

Die Entwicklung der Songtextkomplexität: Eine Analyse

Lukas Reinhardt

2024-09-13

Zusammenfassung

Diese Arbeit [1] beschäftigt sich mit dem Thema, ob die Komplexität der Songtexte über die letzten Jahrzehnte abgenommen hat. Analysiert werden Daten aus den Jahren 1950 bis 2023. Bisherige Studien, wie die von Varnum et al. [2] und Choi und Downie [3], haben sich bereits mit dieser Entwicklung befasst. Diese Arbeit erweitert den Ansatz durch die Metriken TTR, Shannon-Index und ARI sowie die Berücksichtigung einzelner Genres. Ziel ist es, ein ganzheitlicheres Bild der Entwicklung der Textkomplexität zu erhalten. Die Ergebnisse zeigen, dass, wie in den bisherigen Studien [2–3], die lexikalische Vielfalt abgenommen hat, während der ARI eine gestiegene strukturelle Komplexität impliziert. Somit zeigt die Arbeit, dass die Songtexte zwar an lexikalischer Vielfalt verloren haben, gleichzeitig jedoch eine höhere strukturelle Komplexität erreicht wurde.

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung | 2 |
| 1.1 | Problemstellung | 2 |
| 1.2 | Ziele der Arbeit | 2 |
| 1.3 | Aufbau der Arbeit | 2 |
| 2 | Verwandte Arbeit | 2 |
| 2.1 | Varnum et al. (2021) - Why are Song Lyrics Becoming Simpler?[2] | 2 |
| 2.2 | Choi und Downie (2019) - A Trend Analysis on Concretness of Popular Song Lyrics [3] | 2 |
| 2.3 | Beitrag dieser Arbeit | 3 |
| 3 | Hypothese | 3 |
| 3.1 | Forschungsfrage | 3 |
| 4 | Methodik | 3 |
| 4.1 | Datenbeschaffung | 3 |
| 4.2 | Auswahl der Komplexitätsmetriken | 5 |
| 5 | Auswertung der Ergebnisse | 8 |
| 5.1 | Metriken der Billboard Hot 100 | 8 |
| 5.2 | Analyse der Genres im Vergleich zur Billboard Hot 100 | 9 |
| 6 | Fazit | 11 |
| 6.1 | Einschränkungen | 11 |
| 6.2 | Zukünftige Forschung | 11 |
| | Referenzen | 12 |

1 Einleitung

Die Musikindustrie hat sich in den letzten Jahrzehnten gewandelt, was sich möglicherweise auf die Songtexte populärer Musik ausgewirkt hat. Technologische Entwicklungen, wie Streamingdienste, machten Musik zunehmend zugänglicher und führten zu einer stärkeren Kommerzialisierung. Diese und andere Wandlungen könnten die lyrische Komplexität beeinflusst haben. Diese Arbeit untersucht, ob sich die Komplexität der Songtexte, gemessen an lexikalischer Vielfalt, Informationsdichte und struktureller Komplexität, im Laufe der Zeit verändert hat.

1.1 Problemstellung

Diese Arbeit untersucht, ob die Textkomplexität in den populärsten Musikstücken im Laufe der Jahrzehnte kontinuierlich abgenommen hat. Während frühere Studien [2–3] bereits Trends in der Komplexität von Songtexten analysierten, bleibt unklar, wie sich diese Entwicklungen in verschiedenen Genres im Vergleich zu den allgemeinen Top 100 Songs verhalten haben.

1.2 Ziele der Arbeit

Ziel dieser Arbeit ist es, die Entwicklung der Textkomplexität, in den Billboard Hot 100 [4] von 1950 bis 2023 zu analysieren. Dabei soll untersucht werden, ob Songtexte über die Jahrzehnte hinweg Veränderungen in Bezug ihrer Komplexität erfahren haben. Darüber hinaus wird betrachtet, wie diese Entwicklungen in verschiedenen Musikgenre wie Pop, Rock, Rap, R&B und Country im Vergleich zu den allgemeinen Trends der Billboard Hot 100 verhalten.

