs. Rhyming pairs considered on these features match in 84-97.2% of cases, whereas the same features in non-rhyming pairs are equivalent less than 11 percent of the time. Interestingly, the rate at which the final consonant in the coda (last_in_coda) match in rhyming pairs is greatest, at 97.2%, whereas only 85% of vowels (Last_Vowel) match in rhyming pairs. Vowel and coda positions are not only highly associated with rhyme, but also, highly correlated with each other (R=.7). The entire coda (96.6%) is also more indicative of rhyme than the nucleus (88.0%) across rhyming word pairs.

By treating the combination of vowel and coda matching as a single equivalence check (Vowel_Coda) one can notice that the traditional vowel-coda (rime) condition of end-rhyme is true for these annotated pairs in only 84% of cases.

Finally, I check the rate of agreement in the last consonant phoneme in the final onset of each word of a given pair. It may be expected that there will be some similarity avoidance in this position in order to facilitate differentiation between rhyming words. For instance, CAN and PLAN rhyme, but if they agree in onset, such as CAN and CAN, they are no longer phonologically differentiated, and also no longer rhyme. Despite this, agreement in the onset is found to be nearly twice as great in rhyming pairs as in non-rhyming pairs.

Time Periods: 500 Years

Next, I plot the rates of agreement on these same features (Figure 4.9), but grouped by 100 year time blocks. This allows for identifying trends in these rates of agreement over time. In general, the percentage of rhyming pairs that match on the given features increases over time. The relatively flat line at the top of Figure 4.9) shows a slight increase in coda agreement over this time period, from 94.4% in 1450-1550 to 97.7% in 1850-1950. For other features, the change in agreement over time is more variable, but generally increases across all time periods for stress (80% to 88.5%), vowel (75.4% to 87.2%), vowel+coda combined (72.3% to 86.4%), and syllable count

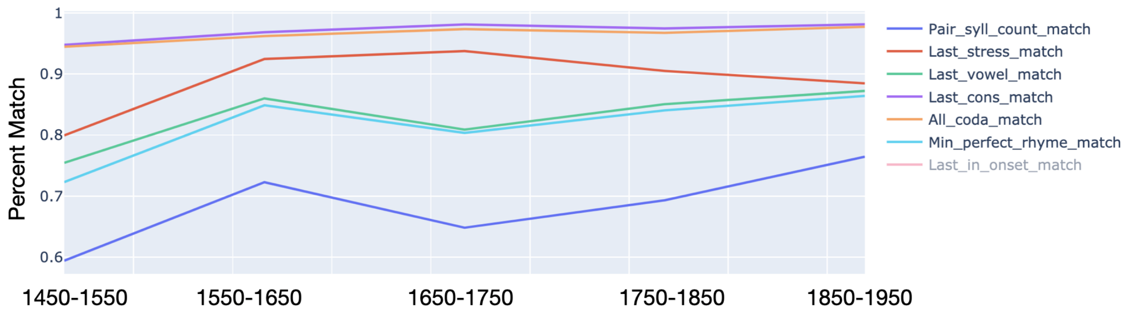


Figure 4.9: Rates of Featural Matching by 100-Year Time Periods

(59.4% to 76.4%). For instance, this exposes that trend that rhyming words from 1450-1500 contain the same number of syllables (often 1) only 59.4% of the time, while from 1850-1950 rhyming words share the same number of syllables 76.4% of the time, a seemingly significant change in rhyming conventions over time.

Line Distance

Finally, in order to understand whether the physical proximity of words within a verse pair plays a role in agreement, I group the phonological features by the line distance between words in each rhyming pair. For example, if a poem is annotated to have an ABBA pattern, corresponding to lines 1-4, this A-A pair have a line distance of 3 (abs(4-1)), and the B-B pair have a line distance of 1 (abs(3-2)). Figure 4.10) shows that most (17,374 rhyme pairs of 22,219) occur in close line distance proximity (1-2). Line distances up to 5 are reported, beyond that the number of rhyme pairs is too low to provide reliable results. Although there are some slight variations that may merit further exploration at a more granular level, the observable trend is that agreement rates across features do not increase or decrease as a function of line distance.

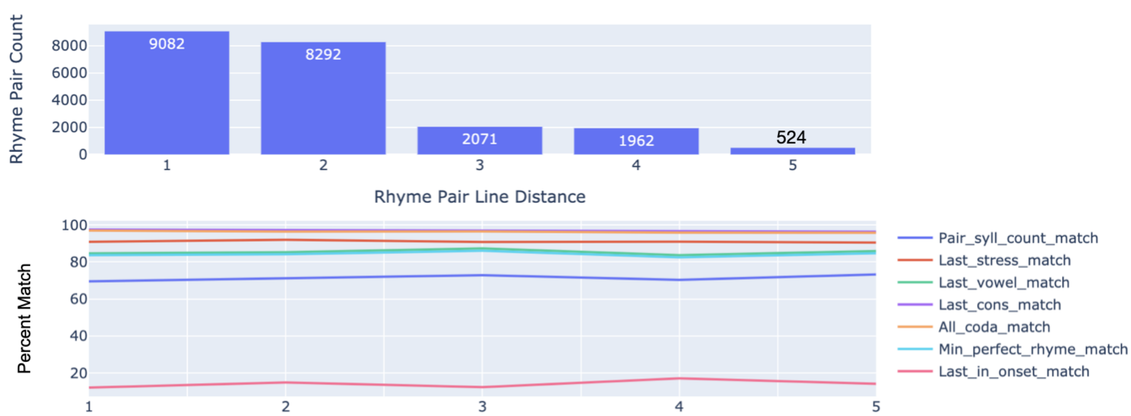


Figure 4.10: Rhyme Pair Line Counts Histogram and Featural Matches by Line Distance

4.2.3 Discussion

These results show that, over 500 years of English poetry, the coda consonants are more indicative of rhyming pairs than vowels, or even vowels and codas combined. This could be due to the changing accuracy of phonetic transcriptions, poetic structure, accents, and/or relative vowel-consonant shift.

Although I suspected that increased physical distance between paired words would reduce rates of agreement in various features, this was not the case. In fact, the line distance between rhymes seemed to have no impact on their level of agreement. This could be a fact about how well (perfect) similarity constraints are preserved across distances, or an artifact relating to a similarity bias human annotators may face when classifying rhyming words.

500 Years of Sound Change

This dataset contains poetry from regions with dramatically different English accents, not reflected in the phonetic transcriptions. Moreover, between the 14th and 18th centuries, many English vowels experienced a dramatic shift, causing long vowels to become raised, and others to merge and dipthongize. While pronunciations have shifted, spellings have largely remained the same [314].

As an example, from 1350-1700 the word ”meat” was pronounced similar to ”met”, then ”mate”, before it took on the modern pronunciation. Meanwhile, the word ”meet” did not follow the same trajectory (it has been pronounced similarly across these time periods).

Because spelling has remained relatively consistent over this period, encoding from phonetic dictionaries or human annotators used in studies such as this have strong biases towards modern pronunciations over older ones.

While English vowels changed drastically during this period, consonants did not. So the relativley higher reliability of coda positions in rhyme pairs over these 500 years is consistent with the historical facts concerning vowel and consonant shift. However, this is specific to the history of English, and trends in other languages will depend heavily on the particular historical trajectory of change each language. For example, data from 950-1450 (which I do not investigate) should contain evidence of consonant shifts (such as pronouncing the ’k’ in ’kn’ words), which would reduce the accuracy if using modern pronunciation look-up tables for that period.

Beyond sound shift related errors, it may also be the case that conventionally, in English, coda consonant matching is simply a more robust marker of perfect rhyme than nucleus matching. However, as I will show in section 3 of the chapter, in multi-syllabic imperfect rhyming sets, vowels and stress patterns are the most predictable positions, while consonant positions are least predictable. This indicates that different kinds of rhyme may exhibit very different matching standards or constraints.

Conclusions

In sum, this study shows that very simple measures of segment agreement are able to capture a surprising amount of the predictability of rhyme. Agreement in various parts of the rhyming syllables do not depend on how close the pair are with any given work, but does increase with more modern time periods. The changing agreement rates of segments in rhyming pairs over time reveals the long running stability of the coda in rhyme. In addition, it is consistent with effects of the great vowel shift. This provides a snapshot of the components of phonological vocabulary at the level of traditional perfect rhyme over this time period.

In some contexts, agreement in coda consonants alone is a more reliable predictor of rhyme than the common operationalized definition of rhyme itself (agreement in both the nucleus and coda). This finding about coda agreement in rhyme may generalize to other English rhyming verse written between 1450-1950, especially in contexts dominated by single-syllable perfect end-rhyme, such as in the Sonderegger data set [9].

Finally, the specific conclusions about coda preference in English rhyme classification may not generalize to other languages or traditions.

Future Work

Follow-ups to this study should expand the scope of the time periods, traditions of rhyme, and languages, to better understand the dynamics of the nucleus and coda in rhyme pairs. Furthermore, analysis at the level of distinctive features, natural classes, or even graded measures of similarity can give more fine grained understanding of rhyme constraints in the wild.

Generalizations about the cross-linguistic relative stability in vowels versus consonants during periods of sound change may also inform the expectations of similar diachronic studies in other languages. Expanding the feature sets in these models to include phonological characteristics and natural classes would also allow for a more fine grained study of imperfect patterns within rhyme. Finally, the imperfect patterns here are studied in the context of mostly perfect rhymes, which sets up a particular standards and expectations - different than the rhyme conventions in other contexts.

4.3 Rhyme Sets: Multi-Syllable Rhyme

How can multi-syllable imperfect rhymes be quantified?

A common literary device, multi-syllable rhyme, is an instance of a broader class of complex repeated phonological structures. These perceptually noticeable and elusive forms, and others like them, are not understood in terms of frequency, complexity, constraints, or learnability. Humans produce and perceive a variety of dynamic phonological patterns in the language arts, particularly in lyrics and rhyming games, yet no robust framework for their investigation exists.

Named and perfect rhyming forms have dominated traditional poetic disciplines for many years. Such devices are often identified by segment or feature matching, a definition-based (Aristotelian) approach to category membership discussed introduced in Chapter 1. However, in recent decades, subcultures like hip-hop have adopted and transformed rhyme into increasingly complex forms [11]. These forms can be imperfect, multi-syllabic, and defy classification by conventional poetic terms. In this section, I begin exploring the fuzzy internal structure of category membership (i.e. family resemblance) [91, 89] by examining variation in segment matching of rhyme sets from a rhyming game.

In this study, I analyze rhyme sets, groups of utterances which can be described by some shared phonological structure. As an example dataset, I present data from a hip-hop rhyming game called 'bar pong', where participants take turns rhyming with a multi-syllable seed phrase. Finally, I demonstrate approaches to analyzing these ordered rhyme sets (dynamical systems) using descriptive and computational methods to reveal and quantify underlying sound structure.

4.3.1 Introduction

Background

Standard poetic analysis considers explicitly defined poetic devices which represent a strict relationship between two words or phrases, (e.g. perfect, masculine, feminine, syllabic rhyme, or patterns such as assonance and consonance). These relations are perceptually and linguistically interesting, but also limit the range of patterns that can be investigated. This is primarily because they pre-specify the exact requirements of a similarity relation (e.g. masculine - last coda and nucleus must agree exactly and both nuclei must present primary stress). Perfect adherence to these strictly defined relations impairs the ability to uncover the true richness of sound patterning in rhyme. This study takes an approach that examines the internal structure of sound patterns [91] within and across sets of rhymes.

Rhyme Phrase Sets

Set 1	Set 2	Set 3
seventy texts	not to mention	Vivian Banks
sesame bread	stocking weapon	video tapes
clever defense	hockey lesson	fill in the blank
Gregory Peck	gothic engine	obsidian tank
never respect	blocks in tetris	miniture gate
memory test	got detention	infinity days
vegetables fresh	not suspension	city is great
yellowy flesh	moxie bredrin'	little bit late
deadly sequence	saucy tendons	live in a maze
anyone's guess	possy henchmen	hit wit' a train

Figure 4.11: Truncated Rhyme Sets (10 members) drawn from three Bar Pong matches

4.3.2 Data

As discussed in Chapter 1, a rhyme set is a collection of words or phrases that are identified to rhyme with each other (ordered or unordered).

Figure 4.11 shows 3 truncated rhyme sets drawn from the data in this study (discussed below). The critical assumption made is that rhyme sets have some underlying phonological structure that can be described. So how can these underlying sound structures be captured? Traditional terminology for poetic analysis cannot well describe these data. Here, I describe these underlying patterns with visualizations, entropy (predictability), and recurrence quantification analysis (dynamics).

Collecting Data

Annotated rhyme schemes from traditional poetry, in the ABAB/AABB style notation, provide a starting point for analysis, but lack the dimensionality of more modern patterns known to exist in the wild. Recently, multi-syllabic patterning has become common in a variety of lyrical genres, none more notable than Hip-Hop. However, large data sets of ordered rhyme sets have not yet been compiled. Battle Rap, in particular, has become a venue for the development of multi-syllable rhyme schemes, with 830+ Battle Rap Leagues having produced over 60,000 rap battles (10-45 min each) [82]. Their transcriptions and annotation can be aided by both human marking as well as recent unsupervised machine learning methods for rhyme identification. These unsupervised techniques have had some success, but the resulting sets of classified matches are also not systematically collected or analyzed.

In particular, a rhyming game that spawned from the battle rap community, Bar Pong, provides a relatively controlled resource for the current purposes. Each round, players take turns saying individual lines that rhyme with the designated 3-5 syllable seed phrase. As mentioned above, sets 1-3 in Figure 4.11 show example rhyme phrases from three different Bar Pong matches.

Bar Pong

Bar Pong is an improvised rhyming game done in a capella format (first developed by a battle rapper named Fredo). The name comes from ‘beer pong’, a turn-taking drinking game. The term ‘bar’ here is a popular term for a line of lyrics in hip-hop culture. To begin, a 3+ syllable seed phrase is chosen, such as “Vivian Banks”, and participants (usually rappers) take turns improvisationally coming up with the next lyric; a punchline, insult, or turn-of-phrase that rhymes with the seed phrase. This game is particularly interesting because rhymes come to be identified by the crowd (of other rappers), who will disqualify a player if they believe the utterance did not rhyme enough with the seed phrase, an informal and revealing kind of demarcation.

This process often continues for dozens of rounds, and only stops when a participant takes too long to come up with something ($\tilde{1}0$ seconds) or says a phrase that the surrounding audience (often rappers) determine is not similar enough (i.e. not a rhyme). There is also a restriction that any given lexical item (word) cannot appear more than two times across all phonologically matched parts of the phrase. This prevents repeated use of the same words to achieve the required sound similarity. Again, it is important to point out that neither the audience nor the participants themselves, have terminology or explicitly agreed upon standards for determining the acceptability of these rhymes. Instead they use a holistic and perceptual sense of acceptability. In these ways, Bar Pong functions as a sort of naturalistic experimental paradigm where a dyad (two people) alternate in producing utterances that are phonologically similar to a target phrase.

4.3.3 Methods

Visualization

Figure 4.12 shows a snapshot of all 8 rhyme sets from this tournament. In terms of their underlying vowels, these patterns fall into two clusters, sets with more vowel agreement (Group A), and sets with less vowel agreement (Group B). For Group B, the underlying stress patterns are also shown. Notice that syllable positions which have a large amount of variation in stress also tend to display a higher degree of vowel variation (e.g. the second column of 'Golden Rule', the second column of 'Vivian Banks'). In some cases, positions that are stressed (stress columns that are mostly blue) seem to be associated with less vowel variation while positions that are unstressed (mostly white) are associated with more vowel variation.

It is notable that the average number of rhymes produced in Group B rhyme sets is

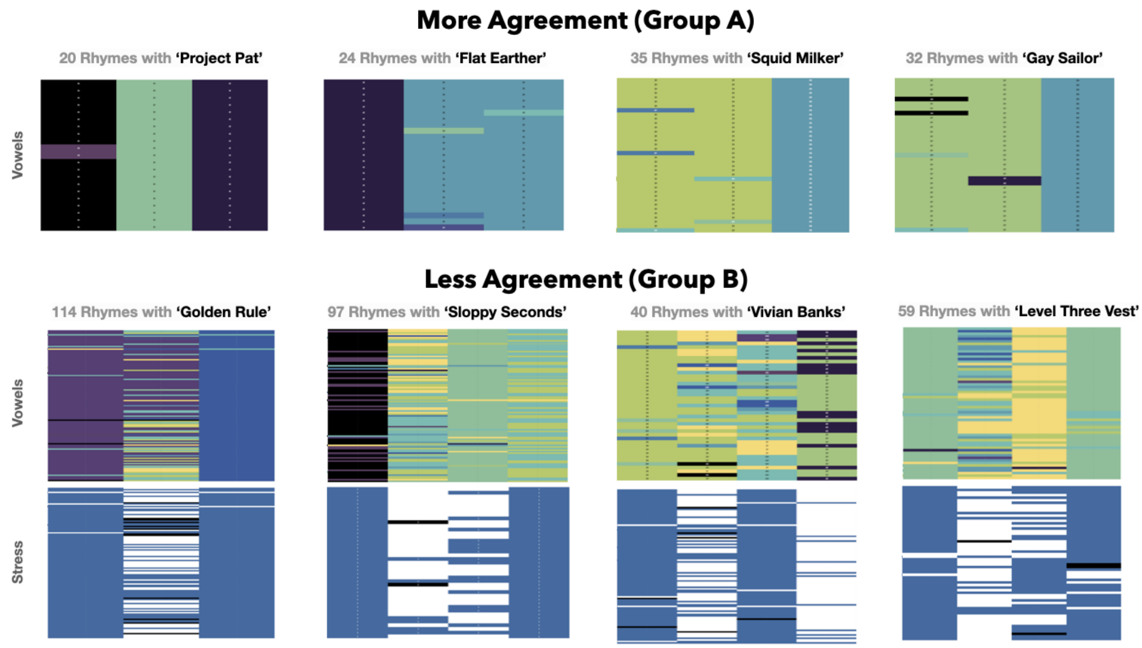


Figure 4.12: Vowel plot of all 8 Bar Pong Matches in a Bar Pong Tournament.

77.5 while the average number of rhymes produces in Group A rhyme sets is 27.7. It may be that the more relaxed matching constraints in the less agreement rhyme sets (Group B) make it easier to satisfy the demands of a given pattern, resulting in more rhymes produced on average. This hypothesis can be tested using both corpus and behavioral studies.

Most Common Pattern

Whether at the level of phonemes, natural classes, or distinctive features, an extremely simple summary of a set can be represented by the most common (frequent) phoneme in each position.

By ignoring some of the interesting variability and dynamics of data like this, we can focus on the simple most common pattern. After counting the frequency of phonemes in each position, the most frequent phoneme in each position can be aggregated to approximate the most common pattern. This is a gross oversimplification, but it allows for comparing these static patterns to their frequency in the lexicon (among other things). Figure 4.13 shows how common the vowel and stress patterns of these seed phrases are in comparison to patterns in the lexicon.

In an English dictionary of 120,000 words, a total of 8119 unique vowel patterns and 265 stress patterns exist. Using the seed phrase from Fredo vs Caustic, 'Vivian Banks', it can be observed that its underlying vowel pattern is the 1233rd most common vowel pattern (of 8119), and its underlying stress pattern is the 78th most common (of 265).

What governs which seed phrases are chosen or deemed worthy of rhyming with? It seems natural to assume that the underlying vowel and stress patterns associated with chosen seed phrases may also be relatively more common in the lexicon, and indeed this intuition is supported in 4.13.

Match	Seed Phrase	Vowel Rank (of 8119)	Stress Rank (of 265)
Caustic vs. FLO	Gay Sailor	861	28
Fredo vs. Irish Rasta	Project Pat	N/A	49
Esem vs. Frak	Squid Milker	177	28
Fredo vs. Caustic	Vivian Banks	1233	78
Caustic vs. Esem	Level Three Vest	4769	139
Dirtbag Dan vs. Reverse Live	Flat Earther	337	28
Fredo vs. Reverse Live	Sloppy Seconds	N/A	26
2 vs. 2	Golden Rule	588	36

Figure 4.13: Frequency Rank of Most Common Vowel and Stress Patterns from Bar Pong Matches

Remember that the selection of seed phrases is an informal process where seed phrases are suggested and rejected until the players agree that *this* is a 'good' or 'reasonable' phrase. For example, "boys with toys" might be passed on, whereas "cats in hats" (a much more common sound pattern) might be accepted. Furthermore, these higher frequency rank patterns may be associated with more words, and therefore lead to longer rhyme sets (more members), at least in the context of bar pong. Figure 4.14 shows the vowel and stress rank plotted against rhyme set length. In these data there does not seem to be a relationship between pattern frequency and number of rhymes generated (rhyme set length).

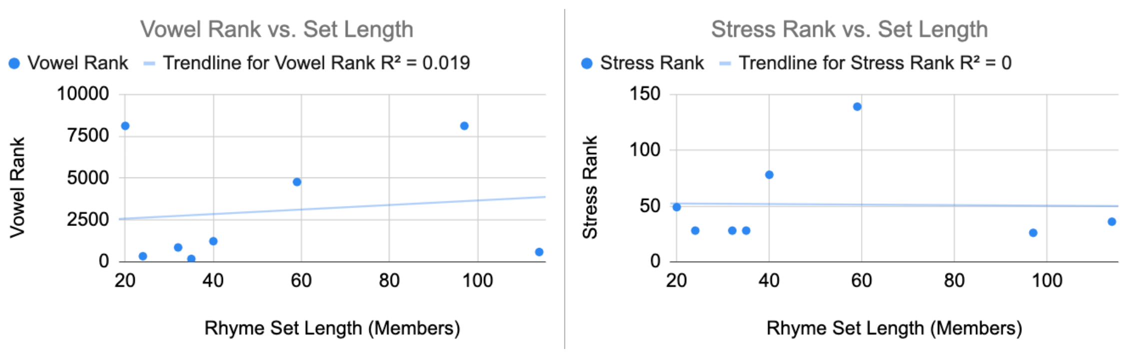


Figure 4.14: Vowel and Stress Rank plotted against Rhyme Set Length i.e. the number of words (or phrases) generated in each Bar Pong match

Positional Entropy

In order to capture the underlying sound structure within a rhyme set, phonemes in aligned positions can be evaluated (e.g. syllables, vowels, onsets, codas).

Shannon Entropy can be used to calculate the uncertainty/predictability within each position across the pattern. The vowel grid representations shown in this dissertation can be considered in terms of their column predictability, and the same approach can be used to consider stress and consonants. All that is required is to count the number of times each phoneme appears in each position across a rhyme set. These positional frequency counts can then be used to calculate the positional entropy.

In Figure 4.15 the vowel and stress plots for the 'Vivian Banks' rhyme set (40 members) are shown on the left. On the right, the entropy of each column in the vowel plot is calculated and recorded in the 'Vowels' row. The same thing is done for stress positions, and consonant cluster positions. The resulting heat-map summarizes the positional predictability of phonemes across all the members of this rhyme set. The higher the entropy (the darker the red), the less predictable the position is. For instance, Figure 4.15A (Left) shows that the syllables in the left and right most stress positions are all the same (black - stressed). This is reflected in 4.15 (Right) where the V1 and V4 stress positions have an entropy of 0, in other words, those positions (columns) are perfectly predictable. On the other hand, the remaining vowel and stress positions in this sample contain multiple phonemes, thus lower the predictability of the position, the and increasing its entropy.

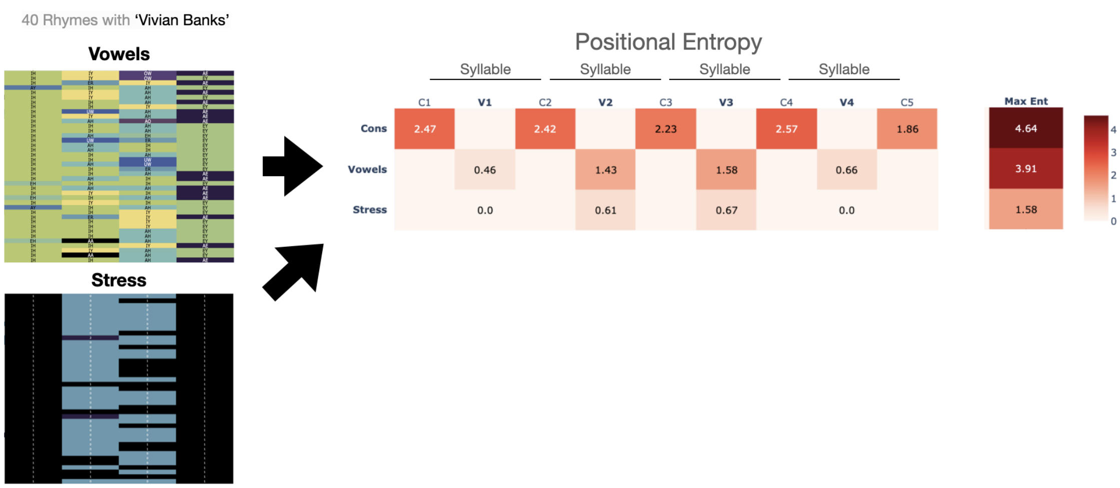


Figure 4.15: Calculating positional entropy of vowel, stress, and consonant components of rhyme sets. Black is stressed, Light Blue is unstressed.

Note that the alphabet size (possible items) for stress, vowel, and consonant categories are are different, so the maximum entropy possible for each respective category (row) in 4.15 is different. Max Entropy for each row (cons, vowels, stress) is shown in a legend on the far right of Figure 4.15.

This approach does not depend on the order of the data (only the frequency), and can be used to characterized both ordered and unordered rhyme sets.

All the rhyme sets in the Bar Pong tournament can be characterized in this manner

Figure 4.16: Entropy Heat-maps of All Matches in a Bar Pong Tournament

to examine and compare their structure. Figure 4.16 shows the related positional entropy heat-maps for these 8 rhyme sets.

A few things are immediately noticeable. First, the final consonant position (blue) of every set is reliably the most predictable consonant position. This is unsurprising since this position coincides with the coda matching of common end-rhyme patterns. This is also the final sound that is heard in the sequence, so there may be a bias to match this element more to facilitate perceptual agreement. Second, there is an observable correlation between the predictability in vowel and stress elements of the same position (green). Finally, among Group B, the 4 sets with less agreement, the least predictable vowel and stress positions (pink) are reliably internal. It seems possible that positions with greater variation tend to be associated with internal vowel positions, rather than at the start or end of a pattern, in order to preserve more predictable boundary positions.

Of course, this is a relatively small sample size (only 8 rhyme sets), and it remains unclear how robust these trends are. Collecting many additional rhyme sets, across genre and size, could lead to revealing more reliable properties about this phenomenon of sound similarity.

4.3.4 RQA

Not only can rhyme sets from this game be examined as unordered members of a set to uncover their underlying positional similarities, but also, they can represent an ordered and unfolding dynamical system where the patterns of imperfect similarity can oscillate and change. In other words, the participants can influence each other as they explore the state space of sound similarity and the lexicon. It should also be noted that any ordered rhyme set coming from even a single individual can also be construed as a dynamical system, influencing itself through choices, state, and path dependencies.

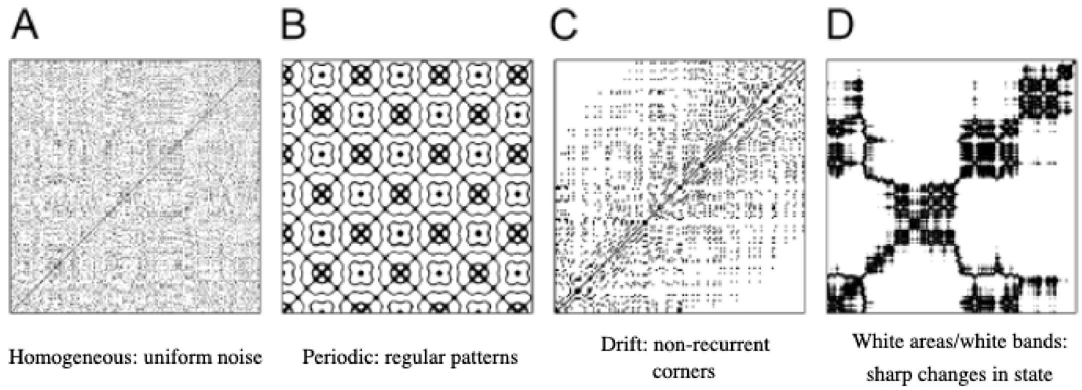


Figure 4.17: High-Level visual features of RQA plots [10]

Recurrence Quantification Analysis was developed in physics for the study of complex dynamical systems [315], and has been used across the behavioral and cognitive sciences to quantify coordination of limb movements and vocalizations in infants [316], coupling and coordination of speakers and listeners [317, 318], bi-manual rhythmic coordination [319], and adaptive decision making strategies [320].