1.3 Aufbau der Arbeit

Diese Arbeit ist in mehrere Abschnitte gegliedert, nach der Einleitung folgen die verwandten Arbeiten, der relevante Studien zusammenfasst und deren Forschungsstand darlegt. In der Methodik werden verwendete Datenquellen, Metriken beschrieben. Anschließend werden die Ergebnisse der Metriken analysiert und interpretiert. Abschließend fasst das Fazit die wesentlichen Erkenntnisse zusammen und gibt Informationen für mögliche zukünftige Forschungen.

2 Verwandte Arbeit

2.1 Varnum et al. (2021) - Why are Song Lyrics Becoming Simpler?[2]

Vanum et al. untersuchten die lyrische Komplexität von über 14.000 Songs der Billboard Hot 100 zwischen 1958 und 2016. Sie nutzten die Komprimierbarkeit der Texte als Maß für lyrische Komplexität und stellten fest, dass in den Jahren mit mehr neuen Musikproduktionen einfachere Texte bevorzugt wurden. Die Analyse basiert auf LZ77-Komprimierungsalgorithmen, wodurch Wiederholungen in den Texten erfasst werden.

Zwar erfasst die Komprimierbarkeit Aspekte der lexikalischen Vielfalt, jedoch werden andere wichtige Dimensionen der lyrischen Komplexität, wie die Informationsdichte, nicht berücksichtigt. Zudem wird der Einfluss verschiedener Genres in dieser Untersuchung nicht näher beleuchtet.

2.2 Choi und Downie (2019) - A Trend Analysis on Concreteness of Popular Song Lyrics [3]

Choi und Downie analysierten die Konkretisierung (Concreteness) der Songtexte populärer Musik von 1965 bis 2015. Concreteness bezieht sich auf die Greifbarkeit und Bildhaftigkeit von Wörtern, dabei sind konkrete Begriffe wie “Auto” klarer definiert als abstrakte Begriffe wie “Liebe”. Die Studie zeigt, dass die Concreteness zunächst abnahm, dann aber mit dem Aufstieg von Genres wie Hip-Hop und Rap wieder anstieg.

Auch wenn die Metrik der Concreteness interessante Ergebnisse liefert, werden andere Aspekte der lyrischen Komplexität, wie die lexikalische Vielfalt und Informationsdichte, in dieser Studie nicht betrachtet. Diese

könnten zusätzliche Perspektiven auf die Entwicklung der Songtexte liefern.

2.3 Beitrag dieser Arbeit

Im Gegensatz zu den genannten Studien verfolgt diese Arbeit einen umfassenderen Ansatz zur Bewertung der lyrischen Komplexität. Während frühere Untersuchungen, wie die von Varnum et al.[2] sowie Choi und Downie [3], spezifische Aspekte wie Komprimierbarkeit oder Konkretisierung betrachteten, wird hier die Vielfalt und Informationsdichte der Songtexte aus mehreren Perspektiven analysiert. Ein weiterer wichtiger Aspekt ist die Untersuchung der Entwicklung der Genres, die in den bisherigen Studien kaum betrachtet wurden. Dabei kann die Untersuchung das Genre zeigen, ob und wie sich die Genres und die Billboard Hot 100 [4] gegenseitig beeinflusst haben. Dadurch trägt diese Arbeit dazu bei, das Verständnis der lyrischen Komplexität zu erweitern und bietet neuen Einblicke auf die Veränderung in der Musikindustrie.

3 Hypothese

Die Hypothese, die in dieser Arbeit geprüft wird, lautet:

Die Qualität, Komplexität der Songtexte hat sich über die letzten Jahrzehnte verschlechtert.

Diese Annahme basiert auf bisherigen Studien, die auf eine Vereinfachung und Anstieg von Wiederholungen in populären Songs hinweisen. Besonders der Einfluss der Kommerzialisierung und die Dominanz bestimmter Musikgenres könnten dazu beigetragen haben, dass die Komplexität der Texte abgenommen hat.

3.1 Forschungsfrage

Die zentrale Forschungsfrage lautet:

Hat sich die Qualität der Songtexte in den letzten Jahrzehnten verringert und wie unterscheidet sich diese Entwicklung im Vergleich zu einzelnen Musikgenres?