Recurrence plots can capture various high-level features of complex systems (Figure 4.17). Distinct classes of these systems can be identified from their RQA plots, include homogeneous, periodic, drift, and disrupted. Note that the rhyme set (vowels) I consider here will broadly fall into the class of periodic patterns.

In the case of the current bar pong tournament, one can represent a complex dynamical system where sequences (here members of a rhyme set) are considered as samples of the system over time. This approach to characterizing ordered rhyme sets can reveal clusters and trends in these data. Instead of collapsing time, as with positional entropy (which loses the time-series order of information), recurrence analysis can be used to capture, quantify, and characterize patterning strategies through their dynamics. Indeed, ordered rhyme sets may contain a great deal of information about how path dependencies and similarity judgments play out in these systems.

RQA treats data in a serialized form (a single row/vector of vowels). In particular, RQA uses serialized data to compute a low-dimensional phase-space embedding of a more high-dimensional dynamical system. In the case of rhyme sets, the number of syllables (or positions) being considered represents the most salient embedding dimension. In other words, embedding is already known and can be assumed during computation (rather than using some approximation to discover it). The embedding dimension is a parameter of RQA computation, and for the analysis of vowels here, the embedding dimension is set to 3 for 3-syllable rhyme sets and 4 for 4-syllable rhyme sets. With an embedding dimension set to 3, each point drawn on the RQA plot represents sequences that are recurrent across a window of 3 measurements (vowels). It should be noted that RQA assumes an embedding dimension will apply to the entirety of the data (that it is stationary). This assumption is satisfied by the current vowel data which have stationary dimension of either 3 or 4 across each rhyme set (corresponding to the 3 or 4 syllable vowel patterns).

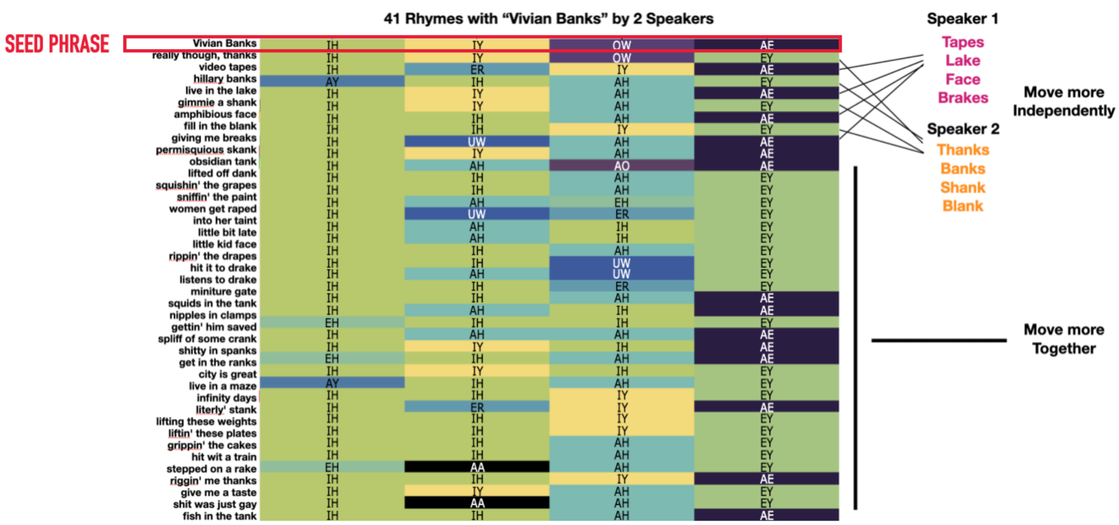


Figure 4.18: Visualization of vowels from a round of Bar Pong. This is a turn-taking game where the first row represents the seed phrase and successive rows are rhymes uttered (turn-taking) by speaker 1 or 2.

In order to build a stronger intuition for why rhyme sets present interesting dynamical systems, in Figure 4.18 I display the underlying vowels of each phrase in the rhyme set seeded by 'Vivian Banks' (also seen in Figures 4.12 and 4.11). Because bar pong rhyme sets unfold over time as two speakers take turns rhyming, vowel grid plots can reveal intuitions about these speakers' coordination over time. For instance, for the first 8 rhymes (after the seed) Speaker 1 only uses /æ/ (AE) in the last vowel (right-most position) while Speaker 2 only uses /eɪ/ (EY) in the same vowel position. This represents a phase where, in the terminal vowel, the speakers seem to be moving more independently. However, after this phase, the speakers seem to move more similarly (with respect to the right-most vowel position).

These data can also be serialized and displayed as RQA plots. Figure 4.19 shows the vowel grid plots and the corresponding RQA plots for 4 of the considered 8 rhyme sets (Group A). Notice how the high degree of vowel similarity and repetition in these 3-syllable patterns are captured in the RQA plots. Diagonal lines in recurrences plots represent a repeated sequence of event states. In these plots, each event state (point) represents a recurrence of a sequence of 3 vowels. Notably, the percentage of

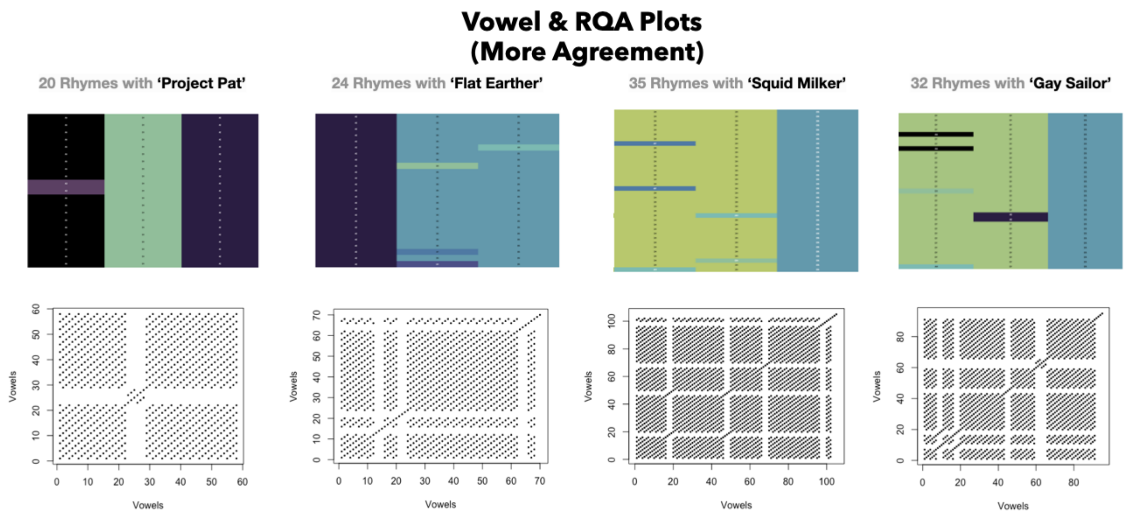


Figure 4.19: RQA Plots of Bar Pong Rhyme Sets with more similarity

recurrence points that occur within diagonal lines (determinism) is extremely high across Group A, above 99% (Figure 4.21). Again, this is expected given the observable 3-dimensional vowel patterns that dominate Group A. At the vowel level, these 4 rhyme sets seem to follow a deterministic process (i.e. few single dots, many long diagonal lines)

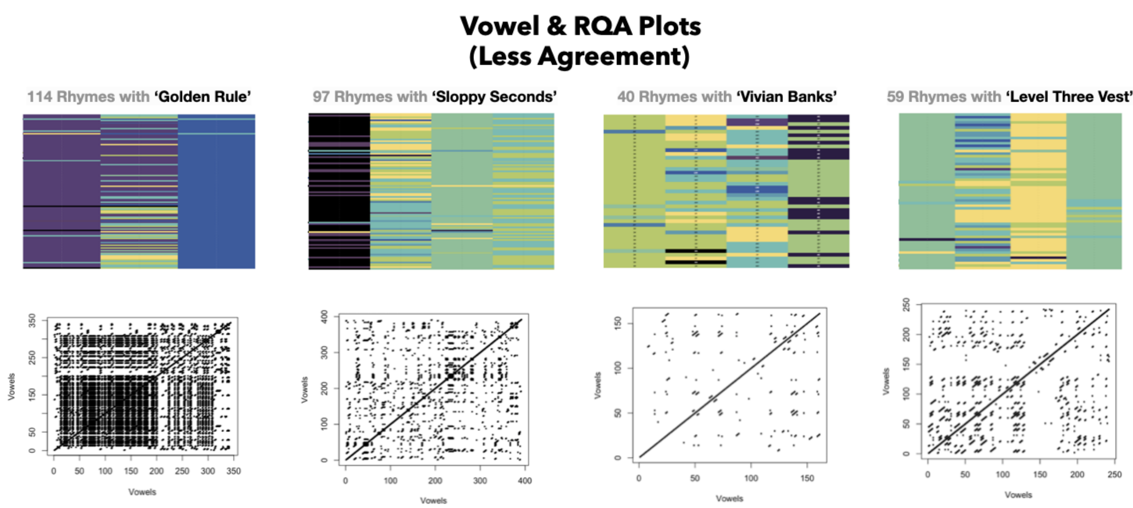


Figure 4.20: RQA Plots of Bar Pong Rhyme Sets with less similarity

However, the 4 rhyme sets of Group B (with 'less' vowel agreement) tell a different story. Figure 4.20 displays the vowel grids and vowel recurrence plots for Group B. The total number of points in the plot relative to the total possible point, also called the recurrence rate, varies significantly across Groups. While Group A sets have a recurrence rate between 21-27%, Group B rhyme sets have recurrence rates of only 2-8%. Moreover, the RQA plots for Group B are visibly less deterministic, displaying fewer long diagonals and many single dots. Group B rhyme sets display determinism between 62-95% (compared to 99%+ for Group A). Finally, Laminarity is the average length of diagonal lines with a plot. In the case of these data, a diagonal line of length 4 would indicate that the same sequence of 3 (or 4) vowel sounds was repeated 4 times in succession. Again, the laminarity of data from Group A (8-14) is much different from that of Group B (3-4), reflecting the average number of successive vowel sequence repetitions. These RQA metrics are also shown in Figure 4.21, along with Entropy. In general, entropy from RQA reflects the complexity (Shannon Entropy) of diagonal lines in a plot. rEntr, shown here, is simply the entropy normalized by number of lines in the RQA plot.

	Group A				Group B			
	Gay Sailor	Project Pat	Squid Milker	Flat Earther	Golden Rule	Vivian Banks	Level 3 Vest	Sloppy Seconds
RR	21.68	27.17	24.68	23.92	8.23	2.49	4.29	3.23
DET	99.28	99.56	99.41	99.66	94.84	74.01	62.37	65.48
L	8.71	14.00	10.06	8.28	3.63	4.00	4.26	3.57
rENTR	0.86	0.98	0.87	0.74	0.57	0.64	0.72	0.63

Figure 4.21: RQA Measurements from Bar Pong Vowel Data, Grouped by Syllable and Degree of Vowel Agreement. Group A (more vowel agreement) and Group B (less vowel agreement)

4.3.5 Conclusion

As a tool, rhyme sets (and alliterative sets and consonance sets, etc.) provide a powerful way to visualize and uncover structure across word forms. Furthermore, unordered rhyme sets can easily accommodate multi-syllabic and positional analysis, while ordered sets are additionally compatible with dynamical approaches that quantify how a system unfolds (e.g. RQA).

I began by introducing multi-syllable rhyme sets as a way to explore the structure and dynamics of complex rhyme. The underlying vowels and stress of Bar Pong rhyme sets were then plotted and the most common patterns were extracted and analyzed in terms of their frequency rank, though no obvious trends were identified when comparing rank to numbers of members in a rhyme set.

The positional predictability of multi-syllable rhyme sets was analyzed using Shannon entropy, uncovering internal structure if individual rhyme sets. Potential regularities we also identified across rhyme sets (more/less vowel agreement, terminal coda predictability, etc.). Finally, RQA revealed both visually observable and measurable differences across multi-syllable rhyme sets (Group A, Group B).

The approach taken here is also amenable to using graded measure of distance as well (instead of the identity based measures). Individual phoneme segments were my focus, but the process can be extended to representations of natural classes, distinctive features, as well as continuous similarity measures derived from signal processing of audio.

Collection and examination of additional rhyme sets would allow for a more systematic description of this phenomenon, from content, to predictability, to dynamics. In addition, these data may shed light on new poetic forms as well as enable various efforts in the study of perception, production, and learning of language.

Chapter 5

Covert Patterns

In this chapter, I examine phonemic differences in language-based music. I begin with an exploratory study which seeks to find if there are differences in phonological patterns across different genres of lyrics. Then, I zoom out, using much larger data, and compare the phoneme frequencies of lyrics with those of fiction, non-fiction, speech, and poetry. This will allow us to see how genres that reliably include language-based music (lyrics, poetry) differ from natural language (fiction, non-fiction, speech).

Questions Covered:

- (Section 5.1 & 5.2) How predictable are phonemes across musical and language genres?
- (Section 5.2) Are there cognitive or task effects related to language-based music?

5.1 Entropy of Sounds: Sonnets to Battle Rap

How predictable are phonemes sequences across musical genres?

Poetry and lyrics across cultures, from Sonnets to Rap, demonstrate an obvious human cognitive capacity for the perception and production of various multi-syllable sound patterns. Here, I use entropy to measure discrete serialized sequences of phonemes to explore the complexity of these sound structures across genres of creative language arts. This constitutes an investigation into the predictability of covert phonological vocabulary. The present exploratory analysis has two main objectives. First, the aim is to broaden the scope of cognitive processes and data that are considered in statistical learning approaches to phonological learning and language acquisition. Second, I hope to provide a basis for more targeted investigations of these patterns. Specifically, I compare the conditional entropy of segment sequences in lyrical genres. In general, Battle Rap (vowels) and Sonnet (stress) maintain noticeably lower entropy than other genres across sequence sizes, while verbal sounds from Electronic music and Hip-Hop display relatively higher entropy.

Sound patterns that use stress, rhyme, assonance, and consonance are common in language art practices across cultures. As genres like Hip-hop, Rap, and improvisational rhyming trend toward use of larger rhyming patterns than their literary cousins, many questions arise about the perception, production, and complexity of these structures.

C.E. Shannon estimated the source entropy of English characters, using human guessing, to be between 0.6 and 1.3 bits per [orthographic] character [321]. In 1965 Kolmogorov noted that while English characters (at the time) had an estimated source entropy of 1.9 bits per character, it is likely that works from artistic disciplines, such as Sonnets, would have more constraints (predictability) and therefore, should have a lower source entropy, between 1.0 to 1.2 bits per character. [12].

Since then, many better estimates of the entropy of English have been calculated [322, 323], along with numerous linguistically driven information theoretic studies [324]. Work focusing on sequences of vowels and consonants has also demonstrated the interdependence of constituent parts like vowels and consonants [325, 326]. Furthermore, the cognitive science of learnability has flourished, reinforcing the desire to explore realms of human patterning in terms of perception, production, and statistical learning [142].

In this section, I measure the conditional entropy of sound segment sequences in lyrics and poetry as shown in Table 5.1. Orthographic representations of texts are collected, but instead of analyzing the serialized orthographic characters of a phrase like "The Atomic Bomb Designer", words are serially encoded into ARPABET form (or some constituent parts: vowel, stress, consonants) to represent the sound information of the text.

Encoding	Example
Words	THE ATOMIC BOMB DESIGNER
IPA	/θə/ /ətɑmɪk/ /bɑm/ /dɪzaɪnɝ/ ...
ARPABET	DH AH0 AH0 T AA1 M IH0...
Vowel	AH AH AA IH AA IH AY ER
Stress	0 0 1 0 1 0 1 0
Cons	DH T M K B M D Z N

Table 5.1: Categories of phonological items derived from orthography. ARPABET encoding, also referred to as ALL in this text, represents the full and faithful transcription from orthography to IPA (ARPABET)

Purpose

Many are familiar with the rhyme and long range metrical constraints on language in the domains of poetry or iambic pentameter [56]. But as various multi-syllabic constraints have become common in arenas like Hip-Hop and Battle Rap, an analysis of the relative complexity of sound sequences across genre is increasingly relevant.

Non-obvious or even unintentional sound patterns throughout language arts often remain uninvestigated as the identification of marked patterns can be difficult, time consuming, and up to interpretation. Here, I propose a targeted information theoretic approach that isolates various streams of sound symbols extracted from 14 genres of lyrical texts, ranging from sonnets, to musical lyrics, to a capella Battle Rap. Evidence of underlying sound sequences, as in the example above, and other repeated phonological patterns, should be detectable through their entropy measures.

Entropy provides a tool to measure the amount of uncertainty or surprise associated with some message. Information Theory tells us that sequences of items with lower conditional entropy (conditioned on some context i.e. n-gram) are indicative of higher predictability of the elements involved. So intuitively, analysis of vowel, stress, or consonant items here can be approximated to describe the predictability of varying sizes of sounds sequences. Reductions in entropy can be understood as 'information gain'.

Approach

A number of linguistic constraints (Semantics, Syntax, Morphology, Articulation, etc...) guide word choice in the normal output of natural language. But in verbal art, phonological patterning can become paramount, giving rise to a variety of perceptually interesting patterns (rhyme, assonance, repetition)

In language arts like lyrics and poetry, multi-term sound patterns do not constrain the entirety of the signal, and authors often maintain commitments to an array of other linguistic constraints. Here, the interest in artistic sound patterns naturally focuses the investigation towards the predictability of multi-term sound structures within lyrics, represented as shown in Table 1. I predict that genres suspected to have the most formal constraints would contain more phonological regularities, and therefore, should have lower conditional entropy in these domains (at relevant sequence sizes). The idea is that, when measuring the conditional entropy of discrete sound items across genres, the statistical regularities of these sequences in a language should be captured together with whatever additional phonological predictability is specific to a given genre, artist, or work. Lower relative entropy could point to the presence of more formal constraints that exist in different streams of phonological information.

Assumptions

In comparing genres, I consider the genre that each artist produced their work within as part of the process that generates sounds with some particular transition probabilities. An author, or even a language itself, is often considered an approximately ergodic source, satisfying an important assumption of information theoretic analysis. Here, I treat each genre of expression as an approximately ergodic source in order to explore the creative sound structures that vary between them.

5.1.1 Data

I collect text from three sources, one poetry data set mined from poetryfoundation.org, song lyrics data from lyricsfreak.com, and 100 rap battles from battlerap.com. The 100 battle rap texts in question were transcribed to orthography either by battlerap.com or the performers themselves.

Sonnets and Battle Rap are the subjects of interest largely because humans can observe multi-term repeated sequences within them. However, these genres are also limiting factors in sampling for two reasons. On the one hand, very few rap battle performances have a corresponding transcript, so the number of transcribed works in this category is quite low (100 to 200 works). On the other hand, although thousands of sonnets were obtained, they have an average of only 165 syllables. To put this in perspective, most genres average 220-400 syllables per poem or song, with Hip-Hop coming in at 499 and battle rap at 4311. For the sake of reasonable comparison across genres, and with the understanding that vastly different sample lengths and alphabet sizes impact entropy scores, I report results below on the basis of data prepared as follows. I randomly select 36 works (song/poem/rap) from each of the 14 genres. From each work I extract the first 100 consecutive phoneme items, and repeat this for each sound item type (Vowels, Cons, Stress, ALL). I use CMU Phonetic Dictionary [313] to transcribe orthographies to ARPABET representaions. This allows for a simple comparison of their information across genres within a set sample size. Limiting ourselves at 100 phoneme items may not allow us to capture certain long range patterns relevant to the structure of some of these genres [327]. But this tradeoff seems acceptable, as the focus here is on the predictability of sequences of short and medium length relevant for perceptually interesting or phonologically patterned language in lyrics.

Methods

For the scope of this study, I focus on the complexity of the basic elements of sound patterns across genres. This is done in order to identify sequence sizes (phonological structures) that may be interestingly different and merit targeted investigation.

Simple information content or Shannon entropy measures (Figure 5.1) can be appropriate for exploring the complexity associated with individual items, or averages over individual items. This gives a framework for describing complexity based on the probability distribution of a variable X, comprised of a list of items x from an alphabet A. Some such studies were recently conducted focusing on the Shannon entropy and vocabulary of phenomena like improvised jazz [328, 329] and humpback whale songs [330].

As described in chapter 4, the entropy (H) is the amount of information or uncertainty in the possible outcomes of a given variable X. The variable X has an alphabet x. The entropy H of X is given by the summation p(x) times the log-base 2 of (1/p(x)) for each x in the alphabet - here giving results in the form of bits. The crucial addition for conditional entropy is that, instead of examining only one variable X in isolation, the entropy is calculated for X, given another variable Y (X—Y). For instance, using Shannon entropy one might observe the number of times the alphabet item /s/ occurs from the list of possible phonemes, however, using conditional entropy one might observe the number of times the item /s/ occurs, *given* the phoneme /p/ (only when /p/ directly precedes it).

$$\mathrm{H}(X) = \sum_{x \in A_x} p(x) \log_2 \frac{1}{p(x)}$$

Figure 5.1: Shannon Entropy

$$\mathrm{H}(X|Y) = \sum_{x \in A_x} \sum_{y \in A_y} p(x,y) \log_2 \frac{1}{p(x|y)}$$

Figure 5.2: Conditional Entropy

Here, I am largely interested in signals associated with multi-term sound patterns, Therefore, it is informative to utilize conditional entropy measures (Figure 5.2) and multiple block sizes to explore larger and larger sequences of sound items upon which to condition the prediction of the next sound item. This is a proxy for asking not just about the predictability of individual items, but about predictability of individual items in context. Following from this, I compare entropy scores by group (genre) and across sound item types (stress, vowels, consonants, ALL - a faithful transcription).

Analysis

I use Markov models of lyrics encoded as ARPABET sound items (stress, vowels, etc...) at many block sizes to extract transitional probabilities and then calculate their conditional entropy from the equation in Figure 5.2. This allows for modeling the complexity of sounds as the sound context upon which I condition increases in size.

I compare the entropy of each genre's 36 samples of 100 phoneme items in two ways. First, as in Figure 5.3, all 36 samples of 100 items for each genre are concatenated and entropy measures taken from the resulting 3600 item sample in each of the 14 genres. Alternatively, I individually take the conditional entropy of each of the 36 samples in each genre and average them to arrive at a mean conditional entropy per genre. Finally, I compare genre entropy scores in pairwise fashion using Tukey HSD pairwise tests and Jensen-Shannon distance metrics and report representative results.

5.1.2 Results

Due to the fact that I am comparing entropy across three relatively large dimensions, segment item type, block size, and genre, I do no not visually report the complete results, but relay representative trends and summaries.

Phoneme Sequence Entropy

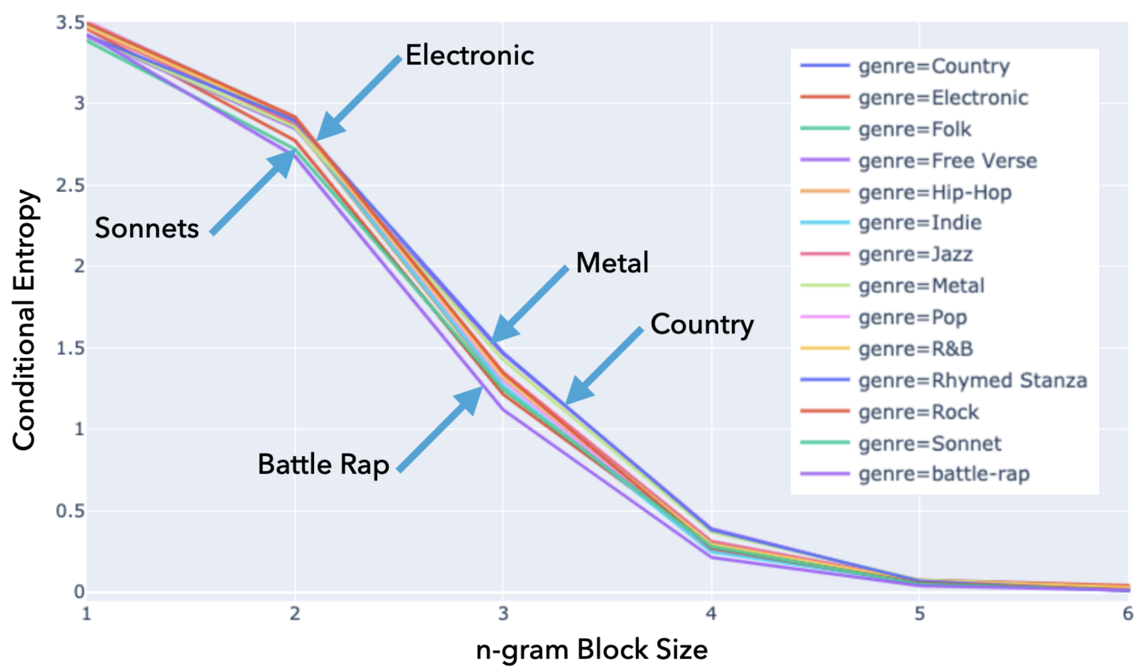


Figure 5.3: Conditional entropy of vowel sound items by genre: Single 3600 item sample per genre (concatenation of 36 samples of 100 sound items per genre). Block Sizes 1-6

Figure 5.3 Shows the decreasing conditional entropy by genre as block (sequence) size increases for the vowel sound items. All genres begin with high entropy at block size 1, but as blocks increase in size, their entropy is reduced, i.e. information is gained. The vertical spread between genres indicates that across some sequence sizes, certain genres have relatively more predictability and/or information gain relative to other genres. It is also interesting to note the sigmoid-like (decreasing function) pattern that all 14 genres follow as a group. The information gain (entropy reduction or f'(x)) from sequences of size 1 to 2 is not as great as the reduction from sequences of sizes 2 to 3 or 3 to 4. However beyond vowel sequences of 4 items, information gain slows down dramatically. This last point is unsurprising as it is not expected to find strong dependencies between phonological items at large distances.

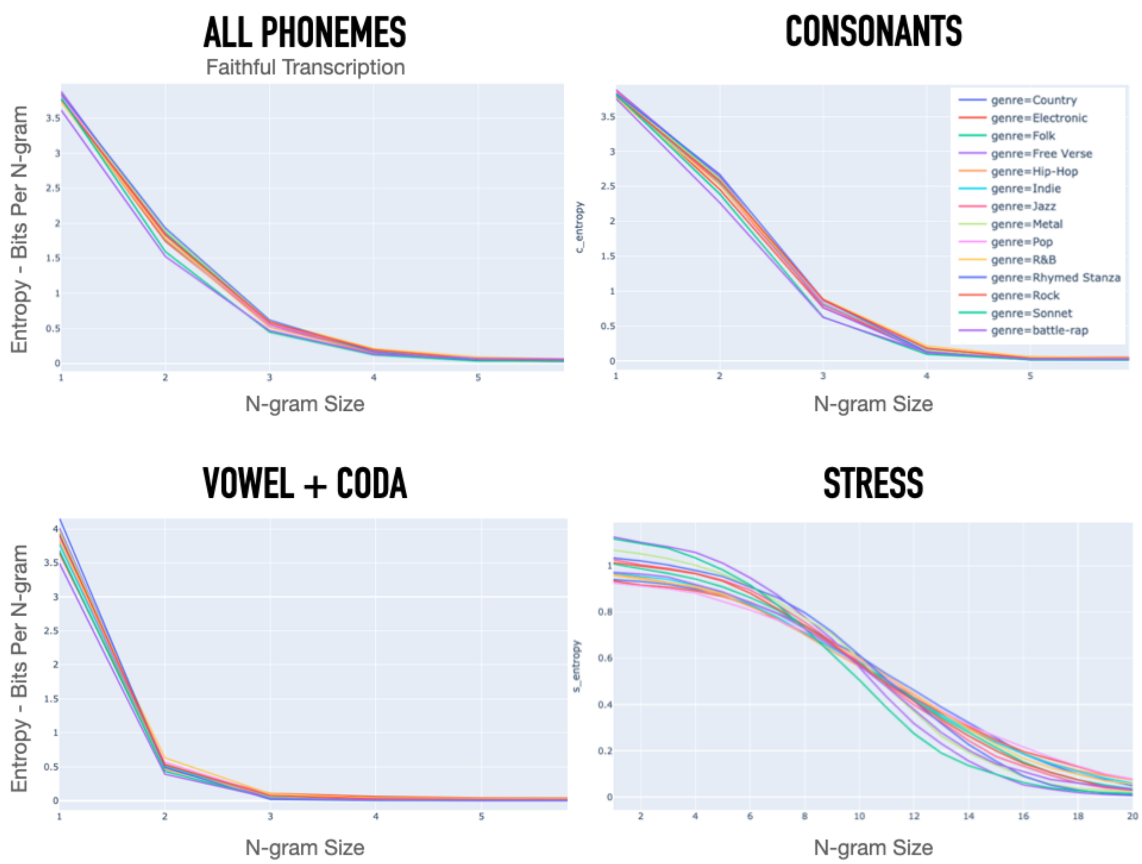


Figure 5.4: Plots of conditional entropy against block (n-gram) size for each category of representation.

Stress Sequence Entropy

In the case of stress sequences, at low n-gram size (1-3) Sonnets and Free Verse maintain the highest entropy, and at larger sequence sizes (n), they display the lowest entropy. This indicates a larger entropy reduction and therefore larger information gain in these genres than others. This is consistent with the notion that when there are larger patterns in a text than a given n-gram size can account for, entropy tends to be overestimated [331]. This would explain why Sonnet stress entropy begins higher, but as ngram sizes increase towards the size of patterns like iambic pentameter (blocks of size 10), Sonnet stress patterns are relatively more predictable, reflected by their lower entropy.

Corresponding plots of stress also display a similar sigmoidally decreasing function, while all other item categories (shown in Figure 5.4) mirror the common strictly decreasing source entropy functions.

Pairwise Tukey HSD

We might continue counting and displaying raw entropy scores for each genre, but in this section I aim to compare entropic measures in order to identify significant differences between genres.

To avoid the build up of error by repeatedly performing ANOVA tests of this kind across genre and block sizes, I use the Tukey HSD test which allows us retain statistical soundness while conducting many pairwise significance tests.

In order to see the quality of these significant relations one can also represent these pairwise significance results in network form as in Figure 5.5. Nodes represent genres that passed some significance test. A pairwise similarity matrix lets us visualize the density and quality of significance relations, where edges indicate specific significant pairs and their directed edges denote the low-high entropy relation. For reference, values within nodes convey averages of conditional entropy across the 36 100-item samples in each genre.

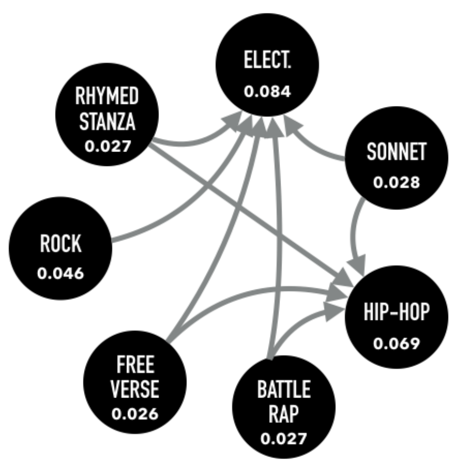


Figure 5.5: Network of passed mean conditional entropy significance tests on vowel sound items, Block Size 4. Edge origin indicates lower entropy of the significant pair, arrow's head indicates the higher entropy. Values in each node show means of conditional entropy across the 36 samples (each of 100 items) per genre. Connection counts same as column 4 of Figure 5.6.

Figure 5.5 illustrates the significant pairwise differences from these tests, but only on vowel items of sequence size 4. This approach can be used to visualize important relations across sound items types and block sizes. A fully connected graph with 14 nodes would indicate that the entropy in each genre is significantly different from every other genre. For vowel items at block size 4, Electronic and Hip-Hop have the relatively higher entropy, forming various pairwise differences with lower entropy genres, Sonnets, Battle Rap, Free Verse, and Rhymed Stanza.

Stepping back to visualize a broader picture of the range of differences across vowel sequence length, I run pairwise Tukey HSD across each genre and block size and report the counts (by genre) of pairwise significance tests passed in Figure 5.6. So then, column 4 of Figure 5.6 represents the counts of Tukey HSD pairwise tests passed and displayed in Figure 5.5. Each cell in Figure 5.6 can have a value of up to 13. A score of 13 would mean that a given genre was statistically different from all other 13 genres at that block size (column).

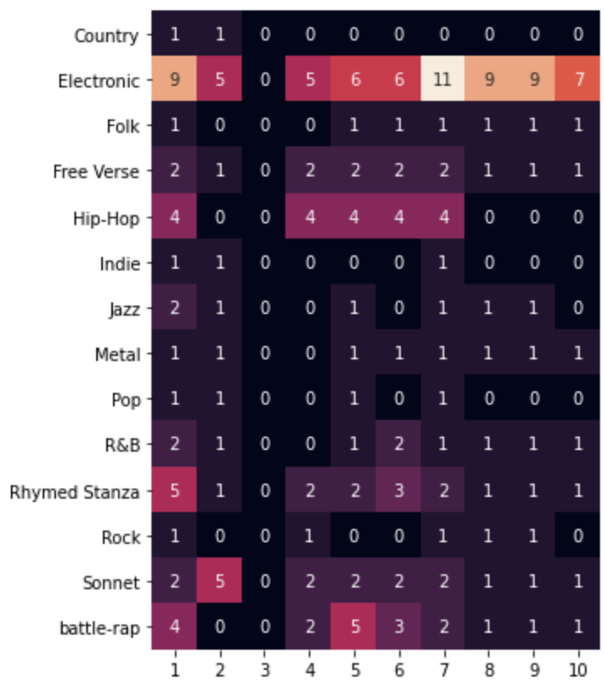


Figure 5.6: Counts of significant pairwise Tukey HSD tests - Vowel sound items - Counts represent number of pairwise tests passed, columns are block sizes

Using this approach loses dimensionality about the valance of specific pairwise relations between genres, but offers the ability to see patterns in the counts of significance tests passed by each genre as block size increases. This can provide information about which genres and which sequence lengths may stand out as interestingly different.

Figure 5.6 shows the two main groups of genres that emerge from counts of their pairwise significance tests. In general, Electronic, Free Verse, Rhymed Stanza, Sonnet, Hip-Hop, Battle Rap participate in many significant pairs. Electronic, and Hip-Hop fall on the higher entropy side while Free Verse, Rhymed Stanza, Sonnets, and Battle

Rap display lower entropy. Figure 5.6 also makes clear that there are block sizes and genres that do not accommodate many significant cross-genre relations, notably block sizes of 3 and genres with mostly 0s and 1s. Conditional entropy differences across genres seem to be described by these two clusters. One where genres find very few significant differences to any others (Country, Folk, Indie, Jazz, Metal, Pop, RB, Rock), and one where multiple pairwise differences occur, because of either high or low entropy.

Jensen-Shannon Distance

In order to arrive at an entropy based measure of similarity between genres, I use Jensen-Shannon Divergence (JSD), Figure 5.7b. JSD is a symmetric measure based on the asymmetric Kullback-Leibler Divergence (KLD) shown in 5.7a, which allows comparison of two probability distributions P and Q. Symmetry is important here because I want each given entropy metric to be the same when calculating P vs Q and Q vs P (e.g. Folk vs Country and Country vs Folk). Jensen-Shannon Divergence smooths and makes symmetric the KLD where M is (P + Q)/2. Lastly, in 5.7c the square root of JSD is taken to arrive at the Jensen-Shannon distance metric.

a) Kullback-Leibler Divergence

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

b) Jensen-Shannon Divergence

$$\frac{1}{2}D(P \parallel M) + \frac{1}{2}D(Q \parallel M)$$

c) Jensen-Shannon Distance

$$\sqrt{\frac{1}{2}D(P \parallel M) + \frac{1}{2}D(Q \parallel M)}$$

Figure 5.7: Expressions for Kullback-Leibler Divergence, Jensen-Shannon Divergence, & Jensen-Shannon Distance

Applying this measure to all genres in a pairwise fashion, Figure 5.8 shows the means of Jensen-Shannon Distance from each genre to all other 13 genres, for a given item type and block size. For example, if I individually calculate the Jensen-Shannon Distance of unigram stress items between Country and each of the other 13 genres, I would get 13 distance measures, averaging them results in a score of 0.038, as shown in the top left of the figure. This process is repeated for each genre and for 1-3 grams sequences across both stress and vowels items. It provides us with an entropy based measure that allows us to notice which genres are, on average, more different from other genres.

The yellow highlighted regions of Figure 5.8 indicate genres with the highest average Jensen-Shannon distance from other genres. In the realm of 1-3 grams stress sequences, Free Verse, Rhymed Stanza, and Sonnets are most different from the other genres, while with respect to 1-3 gram vowel sequences, Electronic, Free Verse, Sonnets, Hip-Hop, and Battle Rap are most differentiated. However, much like in the pairwise Tukey HSD comparisons, these results demonstrate *that* there is a difference, and not what the valence of the difference. For instance, both Electronic and Battle Rap vowel sequences have a relatively high average Jensen-Shannon distance from other genres, but for different reasons. While vowels in Electronic lyrics systematically display relatively higher entropy, vowels from Battle Rap display relatively lower entropy.

	1-3 stress ngrams			1-3 vowel ngrams		
genre	1	2	3	1	2	3
Country	0.038	0.064	0.086	0.061	0.179	0.489
Electronic	0.043	0.073	0.102	0.063	0.200	0.509
Folk	0.033	0.056	0.079	0.058	0.171	0.484
Free Verse	0.075	0.118	0.151	0.091	0.207	0.495
Hip-Hop	0.034	0.060	0.083	0.075	0.197	0.500
Indie	0.038	0.065	0.089	0.067	0.183	0.491
Jazz	0.038	0.067	0.092	0.060	0.179	0.489
Metal	0.049	0.078	0.102	0.074	0.184	0.481
Pop	0.047	0.076	0.102	0.067	0.182	0.489
R&B	0.035	0.061	0.086	0.068	0.185	0.487
Rhymed Stanza	0.061	0.096	0.123	0.086	0.192	0.485
Rock	0.036	0.061	0.085	0.057	0.169	0.479
Sonnet	0.064	0.101	0.130	0.093	0.218	0.514
battle-rap	0.033	0.067	0.092	0.069	0.203	0.528

Figure 5.8: Mean Jensen-Shannon Distances from each genre to all other genres. Columns represent n-gram block sizes.

5.1.3 Discussion

The conventions of language place limits on its structure, and therefore, constrain the space of likely possible messages, resulting in lower entropy. In some genres, explicit constraints are clearly defined. This suggests that there may be genre specific sound patterning constraints that are demonstrated in the predictability of their sound transitions. Although it might have been expected that sonnets and battle rap have low entropy, it is a surprise that free verse, which traditionally does not rhyme or have a regular meter, would have similarly low entropy. These trends are also broadly mirrored when considering consonant and ALL item categories.

Source entropy rates cannot reasonably be estimated the in the limit with samples this small as estimates become deterministic at low values of n. However, the perceptually interesting multi-syllable sound patterns (rhyme, meter) I am interested in comparing across genres may be reasonably represented at these low block sizes 1-10. Even relatively large structures like iambic pentameter (10 syllables per line) seem to be, at least partially, captured in the relatively lower entropy scores displayed by sonnet stress.

Using this sampling approach it seems clear that some genres do exhibit lower entropy than others. This is all the more interesting because they employ drastically different sound patterning conventions. Sonnets often have some iambic meter and end rhyme such as 'ABAB' or 'AABB' constraints. While in Battle Rap, patterns may be large and imperfect, they do not follow a standardized metrical or rhyming structure as sonnets do. However, Hip-Hop rap lyrics, which one might expect to be similar to Battle Rap, consistently exhibit relatively high entropy in both vowel and consonant item categories. Finally, Electronic lyrics stand out as the genre with highest entropy in the case of vowels.

It is also noted that many genres without observed differences from other genres (Country, Jazz, RB, Rock) have a similar historical roots in the Blues. This admittedly anecdotal observation may open a door to exploring sound entropy in terms of genealogy and the development of lyrical sound patterns diachronically.

Future Directions

Follow-ups to this study could include use of larger data sets to compare the complexity of literary genres and individual artists. Due to the limited volume of previously transcribed content, additional rap battles and improvisational performances should also be transcribed to enable further analysis. In the end, some of these qualitative descriptions of genre-based differences must be more rigorously established with larger samples and more exhaustive modeling.

It should be noted that the 100 rap battles considered represent written, not improvised content. However, it would be of particular interest to directly compare findings in the realm of pre-written lyrics to those in similar spontaneous or improvised verbal expression. This could help to tease apart the complexity of improvised vocabularies of creative language from those involving large amounts of human engineering (i.e. explicitly contriving and following some pattern without time constraints).

Many questions remain outstanding. For example, what are these specific phonological constraints, what is their vocabulary, and how are they perceived, learned, and produced? This work should be done in conjunction with various phonological, behavioral, and computational investigations. It is also important to understand how the constraints in creative sound sequences interact with other components of language to produce complex dynamics (syntax, morphology, or phonetic inventory).

As noted in previous work [325, 326], there are interactions between sequences of vowels, consonants. This leads to the natural question, how do the patterns presented here hold when transitioning from simple phonological items (stress, vowels, consonants) to more complex phonological items like syllables (consonant-vowel-consonant - CVC), rhymes (rime - vowel-consonant - VC), or specific variations such as masculine and feminine rhyme.

Conclusion

The present work has described a methodology for systematically investigating entropy of sound along three dimensions, source (genre), sequence size (n-gram blocks), and phonological items (vowels, consonants, and stress sequences). In sum, I have shown that some genres (Sonnets, Battle Rap) have relatively more predictable sequences (lower entropy) at n-gram sizes consistent with their known poetic structures, while other genres, notably Electronic and Hip-Hop, exhibit markedly less predictability.

5.2 Phoneme Frequencies

How predictable are phonemes across musical and language genres?

Are there cognitive or task effects related to language-based music?

While investigating the use of sound sequences in language is important, it is also relevant to document the phonological vocabulary of language-based music at the level of individual phonemes. I investigate the phonemes of 5 genres (Non-Fiction, Fiction, Speech, Poetry, Lyrics) to uncover systematic trends in the language of musical lyrics and poetry. This allows for uncovering insights about how the task and cognitive constraints of forms like musical lyrics and poetry may be reflected in language.

For instance, some sounds are more singable than others, a property sometimes connected with the linguistic concept sonorance. The more sonorant a sound is, the more its production can look similar to a sung note (that can be held). In English, all vowels are sonorant, while only some consonants are sonorant. One might suspect that singers have a bias towards using more singable sounds. An intuition for this bias can be captured in a simple exercise - try singing the following non-sonorant sounds /p/, /t/, or /k/, then try singing the following sonorant sounds /m/, /n/, /l/.

The sonorant consonants can be held as a sung note, while the non-sonorant sounds cannot be. Here, I inspect the sounds of 5 genres to determine if the sounds used in language-based reflect a bias towards sonorant sound segments.

In this section, I first present a small dataset of Diverse Creative English Texts (DCET), which offers an opportunity to explore differences across 5 categories of language (Non-fiction, Fiction, Speech, Lyrics, Poetry). Second,

Given the creative (and sometimes subtle) use of phonological patterns across these categories, I explore phoneme term frequency rate features as a way to check for systematic biases in the data. I also test these features in text classification and compare them to a UCI dataset containing 5 categories of non-fiction [332]. This constitutes simple bag-of-words models using phonemes, letters, and words as feature sets.

5.2.1 Data

All orthographic texts used here are encoded into IPA with the CMU pronouncing dictionary [313]. In cases where words are not available in CMUdict, I employ a Grapheme2Phoneme model to encode misspellings or uncommon words [109].

Diverse Creative English Texts (DCET)

In some domains (Lyrics, Poetry), phonological patterns are employed explicitly and liberally (e.g. poetic devices). In other domains (Non-fiction, Fiction, Speech), phonological patterns are omnipresent, but may be more subtle or simply reflective of the phonology of natural language.

The current dataset represents 60 corpora across a diverse range of language (non-fiction, fiction, speech, musical lyrics, poetry) and serves as a test-bed for exploring a variety of effects related to language, cognition, and creativity. An important distinction between Poetry and Lyrics categories here is that they reflect non-musical and musical samples respectively. DCET provides broad enough samples to test whether phonemic usage differs across genre. More details on the dataset can be found in Appendix Fig. A.1.

For this investigation each corpus is limited to 14K words. From these words, 14K vowels and 22K consonants are extracted. Every unique vowel and consonant segment is present in every corpus (document).

UCI Bag-of-words Dataset

The UCI machine-learning BOW dataset [332] represents 5 sub-categories of documents within Non-fiction, and offers a way to investigate whether phonemic features generalize from a phonologically creative dataset (DCET) to more common scenarios. The UCI database covers a range of Non-fiction where there is little reason to expect that phonemic features should be a strong predictor. I select 5900 documents from UCI across NY Times articles, Enron e-mails, NIPS full articles, PubMed Abstracts, and KOS Blog entries. In general, one might expect that the phonological biases of Musical and Poetic genres do not appear across the UCI data, and that therefore, phonemic features may be less revealing.

Methods

Traditionally, bag-of-words features are encoded using Term Frequency - Inverse Document Frequency (TF-IDF). Inverse Document Frequency is critical to balancing features with bag-of-words as some words appear in most documents, while others appear in no documents. The number of unique phonemes terms in this investigation is small (15 vowels, 24 consonants), and every phonemic term appears in each document. Because of this, I focus on simple Term-Frequency and only introduce TF-IDF processing for lexical items and performance comparison. TF-IDF features are not further transformed.

Phoneme rates are normalized (0,100). Before processing, phoneme rates are additionally scaled to shift the distribution to a mean of zero and a standard deviation of one.

Bag-of-phonemes features are evaluated with both unsupervised (PCA & K-Means) and supervised methods (Logistic Regression, KNN, SVM, Random Forests) .

5.2.2 Distributions

Word uni-grams follow a Zipfian distribution, implying an inverse relationship between word rank and word frequency [333]. This sort of distribution means that the second most frequent item (word) is one-half as frequent as the most frequent item, and the third most frequent item is one-third as frequent as the most frequent item, and so on (the nth most frequent item is 1/n as frequent as the most frequent item). it has also been shown that this Zipfian relationship actually breaks down when there are a huge amount of items [334] (i.e. including all words in a language rather than just the 20 or 30 thousand most common ones). However, Zipfian distributions do apply (again) to the entire lexicon of a language when including phrases (in addition to words) into the rank-frequency plots [335, 336].

It is believed that both letter [337] and phoneme uni-gram frequencies [338] do not follow a Zipfian distribution, but rather a Yule distribution (similar to Zipf, but with preferential attachment - leading to the "rich get richer" phenomenon). It seems that "Zipf's law applies less strongly to phonology", suggesting that sounds are not as well described by scaling laws as words [339].

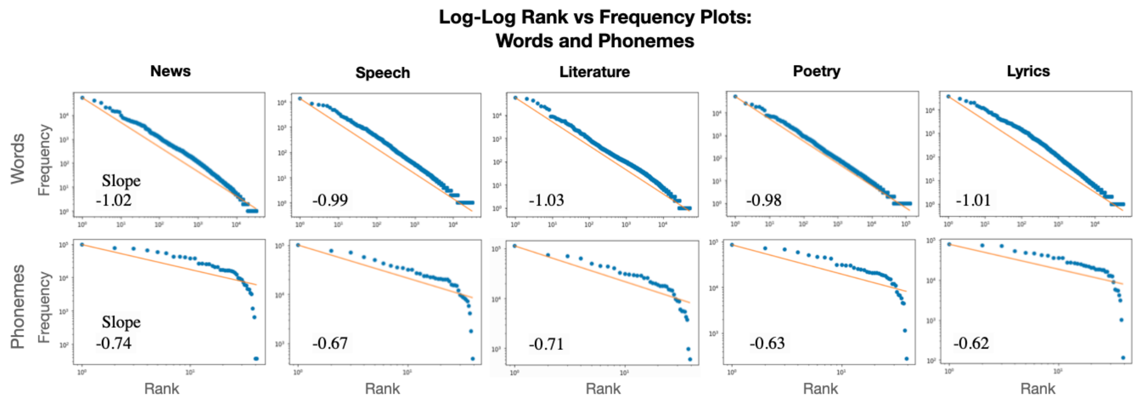


Figure 5.9: Rank vs. Frequency Plots of words, phonemes drawn from words. Number in bottom left is the slope

Figure 5.9 shows the rank-frequency plots of items (words, phonemes [39]) from across these 5 genres. These data are presented in order to demonstrate the generally expected distributions of lexical and phonemic items in this composite data set.

Across DCET words seem to follow a Zipfian distribution. This is usually determined by examining the slope of the log-log plot of rank vs frequency. If the slope is -1, then the distribution is assumed to be Zipfian. Conventionally, the slope of the log-log line has been used as an approximation of alpha in the power law equation $f = (c * (rank + b)^a)$. But there are other, more reliable ways to fit power laws [340, 341]. The distribution of phonemes across these data sets is also visually consistent with that of a Yule distribution.

5.2.3 Results

In this section I report notable trends across Syllables Per Word (SPW) and Consonants Per Vowel (CPV) features before discussing and modeling phoneme term rates (vowels and consonants).

Syllables Per Word (SPW)

The SPW measure uses syllable count to characterized the average length of words in a sample. As shown in Figure 5.10 (Top), Non-fiction boasts the largest average syllable length with a mean of 1.61 syllables per word (SPW). The Lyrics category has the smallest average SPW (1.25). Infant Directed Speech (Speech) and Battle Rap (Poetry) corpora display the fewest SPW with 1.2 and 1.19 respectively. The King James Bible, though classified here as non-fiction, displays a notably lower SPW than all other non-fiction samples.

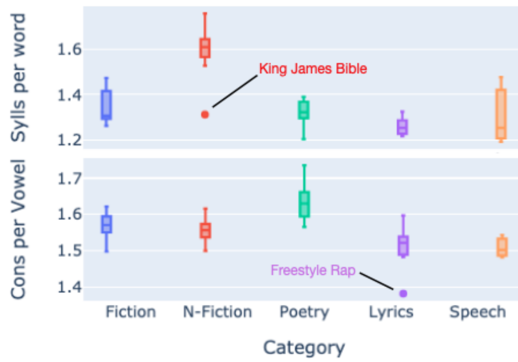


Figure 5.10: Syllables Per Word & Consonants Per Vowel Across 5 Genres

Consonants Per Vowel (CPV)

The CPV measure reveals the relative usage of consonants and vowels across a sample (Figure 5.10 Bottom). In general, the Poetic genre displays the greatest use of consonants with a mean of 1.63 CPV, while Lyrics and Speech utilize the fewest consonants with a mean CPV of 1.57 and 1.51, respectively. The Freestyle Rap (Improvised) corpus has the lowest CPV by a large margin at 1.38.

Phoneme Term-Frequency

One of the simplest forms of characterizing phonemic features is to capture their relative frequencies or rates of appearance in each document. Figure 5.11 shows frequency rates of 2 of the 15 vowel phonemes considered. Note that AH (which represents /ʌ/ + /ə/) is used most in the Non-fiction category, and used least by the Lyrics category. The opposite pattern is seen in Non-fiction and Lyrics for /aɪ/.

All phoneme frequency rates for consonants and vowels in DCET are shown in Appendix Fig. A.2 & A.3. They are processed separately for interpretability. The data in these charts serve as features for further bag-of-phonemes investigation. All 39 phoneme terms are checked for significance (Appendix Figs. A.6 & A.7) using a Bonferroni adjustment (p-value = 0.05/39 or 0.00128). 13 of 15 vowels and 18 of 24 consonant phonemes are shown to have significant group differences across the 5 categories in DCET.

Vowel phoneme rates are shown in the chart below, higher rates are denoted by darker shading. Notice the general visual similarity of phonemes usage rates within genres.

Figure 5.11: Usage Rates for 2 Vowels Across 5 Genres

In addition, some phonemes are over or under-represented in certain genres. For instance, in lyrics, /aɪ/ and /oʊ/ are used at a higher rate than in any other genre. At the same time, lyrics utilize relatively fewer 'ER' sounds (rhotic vowels /ɝ/ and /ɚ/) than other genres. Vowel and consonant term rates for each category in the UCI dataset are also included (Appendix Figs. A.5, A.4).

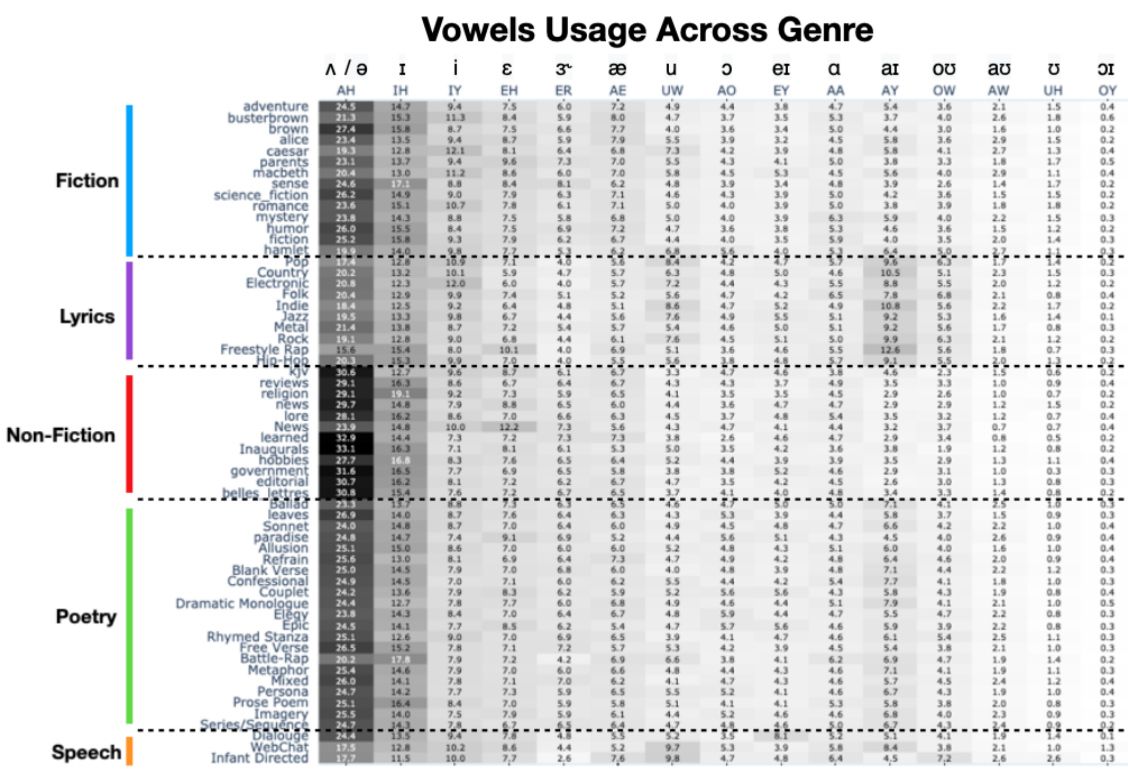


Figure 5.12: DCET Vowel Term Frequency Rates by category and corpus

Principal Component Analysis

Principal component analysis allows for taking this term frequency data and compressing it in a way the results in features which are as uncorrelated as possible. This procedure is particularly useful for determining how many features are required to explain the variance in the original data. From 39 phoneme segments I create 39 principal components (which explain 100% of the variance in the original features). The top 10, top 5, and top 2 PCA features explain 92%, 83%, 67.3% of the variance, respectively.

In order to leverage the interpretability that PCA offers I take the top two components (67.3% of the variance) and plot the 60 corpora in this two-dimensional space in Figure 5.13. The top-left of Figure 5.13 shows how DCET's 5 categories are clustered along these two PCA dimensions. Two phonemes, AH (/ə/ + /ʌ/) and /aɪ/, are most strongly correlated with PCA components. The rest of Figure 5.13 is a zoomed-in version of the top-left where we can more clearly see the *loadings* (Eigenvectors * $\sqrt{Eigenvalues}$). These loadings describe how much each variable (phoneme) contributes to a particular principal component. Large loadings indicate that a particular variable has a strong relation to a principal component; its sign indicates a positive or negative correlation.