Diese Frage zielt darauf ab, sowohl die allgemeine Entwicklung der Songtexte zu untersuchen als auch zu analysieren, ob und wie sich Trends in spezifischen Genres zu den allgemeinen Entwicklungen unterscheiden.

4 Methodik

In diesem Kapitel wird der methodische Ansatz der Arbeit beschrieben, der sich auf die Sammlung, Vorverarbeitung und Analyse der Daten konzentriert.

4.1 Datenbeschaffung

Die Daten für die Arbeit stammen aus mehreren Quellen. Die Billboard API [4] liefert die grundlegenden Daten der Charts, die Genius API [5] wird genutzt, um die Songtexte dieser Daten zu beschaffen. Um eine Analyse der Textkomplexität in den verschiedenen Genres durchzuführen, wurden genrespezifische Songs ebenfalls bezogen. Die MusicBrainz API [6] wurde verwendet, um die fehlenden Genre-Informationen der Billboard Hot 100 zu ergänzen, dadurch konnten die am häufigsten vertretenen Genres identifiziert werden.

4.1.1 Billboard API und Genre-Crawling

Zuerst wurden die allgemeinen Billboard Hot 100 Songs von 1950 bis 2023 über die API bezogen. Diese Daten umfassen die 100 beliebtesten Songs in den jeweiligen Jahren, woraus sich 7399 Songs ergeben haben.

Zusätzlich wurden genrespezifische Charts von Billboard bezogen, um die Top-Songs der Genres zu erfassen. Diese genrespezifischen Songs dienen der Analyse, wie die einzelnen Genres sich entwickelt haben und die Billboard Hot 100 beeinflusst haben.

4.1.2 Auswahl der Genres

Die genrespezifischen Charts wurden basierend auf den fünf am häufigsten vertretenen Genres in den Billboard Hot 100 gewählt. Diese Daten wurden mithilfe der MusicBrainz API[6] ergänzt. Dies war nötig, da weder die Billboard API[4] noch die Genius API[5] Genres geliefert haben. Die von MusicBrainz gelieferten Genres waren jedoch nicht einheitlich, was eine genauere Auswertung der Daten erforderlich machte.

Nach einer Sichtung der initialen Daten wurden die am häufigsten auftretenden Genres aus der Liste der Subgenera und anderen Irregularitäten ausgewählt. Diese Auswahl wurde genutzt, um mittels eines Regex-basierten Verfahrens die inkonsistenten Daten auf, die am häufigsten vorkommen der Genres zu mappen und zu zählen (siehe Tabelle 1).

Tabelle 1

| Genre | Count |
|-------------|-------|
| pop | 4930 |
| rock | 3796 |
| classic | 2148 |
| r&b | 1393 |
| country | 1205 |
| rap/hip-hop | 1191 |
| alternative | 1072 |
| indie | 925 |
| 80s | 546 |
| reggae | 271 |
| latin | 111 |
| edm | 55 |

Tabelle 1: Die Tabelle zeigt die Anzahl der Songs pro Genre, basierend auf den in den Billboard Hot 100 am häufigsten vertretenen Genres.

Diese Liste gibt uns Auskunft über die relevantesten Genres innerhalb der Billboard Hot 100. Wichtig ist zu betrachten, dass es auch doppelt belegte Songs gibt, wodurch aufsummieren der Genres keinen Sinn macht. Auf dieser Basis wurden die zu crawlenden Genre-Charts gewählt: Pop, Rock, R&B, Country, Rap. Diese wurden danach mittels Matching des Titels geprüft, ob sie auch in den Billboard Hot 100 vorhanden sind. Da Classic kein Matching hatte, wurde es aus den Top 5 entfernt und Rap stattdessen hinzugefügt. Durch dieses Matching haben wir ab etwa 1992 eine 60% Vollständigkeit der Billboard Daten in Form der Genres (siehe Abb. 1).