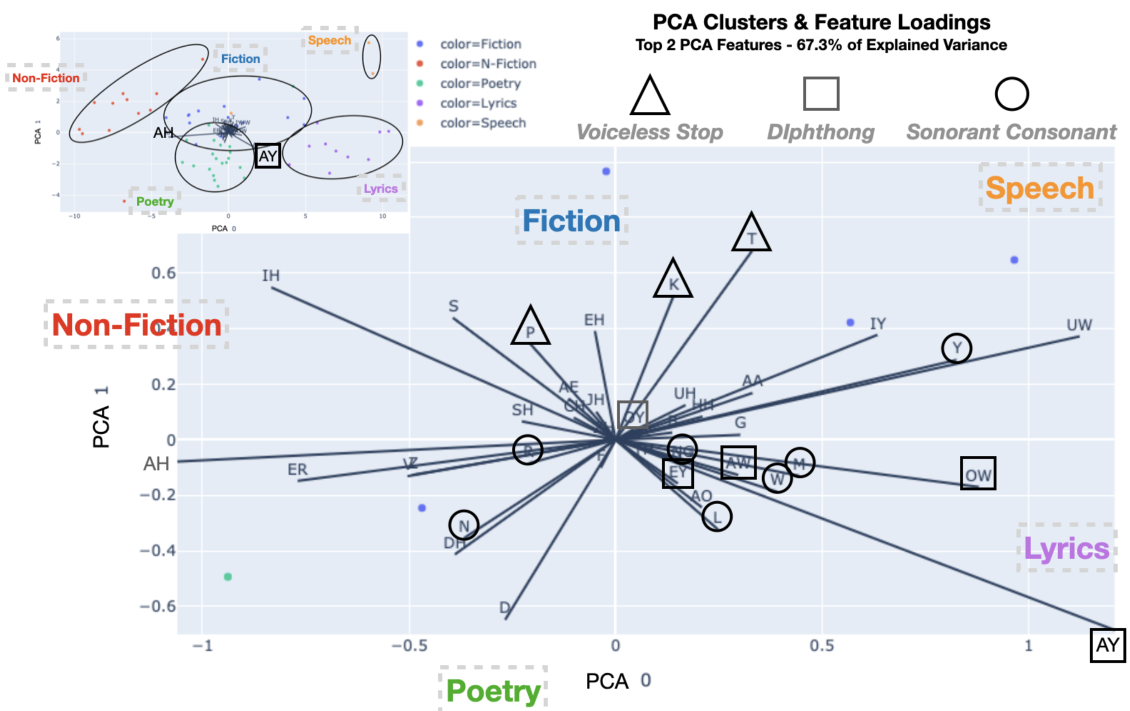


Figure 5.13: PCA 2 Components. Top-left is a zoomed out version of the main figure. Corpora are plotted as color-coded data-points and loadings are plotted to visualize phoneme correlation with PCA features.

Intuitively, one might expect sung lyrics to be positively correlated with sonorant consonants (the most singable consonants) or diphthongs (vowels that start as one sound, and end as another). Furthermore, the task of singing seems to heavily leverage continuous sounds, so classes like voiceless stops may be anti-correlated with Lyrics.

To observe some of the linguistic trends in these data I encode all phonemes that belong to these three select natural classes; Voiceless stops /P, K, T/ in triangles, diphthongs /AY, EY, AW, OW, OY/ in squares, and sonorant consonants /M, N, NG, R, L, W, Y/ in circles. Indeed, Figure 5.13 highlights that most diphthongs and sonorant consonants are positively correlated with Lyrical corpora (loadings point toward the bottom-right, where *all* lyrical samples exist in this 2D PCA space). This is not the case for voiceless stops, which cluster together in the top-region, /P/ being most anti-correlated with Lyrics. In this space, the sonorant consonants /R/ and /N/ also appear somewhat anti-correlated with lyrical corpora, /N/ being positively correlated with Poetry.

K-means

Our 5 human-labeled categories in DCET may not be the best separation of the data with respect to phoneme frequency so the Elbow method is used to determine the optimal number of k (k=13) clusters for K-means. These 13 clusters, shown in Appendix Fig. 5.14, offer intuitive separations of the data. Lyrics, Fiction, and Non-fiction are each highly associated with two clusters. Almost all poetic corpora are grouped together, while Battle Rap (Poetry) and Hip-hop (Lyrics) are also clustered together. Finally, 5 corpora are classed into clusters by themselves.

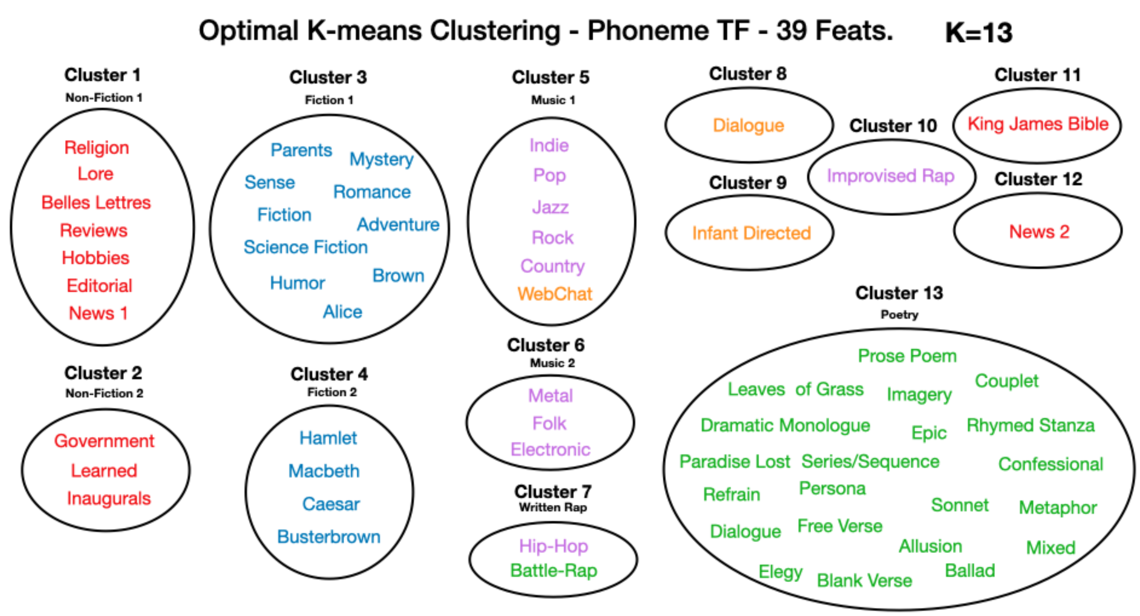


Figure 5.14: Optimal k=13 clusters (elbow method) produced by k-means. Corpora names are color-coded based on DCET category (genre).

	Feature Type	Feats.	**Model**		**Acc %**		
			LR	RF	SVM	KNN	AVG.
DCET	Phoneme TF	39	**86.7**	**86.7**	**86.7**	**83.3**	**85.9**
DCET	Letter TF	26	85.0	81.7	80.0	80.0	81.7
DCET	Word TF-IDF	300	*60.0*	83.3	80.0	80.0	75.8
DCET	Word TF-IDF+PTF	339	86.7	85.0	86.7	83.3	85.4
DCET	Word TF-IDF+LTF	326	85.0	83.3	80.0	80.0	82.1
UCI	Phoneme TF	39	79.5	84.2	87.7	*68.8*	80.1
UCI	Letter TF	26	76.9	84.3	86.1	*73.5*	80.2
UCI	Word TF-IDF	300	**93.2**	**92.3**	94.1	80.3	90.0
UCI	Word TF-IDF+PTF	339	**95.7**	**93.7**	91.2	70.1	87.7
UCI	Word TF-IDF+LTF	326	**96.6**	**95.2**	92.5	75.1	89.9

Table 5.2: 5 Feature Sets. 4 Supervised models using properly held out kfold cross validation: 80/20 train-test. Smaller DCET size dictates k=3 kfolds for cross-validation, UCI cross-validation uses k=10 kfolds. Models are as follows LR - Logistic Regression, RF - Random Forest, SVM - Support Vector Machine, KNN - K-nearest neighbors

4 Models x 5 Feature Sets x 2 Databases

In Table 1, I separate relevant training features into 5 sets and use 4 types of supervised ML models to test classification performance across datasets. The model parameters could be better optimized. In addition, this TF-IDF procedure does not eliminate documents and only considers unigrams.

Across DCET, phoneme term frequency is a better predictor of corpus category than either Letter TF or Word TF-IDF. Models using Word TF-IDF performance were also not improved upon by adding either phoneme or letter features.

Results are different for the UCI dataset. Phoneme and Letter TF alone have similar performance with average accuracy at 80.1% and 80.2% respectively. In addition, in 2 of the models (Log reg. and Rand. Forest), adding phoneme or letter features to a set of TF-IDF lexical features did improve performance.

5.2.4 Discussion

Various effects related to Lyrics and Poetry make the observed associations between certain classes of phonemes and language genres seem defensible; First of all, Poetry utilizes the most consonant sounds, which is consistent with alliterative and rhyme patterning. Musical Lyrics utilize the fewest consonant sounds overall (consistent with related difficulty in singing consonants), but still draws more heavily from sonorant consonants than any other genre. In artistic contexts, phoneme frequency features can even serve as useful features for classification, indicating productive separation of the data in both supervised and unsupervised methods; PCA shows strong correlations between natural classes of phonemes (e.g. diphthongs, sonorants) and particular text categories (e.g. Lyrics);

Task effects may also be at play; Battle Rap may present the fewest SPW since its multi-syllabic rhyming constraints are easier to satisfy by combining shorter words; Improvised Rap utilizes the fewest consonants, perhaps due to a strong online priority for vowel matching. At the least, these results motivate further exploration of the phonological vocabulary of language-based music, not only at the level of individual phonemes, but also at the level of covert sequences and overt poetic devices.

Furthermore, in more conventional contexts (e.g. UCI dataset of all non-fiction) which do not contain language-based music, these strong associations of natural classes and genre are not found, nor do there seem to be performance benefits from using bag-of-phonemes models. However, this result should be tested on a more comprehensive dataset such as Pile [342], especially in comparison to larger data sets of poetry and lyrics.

Chapter 6

Improvisation

Although overt and covert sound patterns in language are an important way to frame the investigation of language-based music, they largely reflect premeditated or written language. A focus on online language production tasks is also critical to understanding this space, and improvised rap is a lyrical online production task that reflects many aspects of language-based music. Although the cognitive mechanisms of improvisational rap have only just begun to be investigated using EEG, the improvised lyrics themselves have yet to be studied. In this chapter, I explore improvised language production in the context of rhyme and rap.

In Chapter 6, I present case studies of improvised rhyming (rapping) samples using methods from both Chapters 4 and 5. First, I consider differences in the phonological vocabulary between written and improvised verses in a rap cipher. Then, I examine samples of improvised rap, using longitudinal data to uncover changes (over time) in phonological and lexical complexity, as well as in semantic relatedness. In both cases I frame these studies in terms of explore-exploit dynamics.

Questions Covered:

- How is improvised rap different from written rap?
- How does improvised rap change over time?

6.1 What's Different About Improvised Rap?

How is improvised rap different from written rap?

Improvised rap is a notoriously elusive skill. Although it is often associated with simple rhymes or imperfect language, it is a cognitive ability that can be acquired to an extremely high level of proficiency. Online speech production (dialogue, conversation, etc...) has been studied from a variety of perspectives, though few efforts have focused on improvised lyricism [343]. Although written rap lyrics have been the target of some investigations, studies usually highlight musical features like meter [24, 25], imperfect rhyme [84, 45], rhyme identification [61, 60], or genre comparison [273]. However, the phonological vocabulary of rap lyrics, both written and improvised, have yet to be thoroughly considered or compared.

Differences between spoken and written language are well documented, in both native speakers [344, 345] and second language learners [346], generally highlighting the presence of larger linguistics structures in written forms (e.g. word length, sentences length, idea units). The ability to phonologically (and otherwise) pattern language on-the-fly is itself a kind of auxiliary skill that one must learn in order to become fluent in improvised rhyming. This cognitive process is related to, but quite distinct from, the processes of writing, memorizing, or performing rap lyrics. And the distinction between these types is often a matter of debate. In many case, rappers will even try to pawn off their written lyrics for improvised ones because it seems more impressive.

Language production, including rap, can be understood as a process of navigating the landscape of language complexity. Broadly speaking, rhyming lyrics can be generated either improvisationally, or through writing. So if we imagine that there is a space of language complexity, we could think of the way improvisers or writer are navigating the space as an explore-exploit problem.

In improvised rap, one is more constrained by the online demands of time and working memory. As a result, improvisers are predicted to travel less far in the landscape than writers and find lower peaks in the space, on average, than writers do.

But rap is not just about rhyme, it is a complex skill, emerging from many dimensions of language and communication. Here, I focus on simple linguistic features drawn from 3 of these domains, phonology, rhyme, and semantics, as a proxy for size or complexity of the language features.

More concretely then, the hypothesis is that, across all our features (discussed below), improvisers will travel less far in the space, which should result in lower variances in our selected features, and that improvisers will find lower peaks in the space on average, resulting in lower means in our features.

In this study, I focus on language features for the comparison of features of phonological vocabulary that may differ across improvised and premeditated modes of production. These features are discussed in the Data section below, but include word length, consonant to vowel ratio (phonemes), *rhyme sets*, rhymed syllable rate, and semantic similarity. Given the improvised vs. written nature of the target data, like in previous studies, it is expected to see a variety of larger linguistic structures in written rap lyrics. In addition, given the relative freedoms of written composition (time, editing, offloading, etc.) and the corresponding constraints of improvised language production (working memory, time) written rap is expected to exhibit more variation across these measures than improvised rap. For each metric, written lyrics are hypothesized to have a greater mean and a greater variance than improvised rap. This may seem somewhat counter-intuitive given the wild or chaotic nature of some improvised rapping. However, at an expert level, improvised rapping is a complex skill that seems to display a large degree self-similarity.

6.1.1 Data

Improvised rap data are uncommon and limited in size. Although no corpus of improvised lyrics exists, the smattering of freestyle rap videos on YouTube provides a reasonable starting point. The current effort utilizes a classic video from *Fresh Coast - The Documentary.* This particular rap cypher [347] is a session from 2007 that includes both improvised and written performances by the same 7 artists (Nocando, Okwerdz, The Saurus, Lush One, D Lor, Tantrum, Franco), some of the most accomplished improvised rappers in the world (though they represent the American West Coast hip-hop scene). In this section, I focus on features drawn from transcriptions of their lyrics and compare them across each artist's written and improvised verses.

This 20+ minute recording offers a window into understanding the constraints of different methods of language production. In three rounds, these seven expert rappers take turns rapping for about 45-60 seconds each round. It should be noted that these verses are all rapped to music (hip-hop instrumentals) rather than a cappella. This is a unique recording in that the first two rounds are improvised and the last one is written, allowing for more direct comparison of mode of production (improvised vs. written). I have transcribed the lyrics by hand and hand-annotated multi-syllable rhyme schemes within them (using vowel matching as the case of minimal imperfect rhyming). All other features (discussed below) are automatically extracted from the orthographic data using NLP tools, CMUdict for phoneme encoding [308] and GloVe word embedding for semantic similarity [348].

Various metrics are used to check for differences between improvised and written rap lyrics. I outline those metrics below. Group means are compared with t-tests, variances are compared using the bartlett test, and distributions are compared with the Kolomogorov-Smirnov test.

Word Length

Certain registers and genres of language have been shown to use longer words [345]. This is in part due to different vocabulary across subgenres (e.g. children's books, fiction, academic writing, etc.) and in part due to cognitive constraints. Given that all these samples are from rappers of the same era and subculture (similar register), and with similar expertise, word length differences here should be due to either individual differences or task constraints (improvised vs premeditated performance). As mentioned above, it is hypothesized that the length of words in written rap will be longer than in improvised rap. Word length here is measured in terms of syllable count.

Phoneme Distribution

The particular sounds (vowels and consonants) used in improvised vs. written rap may also be different. If certain phonemes are preferred due to task or cognitive constraints, or due to rhyme matching constraints such as neighborhood density (how many words are phonologically similar - i.e. possible rhymes), this should be detectable by comparing their phoneme distributions. Here, I use the Kolmogorov-Smirnov test to compare the distribution of written vowels to improvised vowels, and written consonants to improvised consonants.

Consonants Per Vowel - CPV

Consonants per vowel (CPV) captures the relative use of these two types of segments. A focus on patterns of vowel and consonant similarity is characteristic of both rhyme and rap, but to what degree are vowel versus consonant sounds leveraged? Vowel matching (and stress) is central to all forms of rap rhyming. Often, vowel matching alone is considered the most minimal form of rhyme (i.e. imperfect rhyme or a form of assonance). Employing consonant matching, as in conventional rhyme, adds additional constraints that may be more difficult to adhere to while improvising. Given the task constraints of improv, one might expect these performers to have a bias for vowels sounds in their improv lyrics. If vowel sounds are preferred, or consonant sounds are dispreferred while improvised rapping, one should see a difference in the ratio of consonants to vowels (CPV). The relative importance of vowels in rhyming combined with the online demands of creating sound patterns quickly during improvisation should support a bias for vowel. This measure is calculated by counting the number of consonant and vowel sounds in each sample and dividing the consonant frequency by the vowel frequency to get the average number of consonants per vowel.

Again, one might expect written lyrics to use relatively more consonant sounds (higher CPV), while improvised lyrics should use relatively more vowel sounds (lower CPV). This is a trend we already observed in 5.2.3, where Harry Mack's improvised rap was shown to have a dramatically lower CPV (1.38) than any other corpus in DCET ($\tilde{1}$.48-1.73).

Rhyme Sets

Traditional rhyme scheme annotation (e.g. ABABAB or AAABBB) is the default for capturing line-rhyme or end-rhyme patterns in poetry. But in order to document important multi-syllable structure in the rhymes of rap lyrics, a slightly higher dimensional annotation system is needed. Here, I expand on rhyme sets, groups of rhymes which are all identified to rhyme with each other. For example, simple rhyme schemes such as ABABAB, AAABBB or ABAABB can all be minimally described by two rhyme sets [A,A,A] and [B,B,B]. This form can fully represent the original ordering of a rhyme scheme by adding an indexical metadata to the rhyme set (e.g. Scheme: ABABAB $\rightarrow [(A,1),(A,3),(A,5)],[(B,2),(B,4),(B,6)]$.

But simple 1-dimensional sequences of letters (ABABAB) tell us nothing about the specific properties of the underlying rhymes. However, if the letters (e.g. A, B) are replaced with the original utterances themselves, the rhyme sets will look like this: [floor, store, more] and [Santana, jeweled bandana, new Grand Canyon].

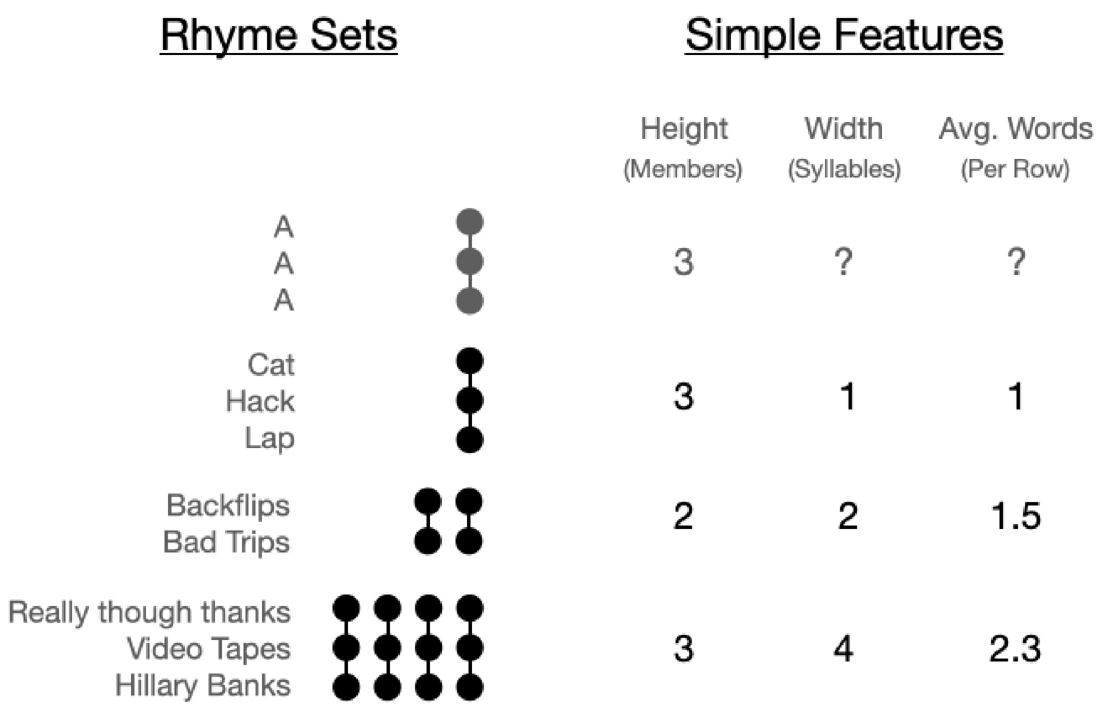


Figure 6.1: 4 example Rhyme Sets created from traditional ABAB annotation

This generalized representation is able to capture traditional rhyme notation while allowing for extracting additional linguistic features from rhyme data. Just as I encoded ABAB rhyme indices as metadata above, one can also extract a variety of other features about rhyming words from sets, including syllable count, word count, part of speech, sound similarity, and semantic features. Some basic rhyme set metrics are defined below.

Set Count is the number of distinct rhyme sets in a sample.

Set Length is the number of rhyming instances in a set.

Set Length Avg. is the average of all Set Lengths in a sample.

Set Syllable Avg. is the average number of syllables-within-rhymes in a set.

Set Word Avg. is the average number of words-within-rhymes in a set.

In this study I utilize Set Length Avg. and Set Syllable Avg. because the aim is to compare the size of these phonological structures across improvised and written lyrics. In other words, we want to know, on average, how many times rhymes are repeated, and how many syllables are being matched in those rhymes.

Rhymed Syllable Rate

In addition, this rhyme annotation (rhyme set) allows for counting the ratio of syllables that are involved in rhyme schemes across the total number of syllables spoken. This percentage is referred to as the Rhymed Syllable Rate. It gives a measure that is correlated with rhyme pattern width, but also takes into account the relative amount of non-rhyming language in each verse. In simple terms, it reveals how densely a given sample is infused with rhyme.

Semantic Similarity

The level of semantic relatedness between words is also of interest. Improv rap should have a lower degree of semantic similarity overall due to the task constraints of Improv. Semantic similarity between any two words can be measured with the standard word embedding and cosine similarity. Here I measure the cosine similarity of not only only adjacent word pairs (n,n+1), but also the similarity within a lexical window of 20 words ((n,n+1)...(n,n+20)). I take the simple average of cosine similarities across this 20 word window for each word in each verse. The word embedding model is trained on Wikipedia (300-D). In the Results section, I also outline the cosine similarity variance by artist and across window sizes (1-20) to justify using the simple average over all 20 windows.

6.1.2 Results

Recall that, across all measures (except phoneme distribution), the hypotheses are the same; the mean and variance of written rap metrics are expected to be greater than in improvised rap.

Segments

Figure 6.2 shows the syllable count for each of the 21 verses, broken down by artist and round. 19 of the verses have between 150 and 250 syllables while two improvised verses are notably longer at 429 and 367 syllables. Although these verses are not of equal length, the metrics chosen to compare them are not dependent on the size of each sample.

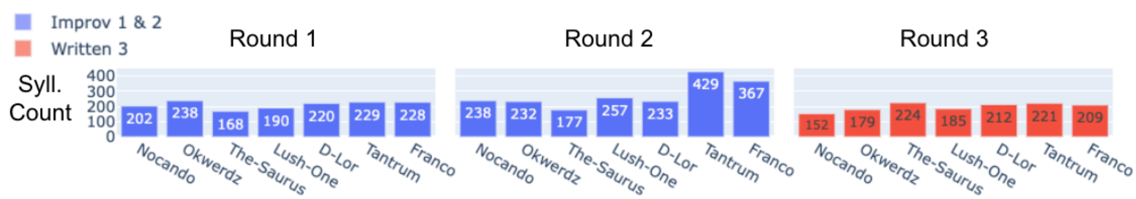


Figure 6.2: Syllable Counts for each verse by round and artist. Improvised verses are blue, written verses are red.

Phoneme Distribution

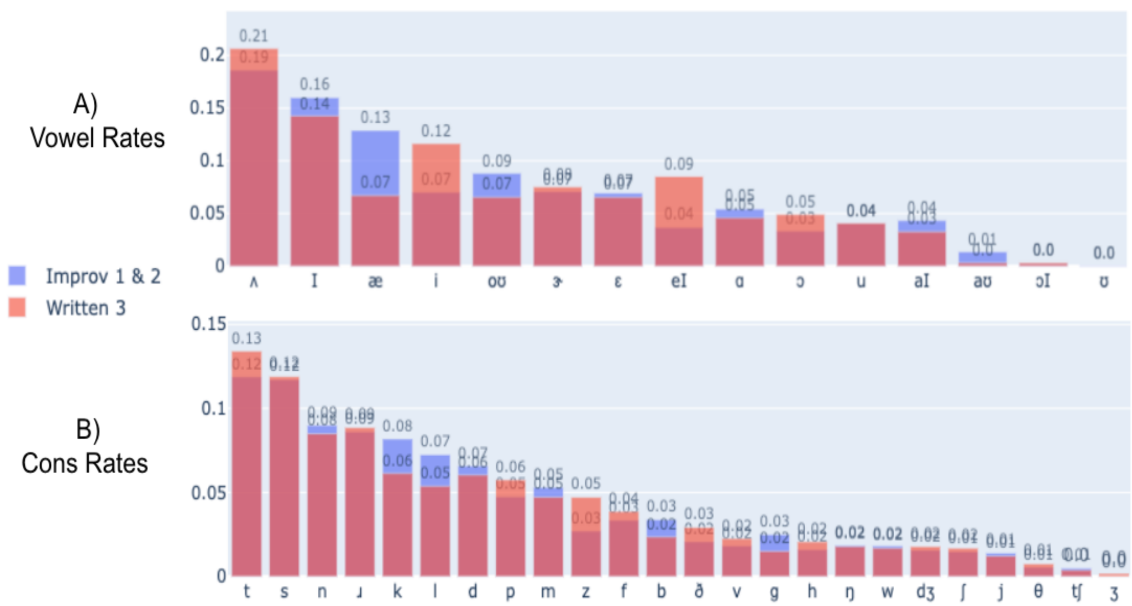


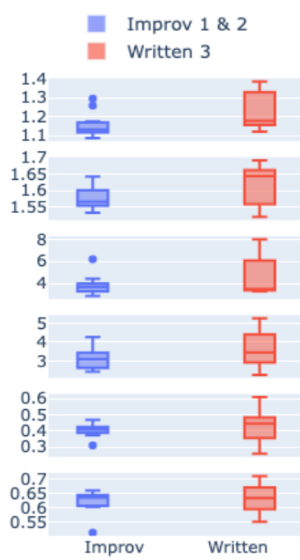
Figure 6.3: Overlapping bar-plots of phoneme frequencies across improvised and written lyrics. Figure A details vowel rates, Figure B details consonant rates.

Next, I describe the distribution of phonemes that participate in rhyme (sets). Figure 6.3A displays the frequency of vowel phonemes, 1382 vowels from improvised rhymes (rounds 1 and 2) and 611 vowels from written rhymes (round 3). This overlapping barplot shows the rates of both improvised (blue) and written vowels (light-red), as well as their overlap (dark-red). For instance, we can notice that the vowel /æ/ was used almost twice as much in improvised rhymes (13%) as in written rhymes (7%)

In similar fashion, Figure 6.3B displays the consonant phonemes rates of these samples, 2377 vowels from improvised rhymes (rounds 1 and 2) and 1086 vowels from written rhymes (round 3). Again, some phonemes such as /z/, are used almost twice as much in written rhymes (5%) as in improvised rhymes (3%).

Although many phonemes are used at similar rates, there are notable differences, as discussed above. The question remains, are the phonemes used in improvised and written samples drawn from the same distribution? The Kolmogorov-Smirnov test allows us to reject the null hypothesis that these two samples are drawn from the same distribution (vowels - pvalue=0.029; consonants - pvalue=0.0135).

The data summarized in Figure 6.3 above only concerns the distribution of phonemes drawn from rhymes. However, this difference between improvisation and written lyrics remains when considering the phoneme distributions across improvised and written verses as a whole (not just within rhymes). This applies to both the consonant distribution (pvalue=0.0154) and vowel distribution (pvalue=0.003).