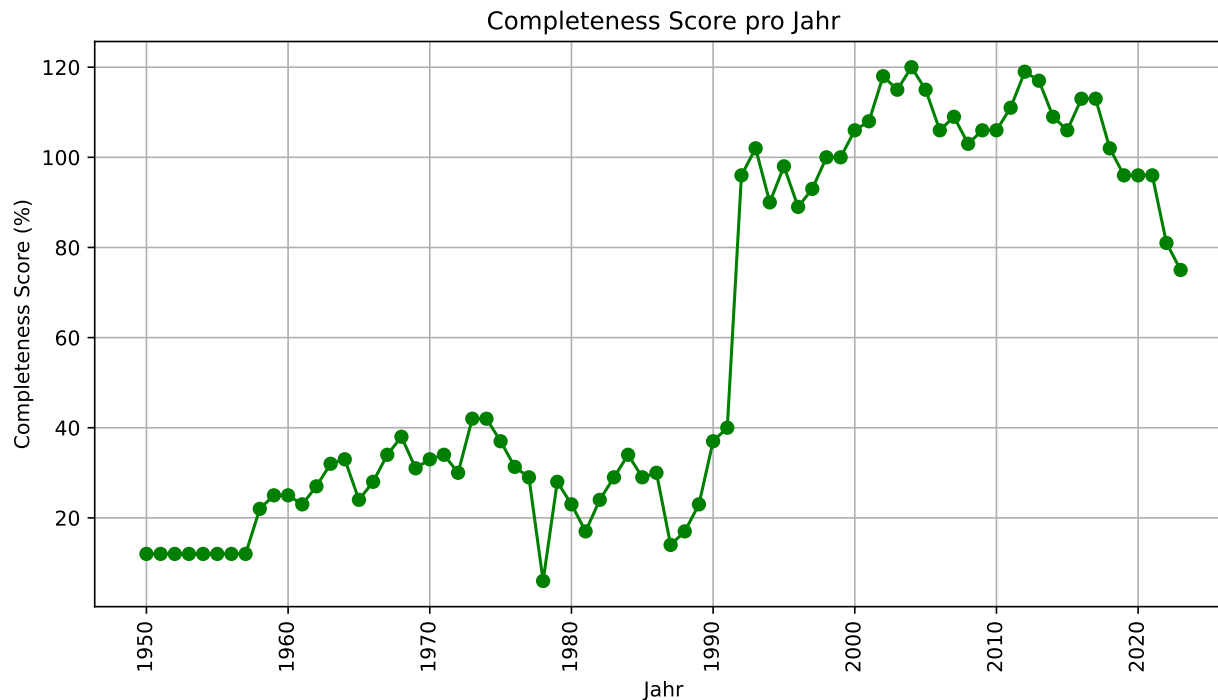


Abbildung 1: Der Plot zeigt den Completeness Score der Top 100 Songs pro Jahr von 1950 bis 2023. Auf der x-Achse sind die Jahre, auf der y-Achse der Completeness Score in Prozent. Ein höherer Score bedeutet einen größeren Anteil genrespezifischer Songs in den Top 100 für das jeweilige Jahr.

4.1.3 Genius API für Songtexte

Nachdem die Billboard- und genrespezifischen Daten erfasst wurden, war die Abrufung der Songtexte über die Genius API[5] der nächste Schritt. Genius bietet eine umfangreiche Datenbank für Songtexte, die für die Untersuchung der linguistischen Komplexität eine zentrale Rolle spielt. Für jeden Song der allgemeinen Billboard Hot 100 und der genrespezifischen Top-Songs wurde eine Anfrage an die Genius-API gestellt, um den vollständigen Songtext sowie weitere Metadaten abzurufen. Daraus resultierte folgende Datenvollständigkeit im Bezug auf die jeweiligen Top-Listen, wie in Tabelle 2 zu sehen ist:

| Genre | Hot 100 | pop | rock | country | rap | r&b |
|-----------------|---------|--------|--------|---------|--------|--------|
| Songanzahl | 7399 | 2960 | 3700 | 4820 | 2265 | 5280 |
| Lyrics | 7165 | 2880 | 3693 | 4673 | 2164 | 5106 |
| Vollständigkeit | 96.84% | 97.30% | 99.81% | 96.95% | 95.54% | 96.70% |

Tabelle 2: Tabelle zeigt die Vollständigkeit der Songtexte in den Top-Listen der Billboard Hot 100 und genrespezifischen Charts basierend auf der Anzahl der Songs mit verfügbaren Songtexten. Die Vollständigkeit gibt den Prozentsatz der Songs an, für die Songtexte über die Genius API abrufbar waren.