	Feature	Mean	t-test p-val	Variance	Bartlett p-val
A)	Word Length	1.150 1.220	0.0660	0.003 0.012	0.0637
B)	Consonants Per Vowel	1.575 1.612	**0.0491**	0.001 0.004	**0.0397**
C)	Set Length Avg.	3.820 4.674	0.3335	0.661 3.547	0.0684
D)	Set Syllables Avg.	3.138 3.646	0.1575	0.297 1.037	**0.0142**
E)	Rhymed-Syllable Rate	0.407 0.433	0.4309	0.002 0.013	**0.0024**
F)	Sem. Connectedness	0.625 0.632	0.7113	0.001 0.003	0.2628

Figure 6.4: Primary Results: Box-plots, means, and variances for 6 metrics across improvised and written rap. pvalues for means are calculated with a t-test, pvalues for variance are calculated with the bartlett test.

Word Length

The Word Length measure in Figure 6.4A displays the expected trend of longer words in written than improv lyrics, though it is not significant (p=0.066). Word Length variance follows the same pattern (Written larger than Improv) but is also not significant (p=0.064).

Consonants Per Vowel

The consonant per vowel metric results are shown in Figure 6.4B, highlighting significant differences in both the mean and variance of relative consonant and vowel phoneme usage. This effect also holds when calculating CPV only on the identified rhyming words (rather than the entire verse) (Means: 1.75 Written, 1.72 Improv).

Set Length Avg. & Set Syllable Avg.

The expected trends (Written > Improv) are also observed in the rhyme set measures (Figure 6.4C and Figure 6.4D). However, only the variance of Set Syllable Avg. is shown to be significantly greater in written than in improvised verses (pvalue=0.0142). Although differences in their means are not significant, the rhyme sets used in written lyrics are, on average, both wider and longer (3.64 syllables, 4.67 members) than those in their improvised lyrics (3.13 syllables, 3.82 members).

Rhymed Syllable Rate

The means of Rhymed Syllable Rate indicate that 40.7% of syllables were involved in rhyme during improvisation, while 43.3% of syllables were involved in rhyme of written lyrics - though again, this difference was not significant. The variance of Rhymed Syllable Rate in written verses is demonstrated (Figure 6.4E) to be larger than improvised verses (pvalue=0.0024). This indicates that written lyrics are more variable in terms of their degree of rhyme saturation than improvised lyrics.

Semantic Similarity

Figure 6.4F highlights the average semantic similarity for each verse as measured by GloVe word embedding cosine similarity over a 20-word neighborhood window. Although neither the mean nor variance of semantic similarity demonstrate significant group differences (p=0.711, p=0.262) on the entire dataset, when removing a single outlier (Nocando Round 1), written lyrics are shown to have greater variance than improvisation (pvalue=0.0026) while their means remain quite similar. Below is a further discussion of this outlier and the striking variance of semantic similarity by verse.

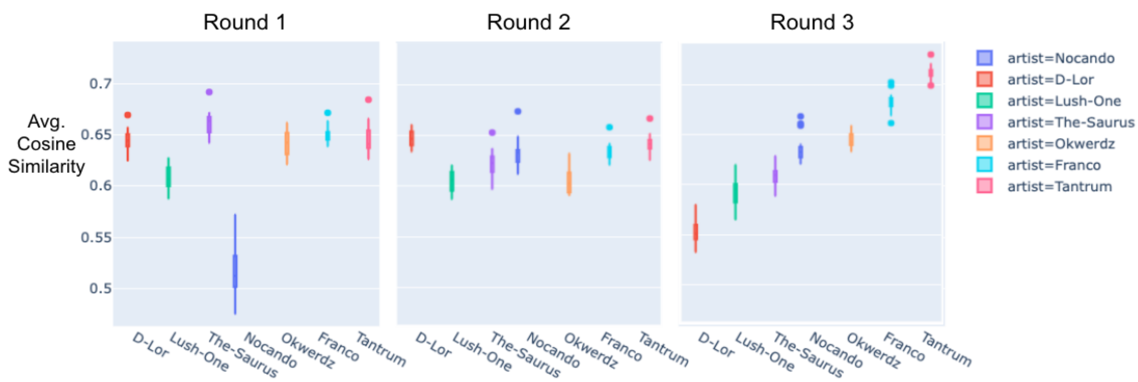


Figure 6.5: Boxplot of semantic similarity measure across each round and individual. Each boxplot contains 20 data points, the average semantic similarity for all words at each lexical distance, from 1 to 20.

Nocando is known for his sometimes off-the-wall rapping style where he incorporates impressive rhymes with surprising or unrelated rhymes. The couplet in Figure 6.6 exemplifies this attribute.

They look Juelz Santana
in a jeweled bandana in a new Grand Canyon.

Figure 6.6: Couplet from Round 1 (improvised) - Nocando

Notice the four-syllable rhyme scheme (highlighted in blue) and the grammatical well-formedness, yet the dramatic lack of semantic similarity. The average semantic similarity for this entire verse is shown in Figure 6.4F (the single blue data point below the box-plot) and in the middle box-plot of Figure 6.5 Round 1 (blue). This verse has a mean semantic similarity that is dramatic lower than other verses, while its variance is much higher (including compared to Nocando's other verses). The particularly low mean and large variance from Nocando here is consistent with the qualitative perception of this verse as all-over-the-place and semantically dissimilar.

Consider another couplet on the other end of the semantic similarity spectrum, Tantrum's 3rd round written verse (6.7). The semantic similarity boxplot for this entire verse can be seen in the right-most box-plot of Figure 6.5 Round 3 (pink). Again, the high amount of apparent semantic similarity in this couplet is reflective of the rest of the verse, which is captured by the high degree of average cosine similarity across the moving 20-word window.

Figure 6.5 shows the box-plots of this semantic similarity measure by each round and individual. Box-plots contains 20 data points, the average semantic similarity for all words at each lexical distance, from 1 to 20. It is striking to notice how the variance

When I blaze the beaten verse, you'll taste defeat and worse
Cause I ain't swallowing my pride I'd rather make you eat your words.

Figure 6.7: Couplet from Round 3 (written) - Tantrum

of semantic relatedness by verse is so small, while the means of each verse or artist can vary dramatically.

6.1.3 Discussion

Previous work has shown that in less familiar settings, improvisation becomes more predictable, as measured by entropy and conditional entropy, pointing to the utilization of more predictable prior knowledge that underlies this complex skill [252]. Although it is not clear whether the setting of this Fresh Coast Cypher would count as 'less familiar' for these rappers, the general trend of less variation in improvised lyrics would seem to indicate higher predictability, consistent with the aforementioned effect.

In addition to finding appropriate rhymes and rhythmically aligning them, one of the primary aims of improvised rapping is to generate lyrics that are coherent, relevant, and even clever. Satisfying all these phonological, grammatical, lexical, and semantic constraints online is an exceedingly difficult task. And often, generating well formed meaningful language can take a backseat to meeting phonological constraints. Understanding these constraints will require examining much larger data-sets as well as the acquisition of improvised rhyming skills in longitudinal studies (which I explore in the next section). Regardless, many of their effects are observable in the metrics chosen for analysis in this study.

In general, the observed trends are consistent with the idea that written rap displays larger structures than improvised rap. Collecting and analyzing more data than is available in this case study may be revealing and allow for higher statistical validity. The results of the Consonants Per Vowel metric suggests that written rap has a higher representation of consonant sounds. This is consistent with the prediction that improvisers should be biased towards using more vowels and fewer consonants. Furthermore, the difference in phoneme frequency distributions across improv and written rap supports the notion that these modes of production draw from different distributions, which may again reflect cognitive or task constraints.

The general trend that written rap exhibit larger variances (by the current metrics) seems plausible. Significant differences are found for Consonants Per Vowel, Set Syllable Avg., Rhymed Syllable Rate, and Semantic similarity (when removing an outlier). This suggests that there is more measurable diversity in written lyrics, while improvised rap seems characterized by less variation across metrics. It is also possible that the affordances of written composition facilitate surplus expression of individual differences in written language while the imposing constraints of improvised rap narrow its overall variability.

It should also be noted that while Set Length Avg. and Set Syllable Avg. means are not found to be significantly different, they are surprisingly larger than one might expect given the difficulty of utilizing even 1-syllable rhyming constraints during improvised language production. Specifically, it is remarkable these improvised rappers find rhymes on-the-fly that phonologically match an average of 3.1 syllables per rhyme, and then repeat those rhymes an average of 3.8 times. This consistent use of multi-syllable structures has been noted in written rap forms, but until now, had not been documented in improvised lyrics.

Contrary to expectations, the similar semantic similarity scores (means) across improv and written lyrics suggests that expert improv rappers are not making concessions on the semantic relatedness of their utterances relative to composed lyrics. However, the relatively low variance in semantic similarity across all verses also suggests that semantic continuity or self-similarity is more highly related to a given performance or verse than to an individual or mode of production. The semantic similarity measure may be reflecting some sense of theme or topicality that itself is particular to each performance (verse). Nocando's first improvised verse, and its low semantic similarity scores, may even indicate a potentially different mode of improvised rap where semantic constraints are dramatically reduced while still meeting other task demands.

Future Work

Future work should focus on collecting larger samples of improvised rap to validate these findings and to ask additional questions that are not possible with data of the scale used in this case study. in addition, collecting speech samples from these same artists, as well as longitudinal data would allow for better understanding of the relationship between the dynamics of rap and natural language, improvisation and composition. Given that these subjects are all experts, we cannot observe how the various constraints met by these artists have changed over time (especially while they were acquiring the skill). This is a point for future investigation and will be considered in the following section. Future studies may also focus on expanding the semantic similarity and rhyme set metrics to more wide-ranging data.

Conclusion

This case study identified and measured various presumed constraints on improvised language production. Using a range of metrics to quantitatively compare features of the phonological vocabulary from improvised and written rap language I uncovered perceptually observable, as well as subtle differences between them. I have shown that, even in small samples such as *The Fresh Coast Cypher*, improvised and written raps seem to be quantitatively different along a variety of dimensions including Phoneme Distribution, Rhyme Set Syllable Avg. (variance), Rhymed Syllable Rate (variance), Consonants Per Vowel (mean & variance), and Semantic Relatedness (variance).

Improvisers also seem to produce surprisingly large multi-syllable patterns on the fly, almost 4 syllables in length - even rhyming 41% of their syllables in total. This further motivates exploring the complexity and dynamics of these improvised rhyming structures, and how we produce and perceive them.

In sum, I have shown that improvised lyrics display language features with smaller means and smaller variances than written lyrics, suggesting that explore-exploit dynamics may be a productive way to frame the cognition of rapping.

6.2 Improvised Rap Dynamics Over 1 Year

How does improvised rap change over time?

Improvised rap presents a case of an understudied, yet complex, cognitive task that exemplifies various aspects of language-based music. Improvised rapping is a skill that is difficult to acquire. Most who have tried it quickly hit various walls when trying to meet one of the many constraints of improvised rapping (e.g. rhyme, rhythm, grammar, meaning, cleverness). The current study investigates the changes in some of these constraints over a one-year period for both a novice and an expert freestyle rapper. I examine how selected features of improvised language change over time in order to investigate the development of this skill.

Like in the previous study, this question can be framed as an explore-exploit problem. In the space of phonological patterns, it is hypothesised that the beginner should be exploiting the space of sound patterns, reflected by decreasing entropy measurements, while the expert is expected to be exploring the phonological landscape, reflected by increasing entropy measurements in this domain. The intuition is that the expert may be more comfortable exploring the space of phonological patterns because he can easily and confidently rhyme with any phrase he lands on, while the beginner may not have a big rhyming or sound pattern matching vocabulary. This results in the beginner being more comfortable with a certain subset of sound patterns, and so they should be expected to leverage and exploit those known patterns more heavily in order to keep up the momentum of rhyming language production. On the other hand, the expert is already comfortable rhyming with many complex sound patterns, so they may be expected to be exploring the space of phonological patterns, rather than only exploiting smaller regions of the space, like the beginner might. In other words, the complexity of the phonological elements for the beginner is expected to decrease (reflecting more exploitation) whereas the complexity of phonological elements for the expert is expected to increase over time (reflecting more exploration).

An additional prediction from this study is that both the semantic similarity of language features for the beginner and the expert will also increase over time. The idea is, as the beginner learns, they will produce more coherent and well-formed rhyming language, and their words will display more similarity (as measured by cosine similarity in GloVe word embeddings). Similarly, over the span of a year, the expert is seemingly integrating additional high level semantic connections in their improvised lyrics, such as plot and narrative, again resulting in more semantic similarity.

As discussed in Chapter 3, musicians (experts) and non-musicians display distinct brain wave activity when listening to music, implying different cognitive processes are at work across skill level. More recently, a study in rhyme facilitation within poetry demonstrated that, while both novices and experts gained memory benefits from rhyme (sub-lexical retrieval cues), only experts were shown to anticipate imminent rhyme [349]. This seems to reveal a degree of learning and cognitive specialization across experts and novices in the domain of language-based music.

Although the current study does not use a direct measure of rhymed-syllable rate (as in Chapter 6.1) since it would require annotated rhymes from these works, there is a proxy. Vowels are a structurally important part of rhyme, and a core element of imperfect multi-syllable rhyming. Therefore, I use frequencies of vowels sequence of different sizes (uni-gram, bi-gram, tri-gram, and quad-gram) to capture information about the phonological (vowel) vocabulary.

6.2.1 Data

An Expert Much like in other disciplines, expert improvisers themselves will tell you things like “I used to be horrible”, or “I have practiced for years”. But how does this rapping skill develop and change over time? One rapper, Harry Mack, a jazz drummer by training, has been practicing rap improvisation for 20 years. He recently built a large online community (improvising live online in various formats) and has a series, called Omegle Bars, where he raps for random strangers on video chat. Each video had between 10 and 25 minutes of improvised content. I collect text transcripts from the lyrics from these videos and extract lexical, phoneme, and semantic information. Texts are transcribed and submitted by YouTube user Nick @plebcrabslayer (& myself).

A Novice A fan and now team member, Ikaanic, was inspired by Harry Mack to begin learning to rap more intentionally. He has taken lessons from Harry and records his journey learning to rap in Freestyle Practice videos on YouTube. These recordings are between 60 and 120 minutes long and are transcribed automatically by the YouTube speech-to-text model.

Despite the differences in the length of these videos and their modes of transcription (human annotated vs. speech-to-text), they are decent comparables as they represent sequential improvised rap sessions over the same 1-year period. The lexical, phonemic, and semantic measures used here are robust to the mode of transcription; results were found to be similar when Harry Mack lyrics were also transcribed using speech-to-text (rather than hand annotated). Given this similarity, I assume the features drawn from speech-to-text transcriptions of Ikaanic Freestyle Practice sessions will be sufficient for comparison (as this content is much too long to transcribe by hand).

6.2.2 Methods

There are three features of language that are measured here in order to describe how improvised rap language changes over time, lexical and phonological complexity, and semantic similarity.

Entropy In the literature, increases in entropy have been associated with encouraging exploration. In human experiments, high entropy is associated with participants switching from exploitation to exploration [350]. The intuition here is that as an environment changes from a predictable one to a less predictable one, the optimal strategy to use is no longer exploitation, but exploitation, as the resource to exploit is less predictable and value can be gained by exploring a new uncertain environment. Even in reinforcement learning models, high entropy encourages exploration, allowing agents to avoiding falling into local optima [351, 352, 353].

For instance, one might assume that the unigram distribution of words is approximately Zipfian. If the complexity (entropy) of individual words (uni-gram) goes up over time, the distribution of individual words used is becoming more uniform (less predictable). In contrast, if lexical uni-gram entropy does go down over time, the distribution of individual words being used is becoming more predictable. In order to understand how the complexity and constraints of language change as one learns to improvised rap, I conduct entropy analyses on uni-gram, bi-gram, tri-gram, and 4-gram word and vowel sequences.

Semantic Similarity Just as in the previous study, the semantic similarity between words is of interest here, too. However, instead of helping to compare written vs improvised lyrics, word embeddings are used here to explore how semantic similarity changes over time. On the one hand, it might be expected that the semantic similarity of the words produced increases over time as rappers learn to overcome rhyming constraints and begin generating more thematic or related content (rather than simply satisfying rhyming constraints in a semantically unrelated way). On the other hand, semantic similarity could decrease over time as greater and greater amounts of phonological patterns (e.g. rhyme) are incorporated into language output. The intuition here is that rhyming words are more inherently semantically unrelated (except for certain cases of sound symbolism, iconicity etc.)

Semantic similarity between any two words can be measured with the standard word embedding cosine similarity. Here, I measure the cosine similarity of not only of adjacent word pairs (n,n+1), but also the similarity within a lexical window of 20 words ((n,n+1)...(n,n+20)). I take the simple average of cosine similarities across these 20-word windows for each verse. The GloVe word embedding model is trained on Wikipedia (300-D).

6.2.3 Results

Word Entropy [Figure 6.8] Over the course of 1 year, the word entropy across all sequence sizes increased for the expert improviser. This same trend of increasing word entropy was also found for the novice, but only with respect to 3 and 4 gram word sequences, for 1 and 2 gram items, no significant trend was found and the R-squared is approximately 0.

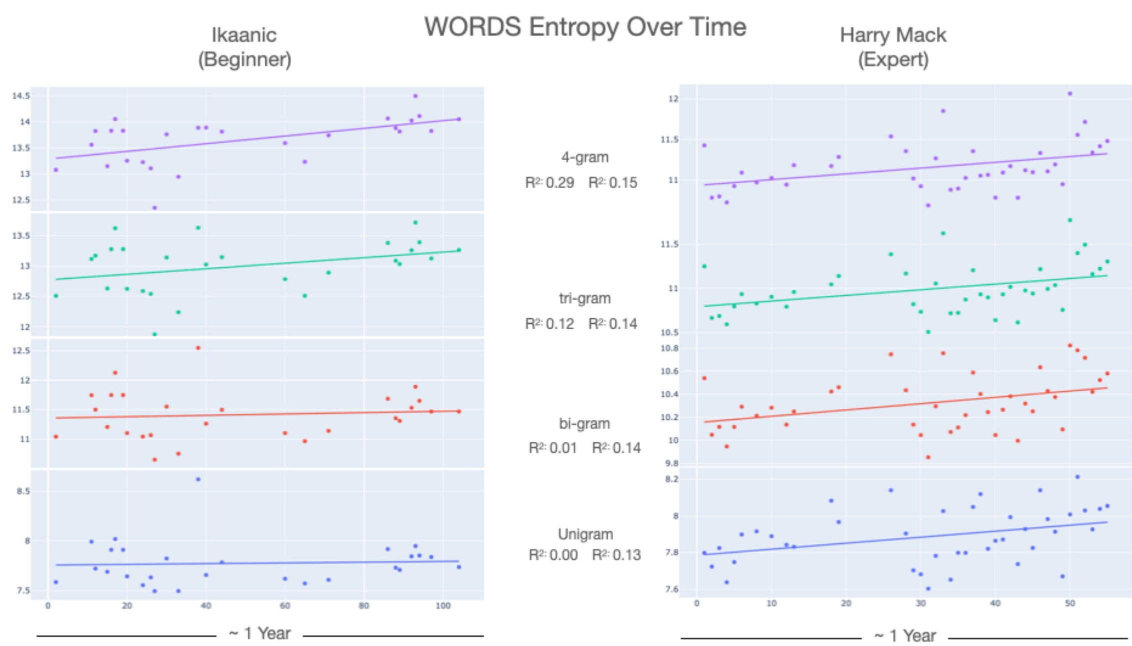


Figure 6.8: Entropy of word n-grams over time and n-gram size. Left: Novice Ikaanic. Right Expert Harry Mack

Vowel Entropy Figure [6.9] In terms of vowel sequences, the improv expert demonstrated increases in entropy over the course of this year across all sizes. These increases are also slightly larger in magnitude than the corresponding increases in word entropy discussed above. However, for the novice, the opposite trend is observed. 1-3 grams vowel sequences produced by Ikaanic are decreasing in entropy over this time period while the distribution of his 4-gram vowel sequences is not changing.

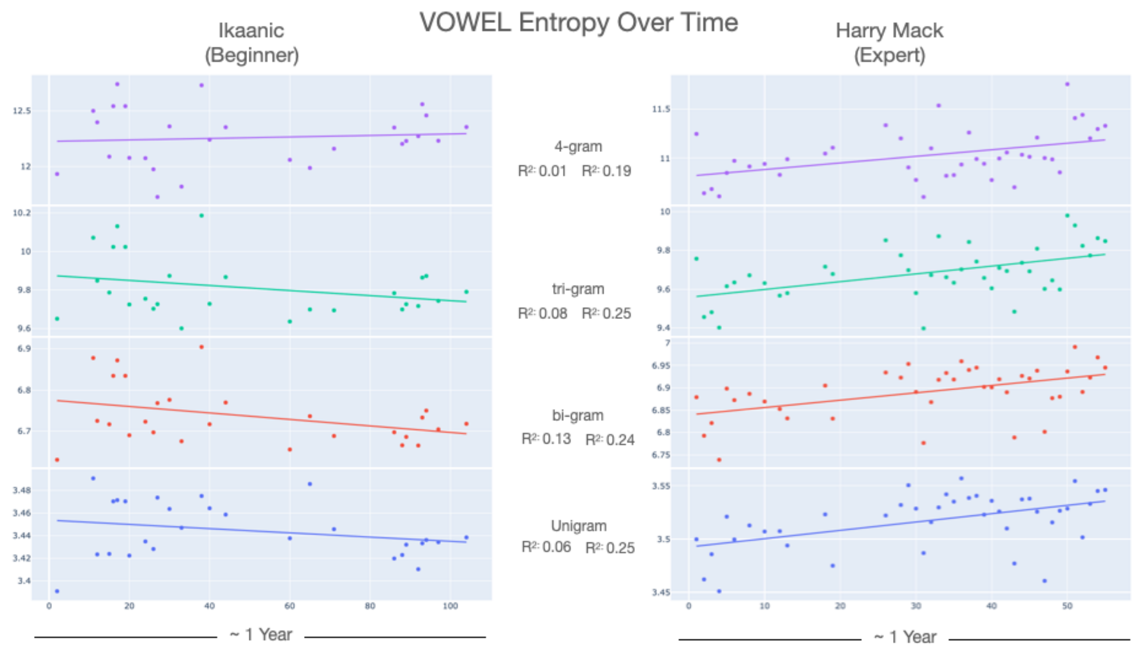


Figure 6.9: Entropy of vowel n-grams over time and n-gram size. Left: Novice Ikaanic. Right Expert Harry Mack

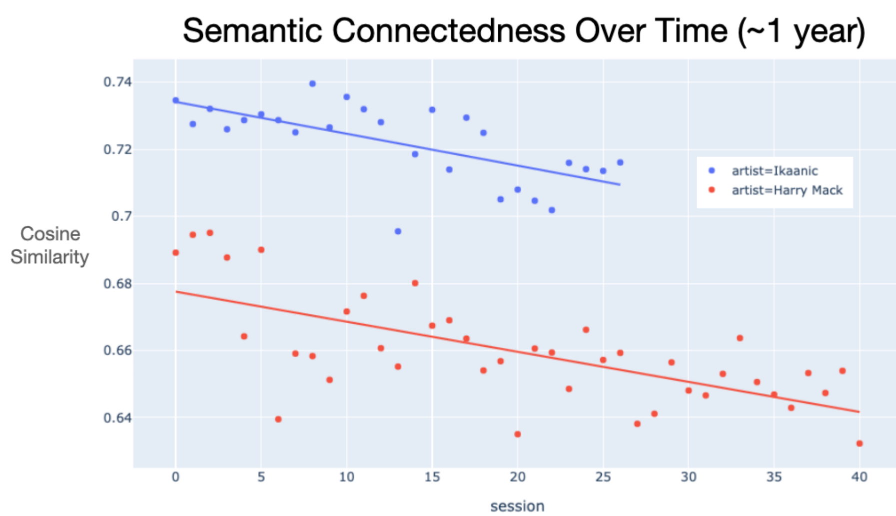


Figure 6.10: Avg Semantic Similarity (Avg. cosine similarity in 20-word window) over time. Blue: Novice Ikaanic. Red Expert Harry Mack
. Each successive rap session is shown along the x-axis.

Semantic Similarity Figure [6.10] For both the novice and expert improviser, the degree of semantic similarity is shown to decrease over time. These decreases are strongly correlated with time (R-squared 0.45) and have similar slopes, however their y-intercepts are shifted. This indicates that, while their trajectories are similar over this period, these two individuals began with very different degrees of semantic similarity, which continued to decrease at a similar rate over the year.

6.2.4 Discussion

A high-level summary of these three sets of results is shown in Figure 6.11. Notice that while semantic similarity is decreasing over time for both individuals, for the expert, word and vowel entropy are consistently increasing (all n-gram sizes). Meanwhile, the novice displays a different profile of entropy trajectories.

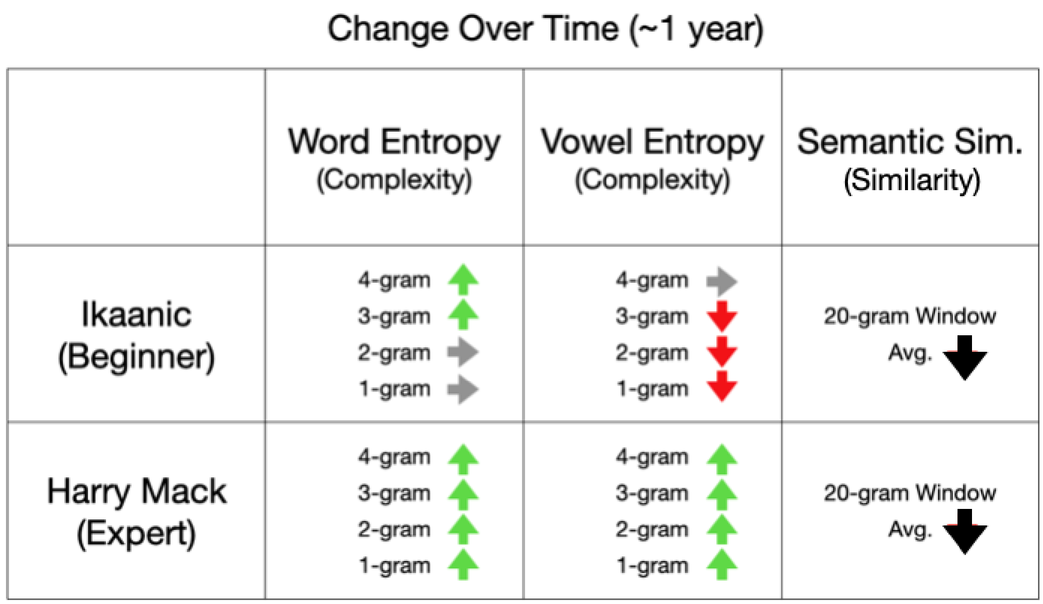


Figure 6.11: Summary of word, vowel, and cosine similarity trends over a course of a year

These data only represent a single year of development, so we are not observing dynamic shifts in these constraints over time (they are monotonic shifts). For example, one might expect that enough longitudinal data (many years) would reveal changing constraints at different periods in development – periods of entropy reduction (towards more exploitation), and other periods where entropy increases (towards more exploration).

In general, word entropy increased for both artists over the course of this year. This may mean that the distribution of used words is becoming increasingly uniform (less predictable), reflecting a more even use of words and phrases. Notably, the distribution of uni-gram and bi-gram word phrases used by Ikaanic does not change over this period. The intuition here is that, while the distribution of uni-grams and bi-grams is not changing over this year, those uni-grams and bi-grams are combined to make novel 3 and 4-gram phrases, which may contribute to Ikaanic's entropy (complexity) increases in 3 and 4-gram word phrases over this same period.

On the other hand, the expert's word entropy is increasing across all n-gram sizes measured. This indicates that, not only is his distribution of 3 and 4-word phrases becoming less predictable (more complex) over time, but his use of 1 and 2-word phrases is also becoming less predictable. Remember that increases in entropy encourage exploration while decreases in entropy encourage exploitation, so this trend across lexical items suggests that both this novice and expert are increasingly exploring (rather than exploiting) the space of words and word phrases in their rap.