4.2 Auswahl der Komplexitätsmetriken

Für die Analyse der Komplexität von Songtexten wurden mehrere Metriken in Betracht gezogen, um Vielfalt, Informationsdichte und Vorhersagbarkeit der verwendeten Wörter zu messen. Insgesamt wurden fünf Metriken evaluiert, von denen drei ausgewählt wurden.

4.2.1 Type-Token-Ratio (TTR)

Die Type-Token-Ratio (TTR) misst die lexikalische Vielfalt eines Textes durch das Verhältnis zwischen der Anzahl der einzigartigen Wörter (Types) und der Gesamtanzahl der Wörter (Tokens). Ein hoher TTR-Wert deutet auf eine große Wortvielfalt hin, während ein niedriger Wert häufige Wiederholungen anzeigt.

$$\text{TTR} = \frac{\text{Anzahl der eindeutigen Wörter (Types)}}{\text{Gesamtanzahl der Wörter (Tokens)}}$$

Diese Metrik wurde aufgrund ihrer Einfachheit und Aussagekraft als ein Hauptmaß für die Analyse der Wortvielfalt in Songtexten ausgewählt. TTR bietet einen direkten Einblick in die sprachliche Vielfalt, die einen zentralen Aspekt der Textkomplexität darstellt. Die Wahl dieser Metrik basiert auf ihrer Fähigkeit, Unterschiede in der Verwendung von Wörtern klar darzustellen und so relevante Aussagen zur Komplexität der Songtexte zu ermöglichen.

4.2.2 Shannon-Index (Entropie)

Der Shannon-Index misst die Unvorhersehbarkeit und Gleichmäßigkeit der Wortverteilung in einem Text. Ein höherer Shannon-Wert deutet auf eine gleichmäßige Verteilung der Wörter hin, während ein niedriger Wert darauf hinweist, dass bestimmte Wörter im Text dominieren.

$$H = - \sum p_i \cdot \log_2(p_i)$$

wobei p_i die Wahrscheinlichkeit des Auftretens jedes Wortes ist.

Im Vergleich zur TTR berücksichtigt der Shannon-Index nicht nur die Anzahl der unterschiedlichen Wörter, sondern auch deren Verteilungshäufigkeit. Dadurch bietet er eine tiefere Analyse der Textstruktur, indem er aufzeigt, ob die Wortverteilung ausgeglichen ist oder von wenigen Wörtern dominiert wird. Diese zusätzliche Dimension der Komplexitätsmessung, die sowohl die Vielfalt als auch die Gleichmäßigkeit erfasst, macht den Shannon-Index zu einer wichtigen Ergänzung der Analyse.

4.2.3 Automated Readability Index (ARI)

Der Automated Readability Index (ARI) bewertet die Lesbarkeit eines Textes basierend auf der Anzahl der Wörter, Zeichen und Sätze. Für die Berechnung wurden Satzzeichen wie Punkte, Fragezeichen und Ausrufezeichen als Satzgrenzen definiert, um stabile und vergleichbare Ergebnisse zu gewährleisten.

$$\text{ARI} = 4.71 \cdot \left(\frac{\text{Anzahl der Zeichen}}{\text{Anzahl der Wörter}} \right) + 0.5 \cdot \left(\frac{\text{Anzahl der Wörter}}{\text{Anzahl der Sätze}} \right) - 21.43$$

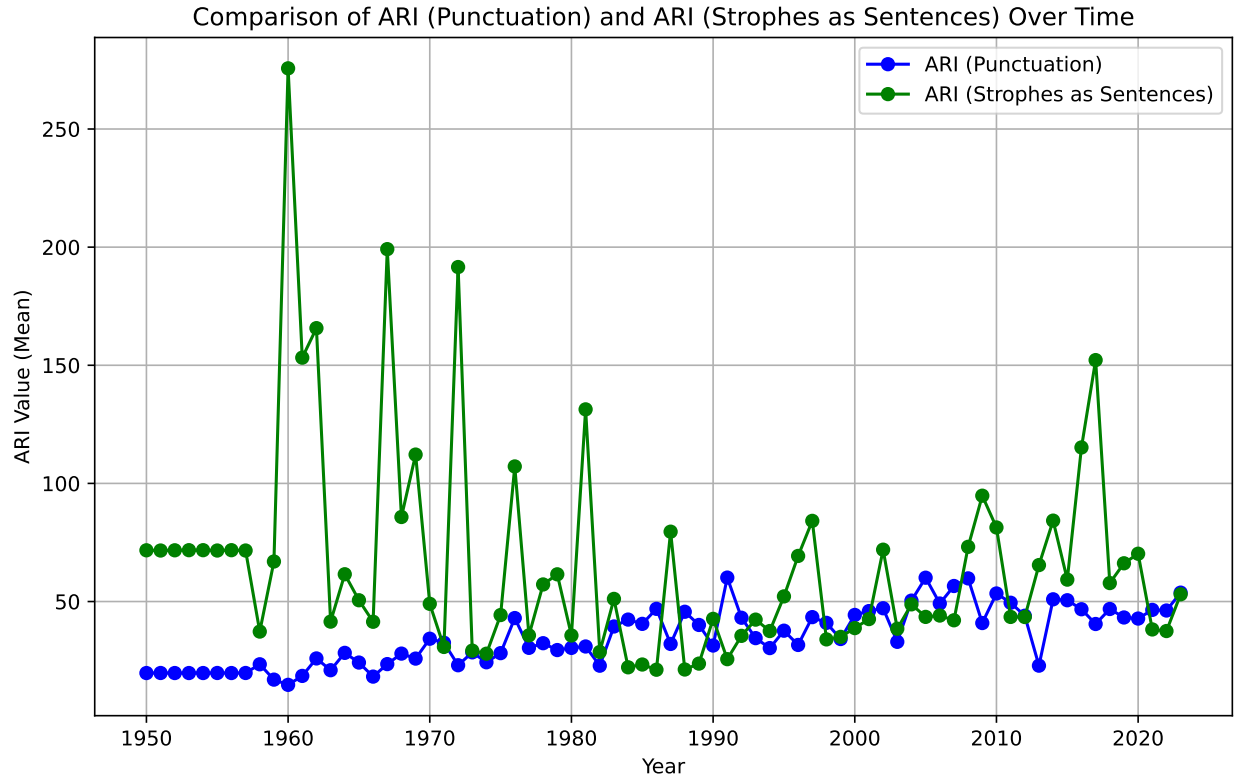


Abbildung 2: Vergleich der durchschnittlichen ARI-Werte pro Jahr basierend auf Zeichensetzung (ARI) und Strophen als Sätze (ARI Strophen). Der Plot zeigt die Entwicklung beider Berechnungsmethoden zur Lesbarkeitsbewertung über den Zeitverlauf.

Diese Methode wurde gegenüber einer alternativen Variante bevorzugt, die Strophen als Satzgrenzen nutzte, da die ungleichmäßige Struktur vieler Songs (z. B. kurze Zeilen oder häufige Strophenwechsel) in dieser Version zu starken Schwankungen führte (siehe Abb. 2). Die Standardmethode erwies sich hingegen als konsistenter und besser geeignet für die Analyse von Songtexten.

Der ARI liefert eine wichtige Perspektive auf die strukturelle Komplexität der Texte und ergänzt die Metriken zur Messung der lexikalischen Vielfalt und Informationsdichte. Er wurde, aufgrund seiner Fähigkeit die Lesbarkeit präzise zu erfassen, als eine der Hauptmetriken ausgewählt.

4.2.4 Chao1- und Chao2-Estimatoren

Die Chao1- und Chao2-Estimatoren schätzen die versteckte Vielfalt eines Textes, indem sie seltene Wörter und Wortkombinationen besonders berücksichtigen. Chao1 fokussiert sich auf die Anzahl seltener Wörter, während Chao2 zusätzlich deren Kombinationen einbezieht.

Beide Metriken basieren auf Schätzungen und bieten theoretische Einblicke in die potenzielle Vielfalt eines erweiterten Textes. Da der Schwerpunkt dieser Analyse jedoch auf der tatsächlichen Verteilung und Nutzung der Wörter liegt, wurden Chao1 und Chao2 zugunsten direkterer Metriken wie TTR, Shannon-Index und ARI verworfen, die eine präzisere Analyse der realen Textkomplexität ermöglichen.