Given that Ikaanic is a novice at freestyle, and just learning how to navigate the constraints and demands of improvised rap, it may be unsurprising that his vowel entropy is decreasing over this period - indicating a trend towards exploiting (rather than exploring) vowel sequences for rhyme. His vowel patterns are becoming more predictable, ostensibly shifting towards an increased exploitation of these sound patterns. On the other hand, Harry Mack's vowel entropy is shown to increase over this same time period, suggesting that he is shifting towards increased exploration of vowel sequences. This is consistent with Harry Mack's self-reported approach to improvisation. Rather than "just rapping with no plan", he practices various word, sound, rhythm, and meaning oriented drills, focusing on what he is bad at and trying to improve, akin to how intentional practicing of an instrument is different from just jamming or performing. It could also be that, because Harry Mack is already fluent in rhyme and an expert at improved rap, he is able to easily rhyme with any given sound pattern. This means that he may not have as strong of a functional bias as Ikaanic towards using specific phrases or sound patterns that he is comfortable with. Instead, Harry Mack can say whatever he likes, and be confident that he will find relevant rhymes with it. This same intuition can be applied to Ikaanic and his downward trending vowel sequence entropy. Since Ikaanic does not know how to fluently rhyme with any given phrase yet, he will likely have a bias to use words and phrases from sound patterns that he is familiar with - an example of exploiting local optima. It seems likely that the trajectory of developing this complex skill involves many alternating periods of exploiting (local optima) and exploring (global optima). Some learners may even get stuck in various local optima. For example, this could manifest as an individual only ever rhyming with $\tilde{5}0$ of the 225 2-syllable vowel sequences [15x15]), potentially limiting development in both lexical and phonological space.

Finally, the degree of semantic similarity across words for both artists decreased over the year, indicating that words, within a window of 20, are becoming less semantically related over time. On the surface, this may seem to be related to the increased word entropy overtime. As the variety of words increases, and their frequencies become more uniform, semantic similarity may tend to decrease due to a more uniform drawing from words across semantic space. However, entropy of word uni-grams and bi-grams do not increase for the novice, indicating that decreases in semantic similarity may not be directly associated with increases in word entropy here. Additionally, the vowel entropy trends are opposite for the novice (decreasing) and the expert (increasing), which also indicates that its interplay with semantic similarity may be minimal.

So why might semantic similarity be decreasing so reliably over time? The vowel entropy measure gives a sense of vowel sequence complexity across an entire verse, that is to say, across all rhyming and non-rhyme words. It may be that the relative about of rhyming words within a verse (or sample) is driving much of this reduction in semantic similarity - The higher the proportion of rhymed syllables (Rhymed Syllable Rate from section 6.1) the more semantically dissimilar its words are likely to become. Why? In order to match words that satisfy particular phonological constraints (e.g. alliteration, rhyme, assonance), they must be selected from specific 'bins' (e.g. all words that rhyme with 'day' or all words that start with '/fl/'). Not only do these 'bins' of possible matches dramatically limit which words can be chosen, but for the most part, their words are semantically unrelated. Due to the linguistic phenomenon of 'arbirariness of the sign' words with similar, or even the same sounds, may have entirely different meanings. As was discussed in Chapter 2, there are some known semantic relations across words with similar sounds. Sound symbolism, iconicity, and phoneaesthetics play a role in sound-meaning correspondences, leading to some reliable semantic relations between classes of sound (particular sound), and some meaning in the world (e.g. /gl/, /sn/). If one were only satisfying rhyming phonological constraints with words that share these sound symbolic or iconic relations, it might be predicted that increased proportions of rhymed (or matched) segments would also increase semantic similarity. This is because when trying to alliterate with /sn/ one would only ever pick words like snort, snout, sniffle, and sneeze, but never add sneakers to that list because sneakers are not nose-related. However, as I mentioned, the vast majority of words with phonological similarity are not so clearly semantically related, or even related at all. Indeed, due to the potential productiveness of maintaining arbitrariness of the sign, most words that sound similar will not have any clear semantic relationship.

For these reasons, when meeting the phonological constraints of language-based-music, the rhyming (or alliterated, etc.) words are likely to be less semantically related. Therefore, the higher the proportion of rhyming words in a sample, the lower the semantic similarity metric may be.

This hypothesis can be tested in future work in a number of ways. First of all, it is important to establish more baselines for this particular semantic similarity metric, especially in longitudinal spoken and written language data that are non-rhyming. Second, rhymes and other poetic devices within longitudinal lyrical data could be annotated (as in Section 6.1) and the connection between 'rhymed syllable rate' and semantic similarity can be directly tested.

Learning to rap improvisationally can be framed as a form of language skill acquisition, a productive and creative kind of expertise that may be learnable in any language. Learning technology has been shown to "generate new linguistic habits" in rappers [174], so both beginner and veteran rappers may even benefit from more formal pedagogical resources, effects which can be measured by approaches like those presented here.

6.2.5 Conclusion

In this section, I focused on creative language patterns (e.g. rhyme) and the holistic practice of improvisational rapping (development and skill). This study has explored the development of lexical, phonological, and semantic constraints in two rappers over a 1-year period of improvisation. I have shown that, although some dimensions of language display similar developmental trajectories across the novice and expert (lexical, semantic), other dimensions of language display opposite trends (phonological complexity). Further investigations may uncover more of the dynamics of phonological development and their interactions with related complex language components and skills (like improvised rap).

Chapter 7

Conclusion

The goal of this dissertation was to motivate and develop the study of language-based music. Our empirical domain has been the exploration of rhyme in multiple genres with a special focus on rap. We demonstrated that an instructive investigation of this space requires the development of both visualization and computational tools. Furthermore, since language-based music patterns can occur without being explicitly noticed, it is also important to use these tools to detect both overt and covert patterns. These resources and their results, in turn, make various data structures more accessible for both linguistic and cognitive investigations. To truly understand the phenomenon of language-based music, it is crucial to relate these patterns to the cognitive mechanisms that constrain their perception and production, and that facilitate their use.

In Chapter 1 I introduced language-based music and an interdisciplinary approach to exploring its forms and underlying cognitive mechanisms. I briefly reviewed the cultural and academic history of rhyme before outlining the two main problem spaces covered in this project, overt and covert sound patterns in language. Specifically, I motivated the study of complex patterns in language by examining a poem and example word sets, highlighting how large sound structures often go unnoticed. Following from this, I highlighted the notion of Rhyme Sets as particularly useful for the study of repeated (and multi-syllable) phonological structures. Finally, I framed the entire exploration in terms of phonological vocabulary, a hierarchical system of phonological elements used by an individual or language.

In Chapter 2 I presented a toolkit, Phonesse, for simple sound segment manipulation and visualization from text corpora as well as a related exploratory interface, Rhymable.com. In particular, I provided MIDI and grid-based representations of language sounds, and a psychologically aligned color-to-vowel mapping that can be used to intuitively reveal patterns of underlying and imperfect sound patterns. Here, I also introduced the majority of the data used in the contained studies including DCET, bar pong, The Fresh Coast Cypher, and longitudinal rap performances.

In Chapter 3 I addressed two primary questions:

- What cognitive mechanisms underlie the perception and production of rhyme forms?
- How have the elements of rhyme been explored using computational tools?

I first reviewed elements related to the perception and production of rhyme, including rhythm, similarity, rhyme acceptability, orthography, priming, words, sound symbolism, and improvisation. Then, I discussed the cognitive mechanisms that support rhyme, covering areas such as memory, language change, phonology and the phonological loop, literacy, expertise, and educational applications. Finally, I reviewed how the elements of rhyme have been explored computationally in the areas of stress, sequences, poetic analysis, rap analysis, and lyric generation.

In Chapter 4 I investigated overt poetic devices and addressed three primary questions:

- How many perfect rhymes (and other poetic devices) are possible?
- How prevalent is imperfect rhyming?
- How can multi-syllable imperfect rhyming (Rhyme Sets) be quantified?

I began by counting all possible perfect poetic device matches (within words in a dictionary) in order to document the range of patterns and pattern matches that are available in English. Then, I looked at 500 years of rhymes, finding large differences in rates of feature agreement over time. In addition, final coda agreement did not change over time, and was shown to be the most predictive rhyme feature across this period (in a context dominated by perfect rhyme). Finally, I expanded on Shaw's *large rhyme sets*, demonstrating how treating these data as complex systems, rather than isolated pairs, can help visualize and describe their sound structure and dynamics.

In Chapter 5 I investigated corpora that contain covert poetic devices and addressed two primary questions:

- How are phonemes distributed across language and musical genres?

- Are there cognitive or task effects related to language-based music?

Here, I began with a comparative case study which demonstrated clear differences between the sound sequences of many genres of music, some of which typify the well known phonological devices used in forms like sonnets (iambic pentameter) or battle rap (multi-syllable rhyme). Then, I expanded the scope of investigation, using the DCET data set and machine learning to identify sound usage differences across musical lyrics, poetry, speech, fiction, and non-fiction. In general, I found significant sound segment biases within poetry (e.g. consonant sounds) and musical lyrics (e.g. sonorant sounds and diphthongs) that seem to reflect task biases and enable reliable text classification (in some contexts).

In Chapter 6 I focused on the language of improvisational rap and addressed two primary questions:

- How is improvised rap different from written rap?
- How does improvised rap change over time?

Using the Fresh Coast Cypher rap session, I compared phonological and semantic differences between written and improvised rap verse. In general, the means of overt phonological structures in improvised rap were found to be smaller (word len, rhyme set height/width, etc.), while improvised variance was also found to be smaller than in written rap. Despite these differences (only some of which were significant), improvised rap lyrics still displayed surprisingly large phonological patterns (rhyme set avg width: 3.14 syllables, rhyme set avg height: 3.82 repeats). Finally, I investigated changes in lexical, phonological, and semantic rap constraints over a one-year time period. For the beginner and expert improvised rappers, lexical and semantic trends were found to be similar, whereas the changes in their phonological constraints over time were in opposition (beginner trends towards exploiting, expert trends towards exploring). This suggests that important changes in phonological constraints may occur in the trajectory of learning to rap improvisationally.

7.0.1 Limitations

As discussed in Chapter 2, though extracting sound elements from orthographic texts provides many conveniences and opportunities for mining existing corpora, there is a limit to the accuracy of these transcriptions. In many cases, phonemes extracted automatically from text should be additionally validated by human annotators. Phonemes can also be automatically extracted from audio files, presenting a more difficult transcription problem (classification) but providing a more faithful correlation to the phones uttered.

While lyrics and textual documents are plentiful in the wild, some phenomena, such as rhyme sets or improvised rap sessions, have only begun to be collected and are limited in quantity. Results from the current studies should be validated and extended on larger data sets and in languages other than English. Furthermore, in some cases here, I focus on vowels as a convenient proxy for the larger scale phonological patterns that exist in language-based music. However, this is a dramatic oversimplification that looses information about consonants, stress, and their associations. There are various phonological relationships between elements of syllables, and including these relations can improve the resolution and validity of the examined data.

7.0.2 Future Work

Beyond addressing the limitations mentioned above, future studies may expand on the form, theory, cognition, learning, or pedagogy of language-based music.

Collecting much larger sets of annotated data would allow for a deeper understanding of common phonologically driven poetic devices as well as the imperfect and multi-syllabic structures that have become so common.

Developing more tools and formal (or annotated) representations of rhyme (and related patterns) can enable better training in supervised models and more surface area for clustering in unsupervised models.

Unsupervised ML approaches may also be used to discover practical boundaries between repeated sound structures or rhyme sets, avoiding the common dependency on only human annotations of rhyme. Machine identified patterns can also be treated as rhyme sets (or other sets) themselves, and can be used to make transparent the underlying sound structure of machine identified rhyme. These data can also be leveraged for academic and practical (e.g. education, art) purposes.

Furthermore, studies that correlate brain activity with annotated devices may reveal important associations between the neural processes and linguistic structures of language-based music.

Bibliography

[1] J Martin. *Speech Sounds and Phonetic Transcription 3 ARPAbet IPA ARPAbet Symbol Symbol Word Transcription*. 2007.

[2] Judith Mcconnell, ; Lynne, and Gloria A Dye. The exceptional parent. *Boston*, 28(12), 1998.

[3] Gretchen McCulloch. How to remember the ipa vowel chart. *https://allthingslinguistic.com/post/67308552090/how-to-remember-the-ipa-vowel-chart*, 2014.

[4] E A Husaiyan. Vowel distinctive feature chart. *https://www.slideshare.net/EmanAlHsaiyan/phonology-phonological-features-of-english-vowels*, 2016.

[5] Piano Visualizer. `https://piano-visualizer.com/home`, 2022.

[6] Erin Banales, Saskia Kohnen, and Genevieve McArthur. Can verbal working memory training improve reading? *Cognitive Neuropsychology*, 32(3-4):104–132, 2015.

[7] Joydeep Bhattacharya and Hellmuth Petsche. Phase synchrony analysis of EEG during music perception reveals changes in functional connectivity due to musical expertise. *Signal Processing*, 85(11):2161–2177, November 2005.

[8] Adiel Mittmann, Aldo von Wangenheim, and Alckmar Luiz dos Santos. A multi-level visualization scheme for poetry. In *2016 20th International Conference Information Visualisation (IV)*. IEEE, July 2016.

[9] Sonderegger. M. Applications of graph theory to an english rhyming corpus. *Computer Speech & Language.*, 25(32):655–678, 2011.

[10] Norbert Marwan, M Carmen Romano, Marco Thiel, and Jürgen Kurths. Recurrence plots for the analysis of complex systems. *Physics reports*, 438(5-6):237–329, 2007.

[11] H Hirjee and D G Brown. Rhyme analyzer: An analysis tool for rap lyrics. In *Proceedings of the 11th International Society for Music Information Retrieval Conference*. 2010.

[12] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problemy Peredachi Informatsii*, 1(1):3–11, 1965.

[13] Kazadi Wa Mukuna. Function of musical instruments in surrogate languages in áfrica: a clarification. *África*, (10):3–8, 1987.

[14] Laura McPherson. Musical surrogate languages in the documentation of complex tone: the case of the sambla balafon. *Proceedings of Tonal Aspects of Language 2018*, 2018.

[15] Laura McPherson. Musical adaptation as phonological evidence: Case studies from textsetting, rhyme, and musical surrogates. *Language and Linguistics Compass*, 13(12):e12359, 2019.

[16] Monroe C Beardsley. Aesthetics. *Harcourt, Brace*, page 12, 1950.

[17] David Crystal. Phonaesthetically speaking. *Engl. today*, 11(2):8–12, April 1995.

[18] Simon Jarvis. For a poetics of verse. *PMLA*, 125(4):931–935, 2010.

[19] H Samy Alim. On some serious next millennium rap ishhh: Pharoahe monch, hip hop poetics, and the internal rhymes of internal affairs. *Journal of English Linguistics*, 31(1):60–84, 2003.

[20] Reuven Tsur. Rhyme and cognitive poetics. *Poetics Today*, pages 55–87, 1996.

[21] Christian Obermeier, Winfried Menninghaus, Martin von Koppenfels, Tim Raettig, Maren Schmidt-Kassow, Sascha Otterbein, and Sonja A Kotz. Aesthetic and emotional effects of meter and rhyme in poetry. *Front. Psychol.*, 4:10, January 2013.

[22] LV Sheng-nan. Study of the co-rhyme of geng & zhen in the book of songs. *Journal of Tianzhong*, 2009.

[23] Robert Walser. Rhythm, rhyme, and rhetoric in the music of public enemy. *Ethnomusicology*, 39(2):193–217, 1995.

[24] Adam Krims and A Krims. *Rap music and the poetics of identity*, volume 5. Cambridge University Press, 2000.

[25] Kyle Adams. On the metrical techniques of flow in rap music. *Music Theory Online*, 15(5), 2009.

[26] Brent Hayes Edwards. Louis armstrong and the syntax of scat. *Critical Inquiry*, 28(3):618–649, 2002.

[27] Novia Permata Silviandari and M Suryadi. Invictus poem by william ernest and its contribution to the social environment during pandemic: Study of sociology literature. In *E3S Web of Conferences*, volume 317, page 03010. EDP Sciences, 2021.

[28] David C Rubin. *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes.* Oxford University Press, Cary, NC, January 1995.

[29] Linda F Wharton-Boyd. The significance of black american children's singing games in an educational setting. *The Journal of Negro Education*, 52(1):46–56, 1983.

[30] Carl J Dunst, Diana Meter, and Deborah W Hamby. Relationship between young children's nursery rhyme experiences and knowledge and phonological and print-related abilities. *Center for Early Literacy Learning*, 4(1):1–12, 2011.

[31] Amy J Shollenbarger, Gregory C Robinson, Valentina Taran, and Seo-eun Choi. How african american english-speaking first graders segment and rhyme words and nonwords with final consonant clusters. *Language, Speech, and Hearing Services in Schools*, 48(4):273–285, 2017.

[32] Marc Lamont Hill. *Beats, rhymes, and classroom life: Hip-hop pedagogy and the politics of identity.* Teachers College Press, 2009.

[33] T Tomás Alvarez III. Beats, rhymes, and life. *Therapeutic uses of rap and hip-hop*, page 117, 2012.

[34] Steven Pinker. Language as an adaptation to the cognitive niche. *Studies in the Evolution of Language*, 3:16–37, 2003.

[35] Samuel A Mehr, Max M Krasnow, Gregory A Bryant, and Edward H Hagen. Origins of music in credible signaling. *Behav. Brain Sci.*, 44(e60):e60, August 2020.

[36] Patrick E Savage, Psyche Loui, Bronwyn Tarr, Adena Schachner, Luke Glowacki, Steven Mithen, and W Tecumseh Fitch. Music as a coevolved system for social bonding. *Behav. Brain Sci.*, 44(e59):e59, August 2020.

[37] John L Locke. Evolutionary developmental linguistics: Naturalization of the faculty of language. *Lang. Sci.*, 31(1):33–59, January 2009.

[38] Varuṇ deCastro Arrazola and Simon Kirby. The emergence of verse templates through iterated learning. *J. Lang. Evol.*, 4(1):28–43, January 2019.

[39] W.E. Rickert. *Rhyme terms.* Style, 1978.

[40] Viktor Maksimovic Žirmunskij. *Introduction to metrics: the theory of verse*, volume 58. Walter de Gruyter GmbH & Co KG, 2016.

[41] Morris Halle. Žirmunskij's theory of verse: A review article. *The Slavic and East European Journal*, 12(2):213–218, 1968.

[42] Roman Jakobson. Linguistics and poetics. In *Style in language*, pages 350–377. MIT Press, MA, 1960.

[43] Dean S Worth. Roman jakobson and the study of rhyme. *Roman Jakobson: echoes of his scholarship*, pages 515–33, 1977.

[44] Arnold M Zwicky. Well, this rock and roll has got to stop. junior's head is hard as a rock. In *Papers from the... Regional Meeting. Chicago Ling. Soc. Chicago, Ill*, number 12, pages 676–697, 1976.

[45] Jonah Katz. Hip-hop rhymes reiterate phonological typology. *Lingua*, 160:54–73, 2015.

[46] A Holtman and E Buckley. *A generative approach to rhyme: An Optimality approach (Doctoral dissertation)*. Utrecht, 1996.

[47] Roman Jakobson. Linguistics and poetics. In *Style in language*, pages 350–377. MIT Press, MA, 1960.

[48] Donca Steriade. Knowledge of perceptual similarity and its phonological uses: evidence from half-rhymes. In *Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona: Universitat Autònoma de Barcelona*, pages 363–366, 2003.

[49] A Kaplan, J Woodmansee, and U S Colloquium. *Imperfect Rhymes as a Measure of Phonological Similarity*. 2018.

[50] J Van Der Schelde. *Phonological and phonetic similarity as underlying principles of imperfect rhyme*. 2020.

[51] Christine A Knoop, Stefan Blohm, Maria Kraxenberger, and Winfried Menninghaus. How perfect are imperfect rhymes? effects of phonological similarity and verse context on rhyme perception. *Psychol. Aesthet. Creat. Arts*, 15(3):560–572, August 2021.

[52] Rajeev Rajan, Aiswarya Vinod Kumar, and Ben P Babu. Poetic meter classification using i-vector-MTF fusion. In *Interspeech 2020*, ISCA, October 2020. ISCA.

[53] Rodolfo Delmonte. A computational approach to poetic structure, rhythm and rhyme. *A computational approach to poetic structure, rhythm and rhyme*, pages 144–150, 2014.

[54] David M Kaplan and David M Blei. A computational approach to style in american poetry. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*. IEEE, October 2007.

[55] 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*, EMNLP '10, pages 524–533, USA, October 2010. Association for Computational Linguistics.

[56] Josh Freeman. Vowel transitions in the sonnets of Shakespeare: an information theoretic analysis. page 26, 2018.

[57] Warren Olivo. Phat lines. *Writ. Lang. Lit.*, 4(1):67–85, March 2001.

[58] M Daniels. The largest vocabulary in hip hop., 2014.

[59] 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: short papers-Volume 2*, pages 77–82. Association for Computational Linguistics, 2011.

[60] Karteek Addanki and Dekai Wu. Unsupervised rhyme scheme identification in hip hop lyrics using hidden markov models. In *International conference on statistical language and speech processing*, pages 39–50. Springer, 2013.

[61] Hussein Hirjee and Daniel G Brown. Automatic detection of internal and imperfect rhymes in rap lyrics. In *ISMIR*, pages 711–716, 2009.

[62] H G Oliveira. A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 11–20. 2017.

[63] S Colton, J Goodwin, and T Veale. Full-FACE poetry generation. In *ICCC*, pages 95–102. 2012.

[64] M L Maher, T Veale, and R Saunders. Proceedings of the fourth international conference on computational creativity. 2013.

[65] Eric Malmi, Pyry Takala, Hannu Toivonen, Tapani Raiko, and Aristides Gionis. DopeLearning: A computational approach to rap lyrics generation. May 2015.

[66] Udo Schlegel, Eren Cakmak, Juri Buchmüller, and Daniel A Keim. G-Rap: interactive text synthesis using recurrent neural network suggestions. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning : ESANN 2018*, 2018.

[67] Marjan Ghazvininejad, Xing Shi, Jay Priyadarshi, and Kevin Knight. Hafez: an interactive poetry generation system. In *Proceedings of ACL 2017, System Demonstrations*, Stroudsburg, PA, USA, 2017. Association for Computational Linguistics.

[68] Nathaniel Condit-Schultz. MCFlow: A digital corpus of rap transcriptions. *Empir. Musicol. Rev.*, 11(2):124, January 2017.

[69] Steven Pinker. How the mind works, first paperback edition edn, 1999.

[70] Cody Moser, Jordan Ackerman, Alex Dayer, Shannon Proksch, and Paul E Smaldino. Why don't cockatoos have war songs? 2021.

[71] Donald E Brown. Human universals, 1991.

[72] Samuel A Mehr, Manvir Singh, Dean Knox, Daniel M Ketter, Daniel Pickens-Jones, S Atwood, Christopher Lucas, Nori Jacoby, Alena A Egner, Erin J Hopkins, Rhea M Howard, Joshua K Hartshorne, Mariela V Jennings, Jan Simson, Constance M Bainbridge, Steven Pinker, Timothy J O'Donnell, Max M Krasnow, and Luke Glowacki. Universality and diversity in human song. *Science*, 366(6468):eaax0868, November 2019.

[73] Daniel S Adler. Archaeology: The earliest musical tradition. *Nature*, 460(7256):695–696, August 2009.

[74] Ruggero D'Anastasio, Stephen Wroe, Claudio Tuniz, Lucia Mancini, Deneb T Cesana, Diego Dreossi, Mayoorendra Ravichandiran, Marie Attard, William C H Parr, Anne Agur, and Luigi Capasso. Micro-biomechanics of the kebara 2 hyoid and its implications for speech in neanderthals. *PLoS One*, 8(12):e82261, December 2013.

[75] Aniruddh D Patel. Musical rhythm, linguistic rhythm, and human evolution. *Music Percept.*, 24(1):99–104, September 2006.

[76] Arthur Waley Estate and Arthur Waley. *The book of songs*. Routledge, London, England, November 2012.

[77] S Li and Y Mai. Proofs that book of songs is rhyme. *Studies of the Chinese Language*, (4), 2008.

[78] Éva M Jeremiás and David Neil MacKenzie. The grammatical tradition in persian: Shams-i fakhrī's rhyme science in the fourteenth century. *Iran*, 48(1):153–162, 2010.

[79] O Jemie. *Yo Mama!: New Raps, Toasts, Dozens, Jokes, and Children's Rhymes from Urban Black America*. Temple University Press, 2003.

[80] Elijah Wald. *The Dozens: a history of rap's mama*. Oxford University Press, 2012.

[81] Jack Beckwith. The Evolution of Music Genre Popularity. `https://thedataface.com/2016/09/culture/genre-lifecycles`, 2016.

[82] VerseTracker: The battle rap culture index. `http://www.versetracker.com`. Accessed: 2022-4-18.

[83] Harry Mack. Harry Mack - YouTube Channel. `https://www.youtube.com/c/HarryMack`, 2022.

[84] Shigeto Kawahara. Half rhymes in japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics*, 16(2):113–144, 2007.

[85] Paul Bauschatz. Rhyme and the structure of english consonants. *English Language & Linguistics*, 7(1):29–56, 2003.

[86] Kristin Hanson. Formal variation in the rhymes of robert pinsky's the inferno of dante. *Language and Literature*, 12(4):309–337, 2003.

[87] EM Edghill et al. *The categories*. Good Press, 2019.

[88] That is Embarrassing Illusion. `https://www.youtube.com/watch?v=8FXQ38-ZQK0`, 2021.

[89] Eleanor Rosch. Wittgenstein and categorization research in cognitive psychology. In *Meaning and the growth of understanding*, pages 151–166. Springer, 1987.

[90] Jaakko Hintikka. Semantical games and aristotelian categories. In *The Game of Language*, pages 201–229. Springer, 1983.

[91] Eleanor Rosch and Carolyn B Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4):573–605, 1975.

[92] J Thomas Shaw. Large rhyme sets and puškin's poetry. *Slavic and East European Journal*, pages 231–251, 1974.

[93] James Milton and Nicola Hopkins. Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review*, 63(1):127–147, 2006.

[94] Oxford English Dictionary. Oxford english dictionary. *Simpson, Ja & Weiner, Esc*, 3, 1989.

[95] Adam Albright and Bruce Hayes. Learning and learnability in phonology. In *The Handbook of Phonological Theory*, pages 661–690. Wiley-Blackwell, Oxford, UK, November 2011.

[96] Adrian P Simpson. Phonetic differences between male and female speech. *Lang. Linguist. Compass*, 3(2):621–640, March 2009.

[97] Cyril R Pernet and Pascal Belin. The role of pitch and timbre in voice gender categorization. *Front. Psychol.*, 3:23, February 2012.

[98] Edward Sapir. Sound patterns in language. *Language*, 1(2):37–51, 1925.

[99] Edward Sapir. Language in: Encyclopaedia of the social sciences. *New York*, 9:155–169, 1933.

[100] Morris Halle. The strategy of phonemics. *Word*, 10(2-3):197–209, 1954.

[101] W Freeman Twaddell. On defining the phoneme. *Language*, 11(1):5–62, 1935.

[102] RT Butlin. An examination of the validity of certain current phonetic ideas. *Transactions of the Philological Society*, 36(1):137–137, 1937.

[103] Michael Ramscar, Robert Port, E Dabrowska, and D Divjak. Categorization (without categories). *Cognitive linguistics–Foundations of language*, pages 87–114, 2019.

[104] J Morais, P Bertelson, L Cary, and J Alegria. Literacy training and speech segmentation. *Cognition*, 24(1-2):45–64, November 1986.

[105] Charles Read, Zhang Yun-Fei, Nie Hong-Yin, and Ding Bao-Qing. The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, 24(1-2):31–44, 1986.

[106] J Burridge and Bert Vaux. Brownian dynamics for the vowel sounds of human language. *Physical Review Research*, 2(1):013274, 2020.

[107] J Ben Falandays and Paul E Smaldino. The emergence of cultural attractors: An agent-based model of collective cognitive alignment. *Proceedings of the cognitive science society 2021*, 2021.

[108] George A Miller. Ambiguous words. *Impacts Magazine*, 2001.

[109] Jongseok Park, Kyubyong Kim. g2pe. `https://github.com/Kyubyong/g2p`, 2019.

[110] Daniel J Levitin. *This is your brain on music: The science of a human obsession*. Penguin, 2006.

[111] Christine Cuskley, Mark Dingemanse, Simon Kirby, and Tessa M Van Leeuwen. Cross-modal associations and synesthesia: Categorical perception and structure in vowel–color mappings in a large online sample. *Behavior Research Methods*, 51(4):1651–1675, 2019.