4.2.5 Zusammenfassung der Metrikauswahl

Die Kombination von TTR, Shannon-Index und ARI wurde gewählt, um eine umfassende Analyse der Songtext-Komplexität zu ermöglichen. TTR misst die Wortvielfalt, der Shannon-Index die Verteilung der

Wörter, und ARI bewertet die Lesbarkeit der Texte. Diese drei Metriken bieten zusammen eine ausgewogene Betrachtung der sprachlichen und strukturellen Komplexität.

5 Auswertung der Ergebnisse

Hier werden die Ergebnisse der Zeitreihenanalyse zusammengefasst. Dabei wurden die durchschnittlichen Werte der Metriken jährlich aggregiert und genrespezifisch verglichen, um festzustellen, ob die Komplexität der Songtexte abgenommen hat und wie die Genres die Billboard Hot 100 beeinflusst haben und ob dieser Einfluss auch umgekehrt stattgefunden hat.

5.1 Metriken der Billboard Hot 100

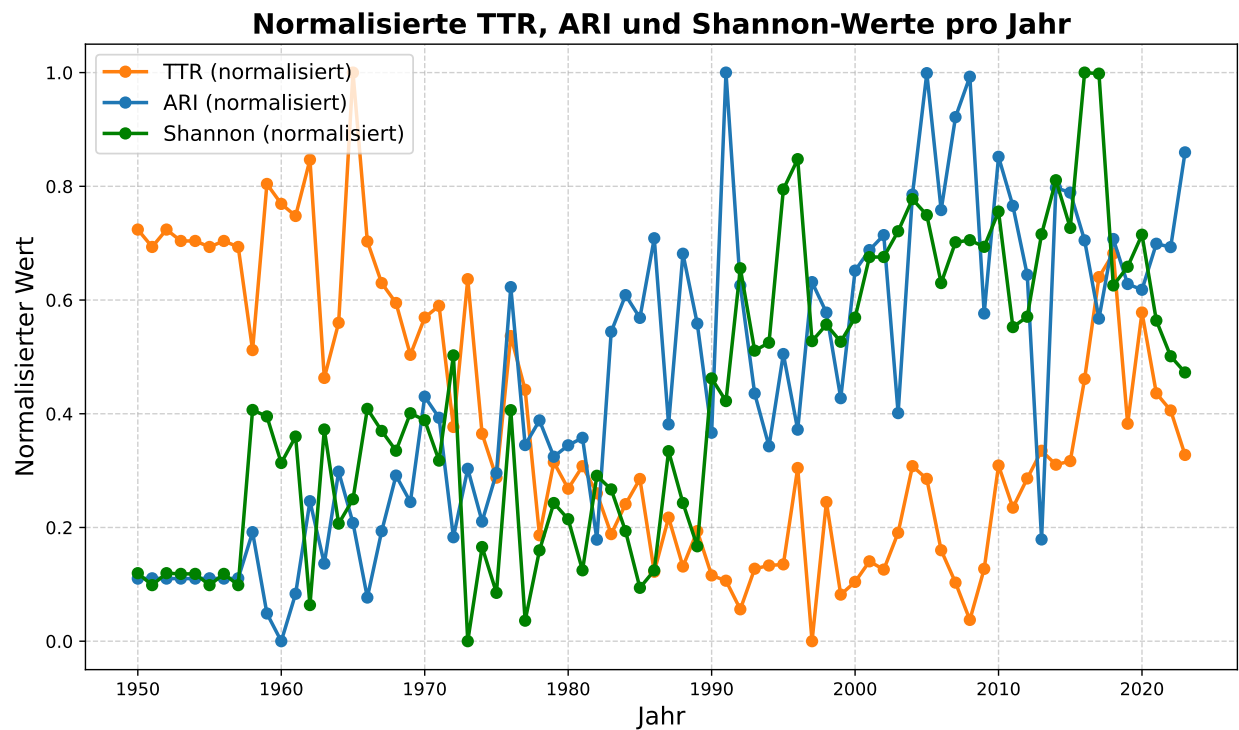


Abbildung 3: Die Grafik zeigt die Entwicklung der normalisierten Werte von TTR (orange), ARI (blau) und Shannon-Index (grün) über die Jahre hinweg, basierend auf den “Billboard Hot 100”. Abgebildet sind die lexikalische Vielfalt (TTR), die Lesbarkeit (ARI) und die Wortverteilung (Shannon) der Songtexte von 1950 bis 2023.