[112] Régine Kolinsky, Pascale Lidji, Isabelle Peretz, Mireille Besson, and José Morais. Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition*, 112(1):1–20, 2009.

[113] Kirsten Read, Megan Macauley, and Erin Furay. The seuss boost: Rhyme helps children retain words from shared storybook reading. *First Lang.*, 34(4):354–371, August 2014.

[114] Michael D Mauk and Dean V Buonomano. The neural basis of temporal processing. *Annu. Rev. Neurosci.*, 27(1):307–340, July 2004.

[115] Shane T Mueller, Travis L Seymour, David E Kieras, and David E Meyer. Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *J. Exp. Psychol. Learn. Mem. Cogn.*, 29(6):1353–1380, November 2003.

[116] Alan D Baddeley. Is working memory still working? 1copyright © 2001 by the american psychological association. reprinted with permission from the original publication: . "is working memory still working?" american psychologist, 56, 849-864. this re-print publication is arranged in recognition of the conferral on dr. baddeley of the aristotle prize at the VIIth european congress of psychology, london, in july 2001, for his outstanding research on human working memory. the original publication in the american psychologist related to prof. baddeley's receipt of the distinguished scientific contribution award of the american psychological association in 2001. *Eur. Psychol.*, 7(2):85–97, June 2002.

[117] Alan Baddeley. The episodic buffer: a new component of working memory? *Trends Cogn. Sci.*, 4(11):417–423, November 2000.

[118] Alan Baddeley. Working memory: theories, models, and controversies. *Annu. Rev. Psychol.*, 63(1):1–29, 2012.

[119] Karl Christoph Klauer and Zengmei Zhao. Double dissociations in visual and spatial short-term memory. *J. Exp. Psychol. Gen.*, 133(3):355–381, September 2004.

[120] Edward E Smith and John Jonides. Working memory: A view from neuroimaging. *Cogn. Psychol.*, 33(1):5–42, June 1997.

[121] M. Denis, R. Logie, C. Cornoldo, and M. Vega. *language and visuo-spatial thinking (Vol*, volume 1. Psychology Press, 2012.

[122] John L Logke and Virginia L Locke. Recall of phonetically and semantically similar words by 3-year-old children. *Psychon. Sci.*, 24(4):189–190, April 1971.

[123] R Conrad and A J Hull. Information, acoustic confusion and memory span. *Br. J. Psychol.*, 55(4):429–432, November 1964.

[124] A D Baddeley. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Q. J. Exp. Psychol.*, 18(4):362–365, November 1966.

[125] Alan D Baddeley, Neil Thomson, and Mary Buchanan. Word length and the structure of short-term memory. *J. Verbal Learning Verbal Behav.*, 14(6):575–589, December 1975.

[126] Stephen A Madigan and Linda McCabe. Perfect recall and total forgetting: A problem for models of short-term memory. *J. Verbal Learning Verbal Behav.*, 10(1):101–106, February 1971.

[127] R G Crowder, M L Serafine, and B Repp. Physical interaction and association by contiguity in memory for the words and melodies of songs. *Mem. Cognit.*, 18(5):469–476, September 1990.

[128] Mitchell G Newberry, Christopher A Ahern, Robin Clark, and Joshua B Plotkin. Detecting evolutionary forces in language change. *Nature*, 551(7679):223–226, November 2017.

[129] Florencia Reali and Thomas L Griffiths. Words as alleles: connecting language evolution with bayesian learners to models of genetic drift. *Proc. Biol. Sci.*, 277(1680):429–436, February 2010.

[130] Christina Pawlowitsch, Panayotis Mertikopoulos, and Nikolaus Ritt. Neutral stability, drift, and the diversification of languages. *J. Theor. Biol.*, 287:1–12, October 2011.

[131] Christian Bentz, Dan Dediu, Annemarie Verkerk, and Gerhard Jäger. The evolution of language families is shaped by the environment beyond neutral drift. *Nat. Hum. Behav.*, 2(11):816–821, November 2018.

[132] F C Bartlett and Cyril Burt. Remembering: A study in experimental and social psychology. *Br. J. Educ. Psychol.*, 3(2):187–192, June 1933.

[133] Kenny Smith, Simon Kirby, and Henry Brighton. Iterated learning: a framework for the emergence of language. *Artif. Life*, 9(4):371–386, 2003.

[134] Stephan Lewandowsky, Thomas L Griffiths, and Michael L Kalish. The wisdom of individuals: Exploring people's knowledge about everyday events using iterated learning. *Cogn. Sci.*, 33(6):969–998, August 2009.

[135] H Cornish, K Smith, and S Kirby. Systems from sequences: An iterated learning account of the emergence of systematic structure in a non-linguistic task. *Proceedings of the annual meeting of the cognitive science society*, 35, 2013.

[136] E.A. Esper. A technique for the experimental investigation of associative interference in artificial linguistic, 1925. material. Language Monographs No. 1.

[137] Limor Raviv and Inbal Arnon. Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition*, 181:160–173, December 2018.

[138] Alain Content, Ruth K Kearns, and Uli H Frauenfelder. Boundaries versus onsets in syllabic segmentation. *J. Mem. Lang.*, 45(2):177–199, August 2001.

[139] A H van der Lugt. The use of sequential probabilities in the segmentation of speech. *Percept. Psychophys.*, 63(5):811–823, July 2001.

[140] M R Brent and T A Cartwright. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61(1-2):93–125, October 1996.

[141] M R Brent. Speech segmentation and word discovery: a computational perspective. *Trends Cogn. Sci.*, 3(8):294–301, August 1999.

[142] Jenny R. Saffran, Elissa L. Newport, and Richard N. Aslin. Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, 35(4):606–621, August 1996.

[143] Laura J Batterink. Rapid statistical learning supporting word extraction from continuous speech. *Psychol. Sci.*, 28(7):921–928, July 2017.

[144] Sven L Mattys, Laurence White, and James F Melhorn. Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exp. Psychol. Gen.*, 134(4):477–500, November 2005.

[145] Lisa D Sanders, Helen J Neville, and Marty G Woldorff. Speech segmentation by native and non-native speakers: the use of lexical, syntactic, and stress-pattern cues. *J. Speech Lang. Hear. Res.*, 45(3):519–530, June 2002.

[146] Juan M Toro, Scott Sinnett, and Salvador Soto-Faraco. Speech segmentation by statistical learning depends on attention. *Cognition*, 97(2):B25–34, September 2005.

[147] James L Morgan. A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, 35(5):666–688, 1996.

[148] Derek Houston, Lynn Santelmann, and Peter Jusczyk. English-learning infants' segmentation of trisyllabic words from fluent speech. *Lang. Cogn. Process.*, 19(1):97–136, February 2004.

[149] Anne Cutler and Sally Butterfield. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *J. Mem. Lang.*, 31(2):218–236, April 1992.

[150] Clément François, Julie Chobert, Mireille Besson, and Daniele Schön. Music training for the development of speech segmentation. *Cereb. Cortex*, 23(9):2038–2043, September 2013.

[151] Bruna Pelucchi, Jessica F Hay, and Jenny R Saffran. Statistical learning in a natural language by 8-month-old infants. *Child Dev.*, 80(3):674–685, May 2009.

[152] Jenny Saffran. Sounds and meanings working together: Word learning as a collaborative effort. *Lang. Learn.*, 64(Suppl 2):106–120, September 2014.

[153] J L Metsala. An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Mem. Cognit.*, 25(1):47–56, January 1997.

[154] U Goswami and P Bryant. *Essays in developmental psychology series.Phonological skills and learning to read.* Lawrence Erlbaum Associates, Inc, 1990.

[155] Usha Goswami. Causal connections in beginning reading: the importance of rhyme. *J. Res. Read.*, 22(3):217–240, October 1999.

[156] Jeffrey R Binder, Sara J Swanson, Thomas A Hammeke, and David S Sabsevitz. A comparison of five fMRI protocols for mapping speech comprehension systems. *Epilepsia*, 49(12):1980–1997, December 2008.

[157] John M Henderson, Wonil Choi, Steven G Luke, and Rutvik H Desai. Neural correlates of fixation duration in natural reading: Evidence from fixation-related fMRI. *Neuroimage*, 119:390–397, October 2015.

[158] Tami Katzir and Juliana Pare-Blagoev. Applying cognitive neuroscience research to education: The case of literacy. *Educational Psychologist*, 41(1):53–74, 2006.

[159] Catherine Moritz, Sasha Yampolsky, Georgios Papadelis, Jennifer Thomson, and Maryanne Wolf. Links between early rhythm skills, musical training, and phonological awareness. *Read. Writ.*, 26(5):739–769, May 2013.

[160] Ulla Richardson, Jennifer M Thomson, Sophie K Scott, and Usha Goswami. Auditory processing skills and phonological representation in dyslexic children. *Dyslexia*, 10(3):215–233, August 2004.

[161] Jennifer M Thomson and Usha Goswami. Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *J. Physiol. Paris*, 102(1-3):120–129, January 2008.

[162] L R Squire and S M Zola. Structure and function of declarative and nondeclarative memory systems. *Proc. Natl. Acad. Sci. U. S. A.*, 93(24):13515–13522, November 1996.

[163] Laura J Batterink, Paul J Reber, Helen J Neville, and Ken A Paller. Implicit and explicit contributions to statistical learning. *J. Mem. Lang.*, 83:62–78, August 2015.

[164] Laura J Batterink, Paul J Reber, and Ken A Paller. Functional differences between statistical learning with and without explicit training. *Learn. Mem.*, 22(11):544–556, November 2015.

[165] Karen M Ludke, Fernanda Ferreira, and Katie Overy. Singing can facilitate foreign language learning. *Mem. Cognit.*, 42(1):41–52, January 2014.

[166] Annie M Paul. Need to remember something? make it rhyme. *Psychology*, 2013.

[167] B Rando, E A O'connor, K Steuerwalt, and M Bloom. Preschool and kindergarten: Rap and young children: Encouraging emergent literacy. *YC Young Children*, 69(3):28–33, 2014.

[168] N Cahill and M Pratt. *More literacy skills through rhyme and rhythm : for the multi-ability classroom*. Oxford University Press, Melbourne, 1998.

[169] R Fink. Rap and technology teach the art of argument. *Learning Disabilities: A Contemporary Journal*, 15(1):39–53, 2017.

[170] Edward Anderson. Positive use of rap music in the classroom. 1993.

[171] M A Jeremiah. Rap lyrics: Instruments for language arts instruction. *Western journal of black studies*, 16:98–102, 1992.

[172] B Segal. Teaching english as a second language through rap music: A curriculum for secondary school students. 55, 2014.

[173] Cristina Aliagas Marín. Rap music in minority languages in secondary education: A case study of catalan rap. *Int. J. Soc. Lang.*, 2017(248), September 2017.

[174] Evelyn Ch'ien. Creative technology and rap. *World Englishes*, 30(1):60–75, March 2011.

[175] Jordan Ackerman. WordSurge - The Smart Dictionary. `https://www.wordsurge.com`, 2013.

[176] Annerose Engel and Peter E Keller. The perception of musical spontaneity in improvised and imitated jazz performances. *Front. Psychol.*, 2:83, May 2011.

[177] Masaru Sasaki, John Iversen, and Daniel E Callan. Music improvisation is characterized by increase EEG spectral power in prefrontal and perceptual motor cortical sources and can be reliably classified from non-improvisatory performance. *Front. Hum. Neurosci.*, 13:435, December 2019.

[178] R. Anderson, S.J. Hanrahan, and C.J. Mallett. Investigating the optimal psychological state for peak performance in australian elite athletes. journal of applied sport psychology 26, 2014.

[179] Alex Dayer and Carolyn Dicey Jennings. Correction to: Attention in skilled behavior: An argument for pluralism. *Rev. Philos. Psychol.*, 12(3):639–639, September 2021.

[180] Siyuan Liu, Ho Ming Chow, Yisheng Xu, Michael G Erkkinen, Katherine E Swett, Michael W Eagle, Daniel A Rizik-Baer, and Allen R Braun. Neural correlates of lyrical improvisation: an FMRI study of freestyle rap. *Sci. Rep.*, 2(1):834, November 2012.

[181] Roger E Beaty. The neuroscience of musical improvisation. *Neurosci. Biobehav. Rev.*, 51:108–117, April 2015.

[182] P Cheyne and A Hamilton. *The Philosophy of Rhythm: Aesthetics*. Oxford University Press, Music, Poetics; USA, 2019.

[183] Jonna K Vuoskoski and Dee Reynolds. Music, rowing, and the aesthetics of rhythm. *The Senses and Society*, 14(1):1–14, 2019.

[184] Jan Stupacher, Michael J Hove, and Petr Janata. Audio features underlying perceived groove and sensorimotor synchronization in music. *Music Percept.*, 33(5):571–589, June 2016.

[185] Maria AG Witek, Eric F Clarke, Mikkel Wallentin, Morten L Kringelbach, and Peter Vuust. Syncopation, body-movement and pleasure in groove music. *PloS one*, 9(4):e94446, 2014.

[186] Guilherme Schmidt Câmara and Anne Danielsen. Groove. 2018.

[187] Aniruddh D Patel. Rhythm in language and music: parallels and differences. *Ann. N. Y. Acad. Sci.*, 999:140–143, November 2003.

[188] Aniruddh D Patel and Joseph R Daniele. An empirical comparison of rhythm in language and music. *Cognition*, 87(1):B35–45, February 2003.

[189] Jessica A Grahn and James B Rowe. Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *Journal of Neuroscience*, 29(23):7540–7548, 2009.

[190] Jessica A Grahn and Matthew Brett. Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.*, 19(5):893–906, May 2007.

[191] Joyce L Chen, Virginia B Penhune, and Robert J Zatorre. Listening to musical rhythms recruits motor regions of the brain. *Cereb. Cortex*, 18(12):2844–2854, December 2008.

[192] David Huron. *Sweet anticipation*. The MIT Press, 2006.

[193] James L Morgan. A rhythmic bias in preverbal speech segmentation. *J. Mem. Lang.*, 35(5):666–688, October 1996.

[194] Anne Cutler and Dennis Norris. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human perception and performance*, 14(1):113, 1988.

[195] Marina Nespor and Irene Vogel. On clashes and lapses. *Phonology*, 6(1):69–116, 1989.

[196] Amalia Arvaniti. Acoustic features of greek rhythmic structure. *J. Phon.*, 22(3):239–268, July 1994.

[197] P J Donegan and D Stampe. *Rhythm and the holistic organization of language structure*. 1983.

[198] K.J. Kohler. Rhythm in speech and language. phonetica, 66(1-2, 2009.

[199] S S Johnsen. *Rhyme acceptability determined by perceived similarity.* 2011.

[200] Mielke. J. A phonetically based metric of sound similarity. *Lingua.*, 122, 2012.

[201] Gallagher. G and Graff. P. The role of similarity in phonology. *Lingua.*, 122, 2012.

[202] Keith Johnson. Vowel perception asymmetry in auditory and phonemic listening. *UC Berkeley Phonology Lab Annual Reports*, 11, 2015.

[203] Linda Polka and Ocke-Schwen Bohn. Asymmetries in vowel perception. *Speech Commun.*, 41(1):221–231, August 2003.

[204] Linda Polka and Ocke-Schwen Bohn. Natural referent vowel (NRV) framework: An emerging view of early phonetic development. *J. Phon.*, 39(4):467–478, October 2011.

[205] S T Mueller. *PSIMETRICA: Tools and techniques for measuring phonological similarity.* 2006.

[206] Viorica Marian, Henrike K Blumenfeld, and Olga V Boukrina. Sensitivity to phonological similarity within and across languages. *J. Psycholinguist. Res.*, 37(3):141–170, May 2008.

[207] K. Tamási, C. McKean, A. Gafos, and B. Höhle. Children's gradient sensitivity to phonological mismatch: considering the dynamics of looking behavior and pupil dilation. *Journal of Child Language*, 46(1):1–23, 2019.

[208] S Seyfarth, M Garellek, G Gillingham, F Ackerman, and R Malouf. Acoustic differences in morphologically-distinct homophones. language. *Cognition and Neuroscience*, 33(1):32–49, 2018.

[209] Natasha Warner, Allard Jongman, Joan Sereno, and Rachèl Kemps. Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from dutch. *J. Phon.*, 32(2):251–276, April 2004.

[210] Diana Deutsch, Rachael Lapidis, and Trevor Henthorn. The speech-to-song illusion. *J. Acoust. Soc. Am*, 124(2471):10–1121, 2008.

[211] Simone Falk, Tamara Rathcke, and Simone Dalla Bella. When speech sounds like music. *J. Exp. Psychol. Hum. Percept. Perform.*, 40(4):1491–1506, August 2014.

[212] Kankamol Jaisin, Rapeepong Suphanchaimat, Mauricio A Figueroa Candia, and Jason D Warren. The speech-to-song illusion is reduced in speakers of tonal (vs. non-tonal) languages. *Frontiers in Psychology*, 7:662, 2016.

[213] J Zhang and C Liu. The neural mechanism of rhyme processing. *Lab of Cognitive Neuroscience and Department of Psychology, Nanjing Normal University, Nanjing, 210097, China*, 2013.

[214] Donna Coch, Tory Hart, and Priya Mitra. Three kinds of rhymes: An ERP study. *Brain Lang.*, 104(3):230–243, March 2008.

[215] Dénes Szűcs and Fruzsina Soltész. Functional definition of the n450 event-related brain potential marker of conflict processing: a numerical stroop study. *BMC neuroscience*, 13(1):1–14, 2012.

[216] Donna Coch, Elyse George, and Natalie Berger. The case of letter rhyming: an ERP study. *Psychophysiology*, 45(6):949–956, November 2008.

[217] Cláudia Cardoso-Martins. Rhyme perception: Global or analytical? *J. Exp. Child Psychol.*, 57(1):26–41, February 1994.

[218] Dennis Norris, James M McQueen, and Anne Cutler. Bias effects in facilitatory phonological priming. *Mem. Cognit.*, 30(3):399–411, April 2002.

[219] Annukka K Lindell and Jarrad A G Lum. Priming vs. rhyming: orthographic and phonological representations in the left and right hemispheres. *Brain Cogn.*, 68(2):193–203, November 2008.

[220] A Khateb, A J Pegna, C M Michel, M C Custodi, T Landis, and J M Annoni. Semantic category and rhyming processing in the left and right cerebral hemisphere. *Laterality*, 5(1):35–53, January 2000.

[221] Rhona S Johnston and Erica A McDermott. Suppression effects in rhyme judgement tasks. *Q. J. Exp. Psychol. A*, 38(1):111–124, February 1986.

[222] Pekka Niemi, Marja Vauras, and Johan Wright. Semantic activation due to synonym, antonym, and rhyme production. *Scand. J. Psychol.*, 21(1):103–107, September 1980.

[223] A Rouibah, G Tiberghien, and S J Lupker. Phonological and semantic priming: evidence for task-independent effects. *Mem. Cognit.*, 27(3):422–437, May 1999.

[224] David N Rapp and Arthur G Samuel. A reason to rhyme: Phonological and semantic influences on lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.*, 28(3):564–571, 2002.

[225] Laetitia Perre, Katherine Midgley, and Johannes C Ziegler. When beef primes reef more than leaf: orthographic information affects phonological priming in spoken word recognition. *Psychophysiology*, 46(4):739–746, July 2009.

[226] Hezekiah Akiva Bacovcin, Amy Goodwin Davies, Robert J Wilder, and David Embick. Auditory morphological processing: Evidence from phonological priming. *Cognition*, 164:102–106, July 2017.

[227] Bob McMurray, Meghan A Clayards, Michael K Tanenhaus, and Richard N Aslin. Tracking the time course of phonetic cue integration during spoken word recognition. *Psychon. Bull. Rev.*, 15(6):1064–1071, December 2008.

[228] W J Levelt. Spoken word production: a theory of lexical access. *Proc. Natl. Acad. Sci. U. S. A.*, 98(23):13464–13471, November 2001.

[229] J T Lurito, D A Kareken, M J Lowe, S A Chen, and V P Mathews. Comparison of rhyming and word generation with FMRI. *Neuroimage*, 7(4):S139, May 1998.

[230] A. Wedel, N. Nelson, and R. Sharp. The phonetic specificity of contrastive hyperarticulation in natural speech. journal of memory and language, 100, 2018.

[231] E. Buz, T.F. Jaeger, and M.K. Tanenhaus. Contextual confusability leads to targeted hyperarticulation. *Proceedings of the Annual Meeting of the Cognitive Science Society (Vol*, 36(36)), 2014.

[232] E. Buz, M.K. Tanenhaus, and T.F. Jaeger. Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. journal of memory and language, 89, 2016.

[233] B. Lindblom. Explaining phonetic variation: A sketch of the hh theory. *Speech production and speech*, page 403–439, 1990.

[234] N.R. Nelson and A. Wedel. The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational english. journal of phonetics, 64, 2017.

[235] R. Wright. Lexical competition and reduction in speech: A preliminary report. research on spoken language processing progress report, 2, 1997.

[236] T Florian Jaeger and Esteban Buz. Signal reduction and linguistic encoding. *The handbook of psycholinguistics*, pages 38–81, 2017.

[237] R J Zatorre, A C Evans, E Meyer, and A Gjedde. Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256(5058):846–849, May 1992.

[238] Robert J Zatorre and Pascal Belin. Spectral and temporal processing in human auditory cortex. *Cerebral cortex*, 11(10):946–953, 2001.

[239] Jan Rayman and Eran Zaidel. Rhyming and the right hemisphere. *Brain Lang.*, 40(1):89–105, January 1991.

[240] D Klein, B Milner, R J Zatorre, E Meyer, and A C Evans. The neural substrates underlying word generation: a bilingual functional-imaging study. *Proc. Natl. Acad. Sci. U. S. A.*, 92(7):2899–2903, March 1995.

[241] Sören Krach and Wolfgang Hartje. Comparison of hemispheric activation during mental word and rhyme generation using transcranial doppler sonography. *Brain Lang.*, 96(3):269–279, March 2006.

[242] D.A. Kareken, M. Lowe, S.H.A. Chen, J. Lurito, and V. Mathews. *Word rhyming as a probe of hemispheric language dominance with functional magnetic resonance imaging. Neuropsychiatry.* Neuropsychology, Behavioral Neurology, 2000.

[243] J. Pressing. Improvisation: methods and models. john a. In Sloboda, editor, *Generative processes in music*, page 129–178. Oxford, 1988.

[244] Jason P Mitchell, C Neil Macrae, and Mahzarin R Banaji. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, 50(4):655–663, May 2006.

[245] K. Kim and M.K. Johnson. Extended self: spontaneous activation of medial prefrontal cortex by objects that are 'mine'. social cognitive and affective, 2014.

[246] Michael D Fox, Abraham Z Snyder, Justin L Vincent, Maurizio Corbetta, David C Van Essen, and Marcus E Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A.*, 102(27):9673–9678, July 2005.

[247] Darya L Zabelina and Jessica R Andrews-Hanna. Dynamic network interactions supporting internally-oriented cognition. *Curr. Opin. Neurobiol.*, 40:86–93, October 2016.

[248] Roger E Beaty, Yoed N Kenett, Alexander P Christensen, Monica D Rosenberg, Mathias Benedek, Qunlin Chen, Andreas Fink, Jiang Qiu, Thomas R Kwapil, Michael J Kane, and Paul J Silvia. Robust prediction of individual creative ability from brain functional connectivity. *Proc. Natl. Acad. Sci. U. S. A.*, 115(5):1087–1092, January 2018.

[249] D.L. Zabelina and M.D. Robinson. Creativity as flexible cognitive control. psychology of aesthetics. *Creativity, and the*, 4(3):136, 2010.

[250] J.P. Guilford. Creativity: Yesterday. *today and tomorrow. The Journal of Creative*, 1(1):3–14, 1967.

[251] Teresa M Amabile. Social psychology of creativity: A consensual assessment technique. *J. Pers. Soc. Psychol.*, 43(5):997–1013, 1982.

[252] Andrew Goldman. Towards a cognitive–scientific research program for improvisation: Theory and an experiment. *Psychomusicology*, 23(4):210–221, 2013.

[253] Charles J Limb and Allen R Braun. Neural substrates of spontaneous musical performance: an FMRI study of jazz improvisation. *PLoS One*, 3(2):e1679, February 2008.

[254] Gabriel F Donnay, Summer K Rankin, Monica Lopez-Gonzalez, Patpong Jiradejvong, and Charles J Limb. Neural substrates of interactive musical improvisation: an FMRI study of 'trading fours' in jazz. *PLoS One*, 9(2):e88665, February 2014.

[255] S S Rahman. *THE NEUROSCIENCE OF MUSICAL CREATIVITY USING COMPLEXITY TOOLS*. 0195.

[256] Jing Lu, Hua Yang, Xingxing Zhang, Hui He, Cheng Luo, and Dezhong Yao. The brain functional state of music creation: An fMRI study of composers. *Sci. Rep.*, 5(1):12277, July 2015.

[257] Venla Sykäri. Beginning from the end: Strategies of composition in lyrical improvisation with end rhyme. *Oral tradit.*, 31(1), 2017.

[258] Raymond A R MacDonald and Graeme B Wilson. Musical improvisation and health: a review. *Psychol. Well Being*, 4(1), December 2014.

[259] J Erkkilä. From the unconscious to the conscious: Musical improvisation and drawings as tools in the music therapy of children. *Nordic Journal of Music Therapy*, 6(2):112–120, 1997.

[260] S C Gardstrom. Practical techniques for the development of complementary skills in musical improvisation. *Music Ther. Perspect.*, 19(2):82–87, January 2001.

[261] B Elan Dresher. Morris halle & Jean-Roger vergnaud (1987). an essay on stress. (current studies in linguistics 15). (cambridge, mass.: MIT press. pp. xi + 300. *Phonology*, 7(1):171–188, May 1990.

[262] M. Oostendorp. The grid of the french syllable. *Linguistics in the Netherlands*, 9:209–221, 1992.

[263] C G Clopper. Frequency of stress patterns in english: A computational analysis. *IULC Working Papers Online*, 2:1–9, 2002.

[264] L Pearl. Learning english metrical phonology: When probability distributions are not enough. In *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition*, pages 200–211. North America, 2008.

[265] Brett Kessler and Rebecca Treiman. Syllable structure and the distribution of phonemes in english syllables. *J. Mem. Lang.*, 37(3):295–311, October 1997.

[266] W.N. Francis and H. Kucera. Brown corpus manual. letters to the editor, 5(2, 1979.

[267] Sharon Rose and Rachel Walker. Harmony systems. In *The Handbook of Phonological Theory*, pages 240–290. Wiley-Blackwell, Oxford, UK, November 2011.

[268] Adam Wayment. Assimilation as attraction: Computing distance, similarity, and locality in phonology, September 2009.

[269] Tania S Zamuner, Louann Gerken, and Michael Hammond. Phonotactic probabilities in young children's speech production. *J. Child Lang.*, 31(3):515–536, August 2004.

[270] Rachel Walker and Michael Proctor. The organisation and structure of rhotics in american english rhymes. *Phonology*, 36(3):457–495, August 2019.

[271] Bruce Hayes and Colin Wilson. A maximum entropy model of phonotactics and phonotactic learning. *Linguist. Inq.*, 39(3):379–440, July 2008.

[272] C.E. Cairns. Phonotactics, markedness and lexical, 1988.

[273] J Ackerman. Entropy of sounds - sonnets to battle rap. *Proceedings of the Cognitive Science Society*, 2020.

[274] P A Luce and D B Pisoni. Recognizing spoken words: the neighborhood activation model. *Ear Hear.*, 19(1):1–36, February 1998.

[275] Benjamin Munson and Nancy Pearl Solomon. The effect of phonological neighborhood density on vowel articulation. *J. Speech Lang. Hear. Res.*, 47(5):1048–1058, October 2004.

[276] Michael S Vitevitch and Paul A Luce. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J. Mem. Lang.*, 40(3):374–408, April 1999.

[277] Michael S Vitevitch, Jonna Armbruster, and Shinying Chu. Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. *J. Exp. Psychol. Learn. Mem. Cogn.*, 30(2):514–529, March 2004.

[278] Holly L Storkel, Jonna Armbrüster, and Tiffany P Hogan. Differentiating phonotactic probability and neighborhood density in adult word learning. *J. Speech Lang. Hear. Res.*, 49(6):1175–1192, December 2006.

[279] Susanne Gahl, Yao Yao, and Keith Johnson. Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *J. Mem. Lang.*, 66(4):789–806, May 2012.

[280] U H Frauenfelder, R H Baayen, and F M Hellwig. Neighborhood density and frequency across languages and modalities. *J. Mem. Lang.*, 32(6):781–804, December 1993.

[281] Judith A Gierut. Phonological complexity and language learnability. *Am. J. Speech. Lang. Pathol.*, 16(1):6–17, February 2007.

[282] O Lavi-Rotbain and I Arnon. *Visual Statistical Learning Is Facilitated in Zipfian Distributions. 1.*

[283] O. Lavi-Rotbain and I. Arnon. The learnability consequences of zipfian distributions in, 2022.

[284] Bruno de Cara and Usha Goswami. Phonological neighbourhood density: effects in a rhyme awareness task in five-year-old children. *J. Child Lang.*, 30(3):695–710, August 2003.

[285] V M Garlock, A C Walley, and J L Metsala. Age-of-acquisition, word frequency and neighbourhood density effects on spoken word recognition by children and adults. *Journal of Memory & Language*, 45:468–492, 2001.

[286] Amanda C Walley. Spoken word recognition by young children and adults. *Cogn. Dev.*, 3(2):137–165, April 1988.

[287] Joseph C Toscano, Nathaniel D Anderson, and Bob McMurray. Reconsidering the role of temporal order in spoken word recognition. *Psychon. Bull. Rev.*, 20(5):981–987, October 2013.

[288] G Grossi, D Coch, S Coffey-Corina, P J Holcomb, and H J Neville. Phonological processing in visual rhyming: a developmental erp study. *J. Cogn. Neurosci.*, 13(5):610–625, July 2001.

[289] Hoft. H. F. W. Counting and visualizing rhyme patterns in sonnets. *Combinatoria Poetica.*, 2009.

[290] Jordan A Ackerman. Of pieces and patterns: Modeling poetic devices. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 43, 2021.

[291] Rudolf Mayer, Robert Neumayer, and Andreas Rauber. Rhyme and style features for musical genre classification by song lyrics. In *Ismir*, pages 337–342, 2008.

[292] R. Rajan and A.A. Raju. Poetic meter classification using acoustic cues. In *2018 International Conference on Signal Processing and Communications*, page 31–35. IEEE, 2018-07.

[293] Rodolfo Delmonte. Exploring shakespeare's sonnets with SPARSAR. *Linguist. Lit. Stud.*, 4(1):61–95, January 2016.

[294] H F Höft. Visualizing rhyme patterns in sonnet sequences. In *Proceedings of Bridges 2015: Mathematics*, pages 363–366. Tessellations Publishing, Music, Art, Architecture, 2015.

[295] L J George and H F Höft. Visualization of rhyme patterns in two sonnet sequences. *Bridges Leeuwarden*, pages 265–266, 2009.

[296] Mitchell Ohriner. Metric ambiguity and flow in rap music: A corpus-assisted study of outkast's "mainstream" (1996). *Empir. Musicol. Rev.*, 11(2):153, January 2017.

[297] J. Gran. Two corpus-based approaches to rap flow. empirical. *Musicology*, 11(2):185, 2016.

[298] Eric. Algorithm that counts rap rhymes and scouts mad lines. `https://mining4meaning.com/2015/02/13/raplyzer/`, February 2015. Accessed: 2022-4-18.

[299] S Nathaniel and S Rastogi. Identifying assonant clusters in poetry.

[300] H Hirjee and D G Brown. AUTOMATIC DETECTION OF INTERNAL AND IMPERFECT RHYMES IN RAP LYRICS. *Oral Session*, 2009.

[301] Karteek Addanki and Dekai Wu. Unsupervised rhyme scheme identification in hip hop lyrics using hidden Markov models. In *International Conference on Statistical Language and Speech Processing*, pages 39–50. Springer, 2013.

[302] A Das and B Gambäck. Poetic machine: Computational creativity for automatic poetry generation in bengali. 9, 2014.

[303] Xingxing Zhang and Mirella Lapata. Chinese poetry generation with recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Stroudsburg, PA, USA, 2014. Association for Computational Linguistics.

[304] Zhe Wang, Wei He, Hua Wu, Haiyang Wu, Wei Li, Haifeng Wang, and Enhong Chen. Chinese poetry generation with planning based neural network. October 2016.

[305] Peter Potash, Alexey Romanov, and Anna Rumshisky. GhostWriter: Using an LSTM for automatic rap lyric generation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, 2015. Association for Computational Linguistics.

[306] H Manurung, G Ritchie, and H Thompson. *Towards a computational model of poetry generation*. 2000.

[307] H. Manurung. An evolutionary algorithm approach to poetry generation, 2004.

[308] Cmu pronunciation dictionary. *Carnegie Mellon University*, pages 79–86, 2000.

[309] B Winter and A Wedel. The co-evolution of speech and the lexicon: The interaction of functional pressures, redundancy, and category variation. *Topics in cognitive science*, 8(2):503–513, 2016.

[310] P Li and M C Yip. Context effects and the processing of spoken homophones. *Cognitive processing of the Chinese and the Japanese languages*, pages 69–89, 1998.

[311] C. W. M. Yip. Spoken word recognition of chinese homophones: The role of context and tone neighbors. *Psychologia*, 43(2):135–143, 2000.

[312] Reddy. S and Knight K. Unsupervised discovery of rhyme schemes. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies:*, 2, 2011.

[313] Weide R. L. Carnegie mellon pronouncing dictionary. *http://www.speech.cs.cmu.edu/cgi-bin/cmudict*, release 0.7b, 1998.

[314] Wolfe, P. M. *Linguistic change and the great vowel shift in English*, volume 42. Univ of California Press., 1972.

[315] Edward N Lorenz. Deterministic nonperiodic flow. *Journal of atmospheric sciences*, 20(2):130–141, 1963.

[316] Drew H Abney, Alexandra Paxton, Rick Dale, and Christopher T Kello. Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, 143(6):2304, 2014.

[317] Daniel C Richardson and Rick Dale. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science*, 29(6):1045–1060, 2005.

[318] Michael J Richardson. *Distinguishing the noise and attractor strength of rhythmic and coordinated limb movements using recurrence analysis*. University of Connecticut, 2005.

[319] Kevin Shockley and Michael T Turvey. Encoding and retrieval during bimanual rhythmic coordination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5):980, 2005.

[320] Erin McCormick, Leslie M Blaha, and Cleotilde Gonzalez. Exploring dynamic decision making strategies with recurrence quantification analysis. In *CogSci*, 2020.

[321] Claude E. Shannon. Prediction and entropy of printed English. *Bell System Technical Journal*, 30:50–64, 1951.

[322] David J C MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2005.

[323] T. Cover and R. King. A convergent gambling estimate of the entropy of English. *IEEE Transactions on Information Theory*, 24(4):413–421, July 1978.

[324] Marcelo A. Montemurro and Damián H. Zanette. Universal entropy of word ordering across linguistic families. *P L o S One*, 6(5), 2011.

[325] A A Markov. An example of statistical investigation of the text eugene onegin concerning the connection of samples in chains. *Science in Context*, 19(4):591–600, 2006.

[326] J. Goldsmith and J. Riggle. Information theoretic approaches to phonological structure: the case of finnish vowel harmony. *Natural Language Linguistic Theory*, 30(3):859–896, 2012.

[327] Werner Ebeling and Thorsten Poeschel. Entropy and Long range correlations in literary English. *Europhysics Letters (EPL)*, 26(4):241–246, May 1994. arXiv: cond-mat/0204108.

[328] Scott J Simon. A multi-dimensional entropy model of jazz improvisation for music information retrieval. *University of North Texas*, page 196, 2005.

[329] Scott J. Simon. MEASURING INFORMATION IN JAZZ IMPROVISATION. 2007.

[330] Ryuji Suzuki, John R. Buck, and Peter L. Tyack. Information entropy of humpback whale songs. *The Journal of the Acoustical Society of America*, 119(3):1849–1866, February 2006.

[331] Pierce, John R. *An Introduction to Information Theory: Symbols, Signals and Noise*. 1980.

[332] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.

[333] Ramon Ferrer i Cancho and Ricard V Sol. Two regimes in the frequency of words and the origins of complex lexicons: Zipf?S law revisited. *J. Quant. Linguist.*, 8(3):165–173, December 2001.

[334] Matthieu Cristelli, Michael Batty, and Luciano Pietronero. There is more than a power law in zipf. *Sci. Rep.*, 2(1):812, November 2012.

[335] Jake Ryland Williams, Paul R Lessard, Suma Desu, Eric Clark, James P Bagrow, Christopher M Danforth, and Peter Sheridan Dodds. Zipf's law holds for phrases, not words. June 2014.

[336] Le Quan Ha, Philip Hanna, Ji Ming, and Francis Jack Smith. Extending zipf's law to n-grams for large corpora. *Artificial Intelligence Review*, 32(1):101–113, 2009.

[337] Colin Martindale, S M Gusein-Zade, Dean McKenzie, and Mark Yu Borodovsky. Comparison of equations describing the ranked frequency distributions of graphemes and phonemes. *J. Quant. Linguist.*, 3(2):106–112, August 1996.

[338] Y Tambovtsev and C Martindale. *PHONEME FREQUENCIES FOLLOW A YULE DISTRIBUTION*. 2007.

[339] A Baumann, T Matzinger, and K Kaźmierski. Phonotactics is affected by statistical scaling laws less than the lexicon. 1, 2019.

[340] Aaron Clauset, Cosma Rohilla Shalizi, and M E J Newman. Power-law distributions in empirical data. *SIAM Rev. Soc. Ind. Appl. Math.*, 51(4):661–703, November 2009.

[341] Anna Deluca and Álvaro Corral. Fitting and goodness-of-fit test of non-truncated and truncated power-law distributions. *Acta Geophysica*, 61(6):1351–1394, 2013.

[342] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*, 2020.

[343] Siyuan Liu, Ho Ming Chow, Yisheng Xu, Michael G Erkkinen, Katherine E Swett, Michael W Eagle, Daniel A Rizik-Baer, and Allen R Braun. Neural correlates of lyrical improvisation: an fmri study of freestyle rap. *Scientific reports*, 2(1):1–8, 2012.

[344] Peter B Denes. On the statistics of spoken english. *The Journal of the Acoustical Society of America*, 35(6):892–904, 1963.

[345] Geoffrey Leech, Paul Rayson, et al. *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge, 2014.

[346] Cristiano Gino Furiassi. Spoken and written learner english: A quantitative analysis of icle-it and lindsei-it. 2004.

[347] Kyle Gray. Fresh coast all star cypher, 2008.

[348] Radim Rehurek and Petr Sojka. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2), 2011.

[349] R Brooke Lea, Andrew Elfenbein, and David N Rapp. Rhyme as resonance in poetry comprehension: An expert–novice study. *Memory & Cognition*, 49(7):1285–1299, 2021.

[350] Samuel F Feng, Siyu Wang, Sylvia Zarnescu, and Robert C Wilson. The dynamics of explore–exploit decisions reveal a signal-to-noise mechanism for random exploration. *Scientific reports*, 11(1):1–15, 2021.

[351] Mauricio Fadel Argerich. Entropy Regularization in Reinforcement Learning. `https://towardsdatascience.com/entropy-regularization-in-reinforcement-learning-a6fa6d7598df`, 2020.

[352] Markus Wulfmeier, Peter Ondruska, and Ingmar Posner. Maximum entropy deep inverse reinforcement learning. *arXiv preprint arXiv:1507.04888*, 2015.

[353] Jingbin Liu, Xinyang Gu, and Shuai Liu. Policy optimization reinforcement learning with entropy regularization. *arXiv preprint arXiv:1912.01557*, 2019.

Appendix A

Appendix Title

Catgegory	Corpus	Source	Word Count
Speech	Dialogue	UCSB Speech Corpus	130,649
Speech	WebChat	NPS Chat	45,010
Speech	Infant Directed Speech	CHILDES	14,468
Fiction	Adventure	Gutenberg	69,342
Fiction	Buster Brown	Gutenberg	18,963
Fiction	chesterton-brown	Gutenberg	86,063
Fiction	Alice In Wonderland	Gutenberg	34,110
Fiction	Edgeworth Parents	Gutenberg	210,663
Fiction	Sense and Sensibility	Gutenberg	141,576
Fiction	Hamlet - Shakespeare	Gutenberg	37,360
Fiction	Macbeth - Shakespeare	Gutenberg	23,140
Fiction	Caesar - Shakespeare	Gutenberg	25,833
Fiction	Science Fiction	BrownCorpus	14,470
Fiction	Romance	BrownCorpus	70,022
Fiction	Mystery	BrownCorpus	57,169
Fiction	Humor	BrownCorpus	21,695
Fiction	Fiction	BrownCorpus	68,488
Musical Lyrics	Pop	lyricsfreak.com	493,213
Musical Lyrics	Country	lyricsfreak.com	377,029
Musical Lyrics	Electronic	lyricsfreak.com	387,182
Musical Lyrics	Folk	lyricsfreak.com	363,884
Musical Lyrics	Indie	lyricsfreak.com	390,271
Musical Lyrics	Jazz	lyricsfreak.com	340,692
Musical Lyrics	Metal	lyricsfreak.com	348,383
Musical Lyrics	Rock	lyricsfreak.com	384,744
Musical Lyrics	Hip-Hop	lyricsfreak.com	991,923
Musical Lyrics	Improvised Rap (Harry Mack Omegle)	Annotated by Author	46,194
Non-Fiction	Editorials	BrownCorpus	61,604
Non-Fiction	Government Documents	BrownCorpus	70,117
Non-Fiction	Hobbies	BrownCorpus	82,345
Non-Fiction	Inaugural Addresses	inaugural Addresses	145,735
Non-Fiction	Belles Lettres	BrownCorpus	173,096
Non-Fiction	Learned - Academic	BrownCorpus	181,888
Non-Fiction	Lore	BrownCorpus	110,299
Non-Fiction	News 1	BrownCorpus	100,554
Non-Fiction	News 2	Reuters	1,253,696
Non-Fiction	Religion	BrownCorpus	39,399
Non-Fiction	Reviews	BrownCorpus	40,704
Non-Fiction	King James Bible	BrownCorpus	1,010,654
Poetry	Sonnets	poetryfoundation.org	61,140
Poetry	Allusion	poetryfoundation.org	14,091
Poetry	Ballads	poetryfoundation.org	34,973
Poetry	Refrain	poetryfoundation.org	18,533
Poetry	Blank Verse	poetryfoundation.org	123,597
Poetry	Confessional	poetryfoundation.org	17,797
Poetry	Couplet	poetryfoundation.org	146,604
Poetry	Dramatic Monologue	poetryfoundation.org	36,327
Poetry	Elegy	poetryfoundation.org	45,308
Poetry	Epic	poetryfoundation.org	131,910
Poetry	Rhymed Stanza	poetryfoundation.org	276,460
Poetry	Free Verse	poetryfoundation.org	640,017
Poetry	Imagery	poetryfoundation.org	17,936
Poetry	Metaphor	poetryfoundation.org	35,279
Poetry	Mixed	poetryfoundation.org	21,446
Poetry	Persona	poetryfoundation.org	32,030
Poetry	Prose Poem	poetryfoundation.org	14,140
Poetry	Series/Sequence	poetryfoundation.org	48,407
Poetry	Battle Rap	battlerap.com	19,116
Poetry	Paradise Lost	Gutenberg	96,825
Poetry	Leaves of Grass	Gutenberg	154,883

Figure A.1: Short Descriptions of Corpora collected for Diverse Creative English Texts dataset. For the current exploration, I only take the first 14K vowels and 22K consonants from each corpus.

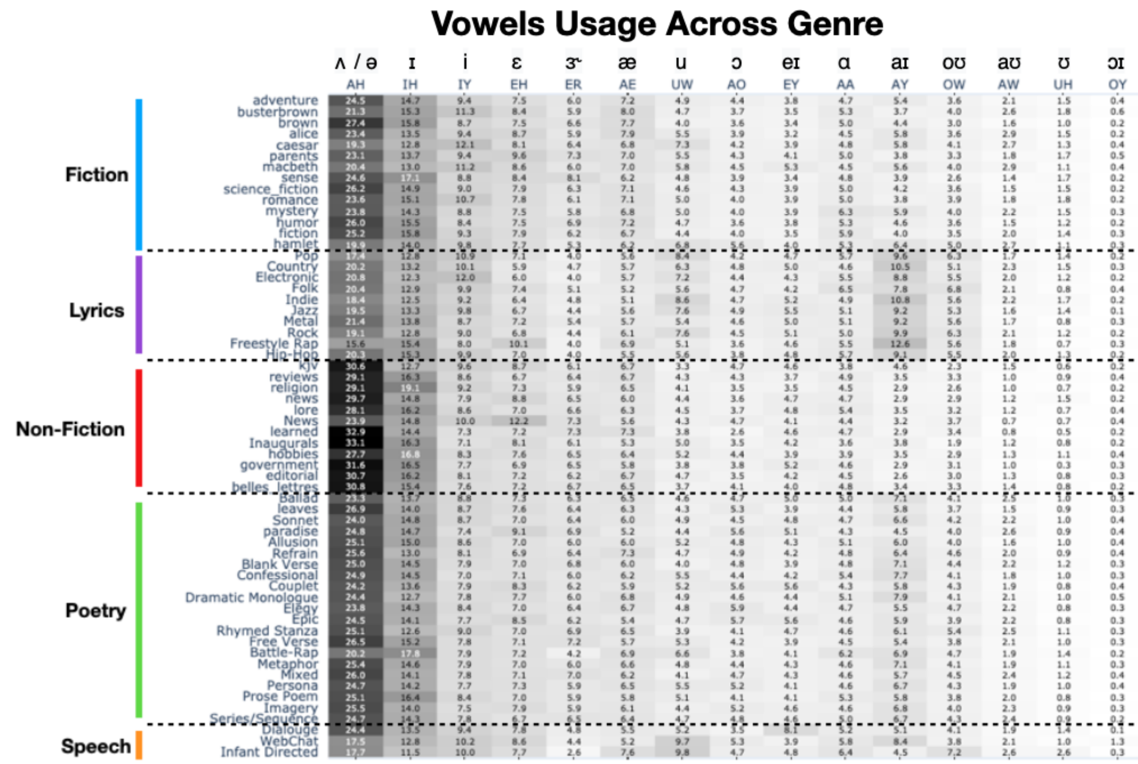


Figure A.2: DCET Vowel Term Frequency Rates by category and corpora (document)

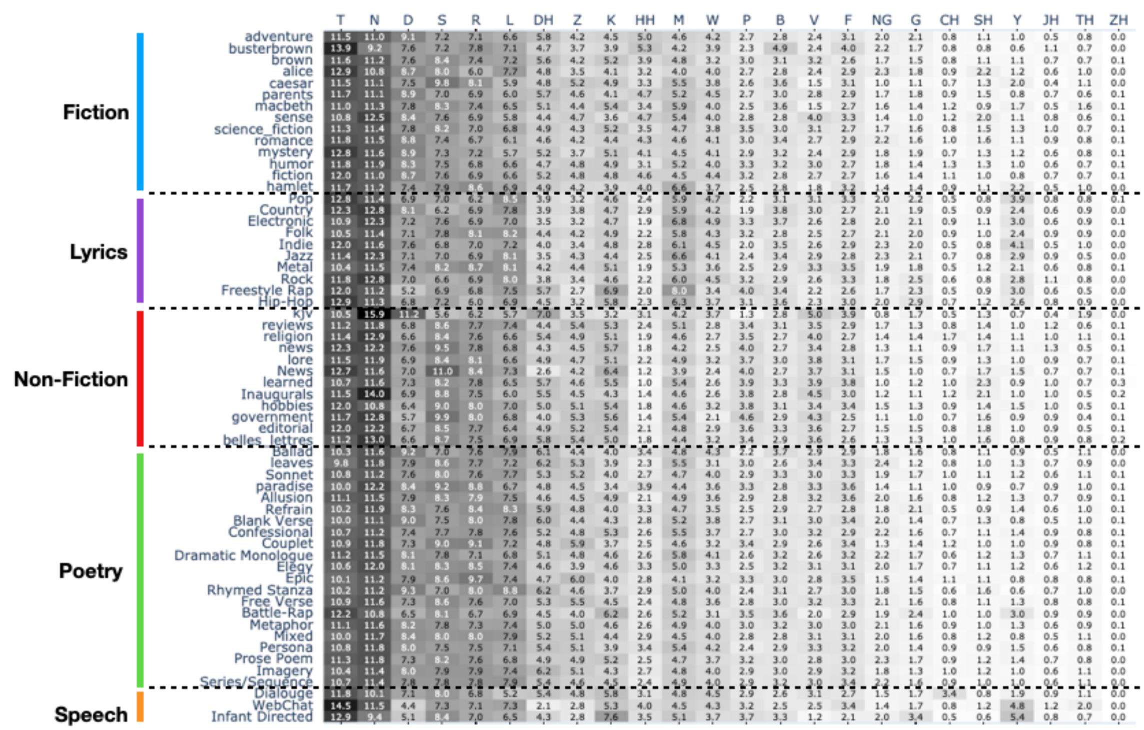


Figure A.3: DCET Consonant Term Frequency Rates by category and corpora (document)

	AH	IH	ER	IY	EH	AA	AE	EY	OW	AO	AY	UW	AW	UH	OY
PubMed	30.8	15.3	8.7	8.4	6.1	6.0	5.9	4.6	4.6	3.8	2.6	2.4	0.4	0.3	0.2
NYTimes	28.7	15.3	9.1	8.3	7.8	5.4	5.2	4.9	3.5	3.5	3.4	2.7	1.0	0.8	0.3
NIPS	27.8	16.4	9.5	9.4	7.0	6.0	5.4	5.0	3.2	2.7	2.5	2.5	1.6	0.7	0.4
KOS	24.8	15.1	10.4	8.9	8.1	5.9	5.4	4.8	4.4	3.3	3.3	2.7	1.3	1.2	0.3
Enron	28.0	10.8	10.4	9.6	7.2	6.8	6.8	6.0	4.8	3.2	3.2	1.6	1.2	0.4	0.0

Figure A.4: UCI BOW Vowel Term Frequency Rates by category

	T	N	S	R	K	L	D	P	M	F	B	Z	V	NG	SH	JH	G	W	Y	HH	CH	TH	ZH	DH
PubMed	12.2	11.9	10.6	9.5	8.5	8.1	6.8	5.7	5.1	3.2	2.7	2.6	2.4	2.3	1.6	1.4	1.2	1.0	0.9	0.7	0.7	0.7	0.2	0.0
NYTimes	11.3	10.9	9.1	7.9	7.8	7.8	7.6	6.6	5.1	5.1	3.0	2.8	2.1	2.0	2.0	1.9	1.5	1.5	1.1	1.1	1.0	0.4	0.2	0.1
NIPS	13.5	12.4	9.3	8.5	8.4	8.3	6.0	5.9	5.7	2.7	2.6	2.6	2.2	2.1	1.9	1.6	1.6	1.3	1.3	0.8	0.7	0.4	0.3	0.2
KOS	11.9	11.2	9.5	8.8	8.0	7.8	7.0	5.4	5.1	4.6	3.3	2.5	2.4	2.4	1.8	1.6	1.6	1.4	1.1	1.0	0.9	0.4	0.1	0.1
Enron	12.1	11.3	9.7	9.0	8.3	7.8	7.1	5.7	5.2	3.0	2.8	2.5	2.4	2.3	2.1	1.7	1.7	1.4	1.2	1.0	0.9	0.6	0.1	0.1

Figure A.5: UCI BOW Consonant Term Frequency Rates by category

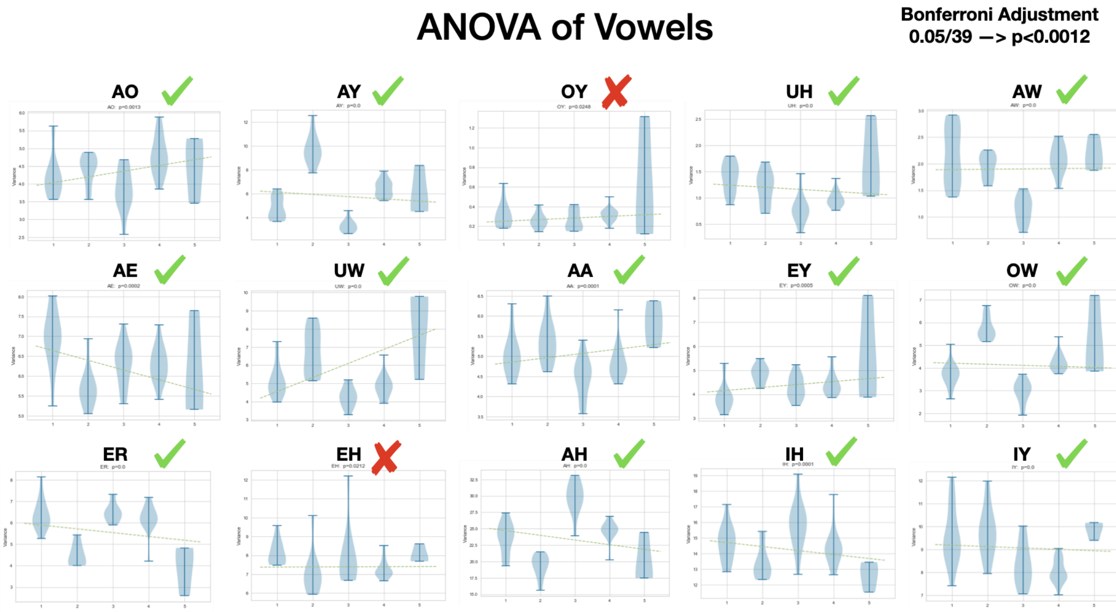


Figure A.6: Plots of Vowel Usage Rates by phoneme and category. Y-axis is variance, columns on X-axis are 1-Fiction, 2-Lyrics, 3-Non-Fiction, 4-Poetry, 5-Speech. A green check indicates significant group differences.

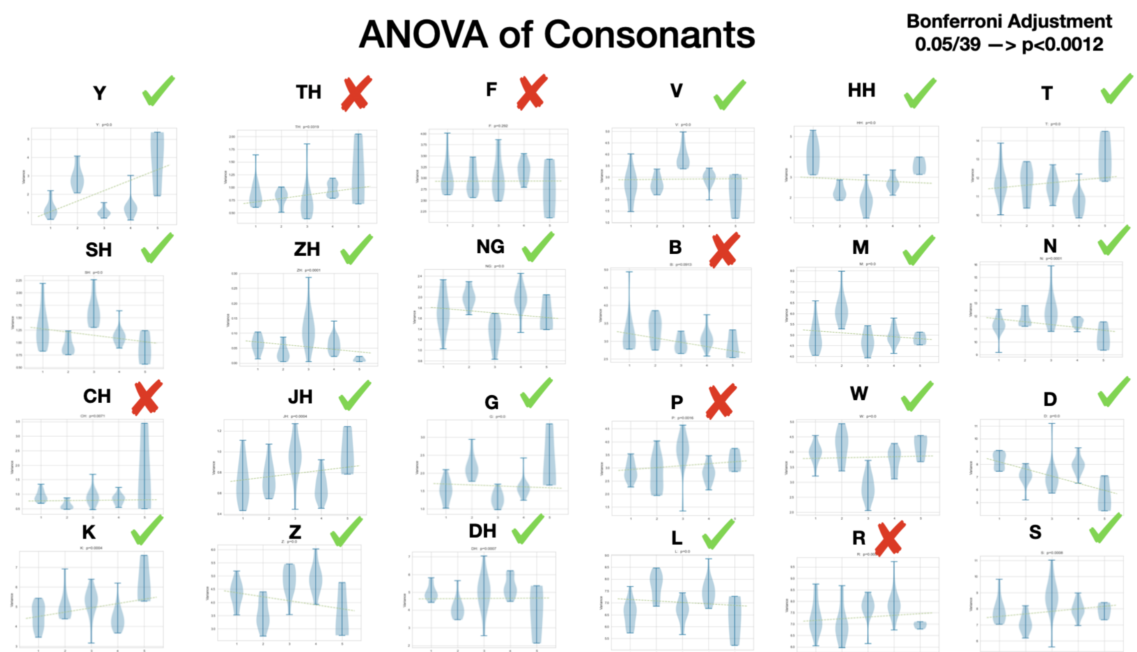


Figure A.7: Plots of Consonant Usage Rates by phoneme and category. Y-axis is variance, columns on X-axis are 1-Fiction, 2-Lyrics, 3-Non-Fiction, 4-Poetry, 5-Speech. A green check indicates significant group differences.