Die Analyse der TTR, welche die lexikalische Vielfalt in den Songtexten über die Jahre hinweg widerspiegelt, repräsentiert sich mit einer stabilen Entwicklung in den 1950er Jahren, was folgern lässt, dass sich die Verwendung der unterschiedlichen Wörter in der Zeit wenig verändert hat. Ab den 1960er Jahren ist jedoch ein deutlicher Anstieg zu verzeichnen, der seinen Höhepunkt um 1965 erreicht. Dies kann auf eine Phase kreativer Experimente hinweisen, in der abwechslungsreichere Wörter in den Songtexten verwendet wurden. Ab 1965 ist ein Rückgang der TTR zu erkennen, bis in die frühen 2000er Jahre. In dieser Phase wurden Songs potenziell immer weniger vielfältig. Der Rückgang des Vokabulars lässt auf eine Vereinfachung der Texte schließen. In den 2000er erlebt die TTR wieder einen Aufwärtstrend, was auf eine steigende lexikalische Vielfalt hinweist. Dieser Trend hielt aber nur bis 2017 an. Von da an begibt die TTR sich auf einen Abwärtstrend, den sie bis einschließlich 2023 beibehält, was auf einen erneuten Trend zu einem weniger umfangreichen Vokabular schließen lässt. Insgesamt ist die TTR seit 1950 gefallen.

Der Average Readability Index (ARI) beginnt in den frühen 1950er Jahren mit niedrigen Werten. Dies lässt

auf eine geringe strukturelle Komplexität in dieser Zeit deuten. Ein Tiefpunkt wird um das Jahr 1960 erreicht, was darauf schließen lässt, dass Songtexte aus dieser Zeit besonders leicht zu lesen waren. Ab den 1960er Jahren gab es aber einen Aufwärtstrend und man konnte einen Anstieg des ARI verfolgen. Dieser Anstieg der Textkomplexität wird fortgeführt bis einschließlich 1990. 1990 und zwischen 2000 und 2010 gab es vermehrt starke Spikes in dem ARI, was darauf hindeutet, dass die Texte mehr und mehr an Komplexität gewinnen. Diese Phase flachte aber schnell wieder ab, jedoch bleibt der ARI seit 2010 auf einem höheren Niveau, was darauf schließen lässt, dass die Komplexität der Texte im Vergleich zu den früheren Jahrzehnten angestiegen ist. Der Stand von 2023 liegt nahe an den historischen Höchstwerten und zeigt, dass die Songtexte über die Zeit hinweg eine insgesamt höhere Komplexität erreicht haben.

Der Shannon-Index zeigt eine ähnliche Entwicklung wie der ARI. In den 1950er Jahren bis 1970er Jahren steigt der Wert kontinuierlich an, was auf eine zunehmende Vielfalt in der Wortverteilung schließen lässt. Zwischen 1977 und 1983 fällt der Index jedoch in eine Tiefphase, was darauf schließen lässt, dass die Texte in dieser Zeit eine geringere Vielfalt aufweisen und von häufig wiederholten Wörtern dominiert wurden. Nach dieser Phase erholt sich der Index wieder und bleibt bis in die 2000er Jahre stabil. Seit 2020 zeigt sich allerdings ein Rückgang des Shannon-Index, was auf einen Rückgang der Vielfalt in der Wortverteilung hindeutet. Dieser Rückgang unterscheidet sich von der Entwicklung des ARI, der seit 2020 weiterhin ansteigt. Während der ARI auf eine zunehmende strukturelle Komplexität der Songtexte hinweist, deutet der sinkende Shannon-Index darauf hin, dass die Vielfalt der verwendeten Wörter abnimmt und bestimmte Wörter häufiger wiederholt werden.

5.2 Analyse der Genres im Vergleich zur Billboard Hot 100

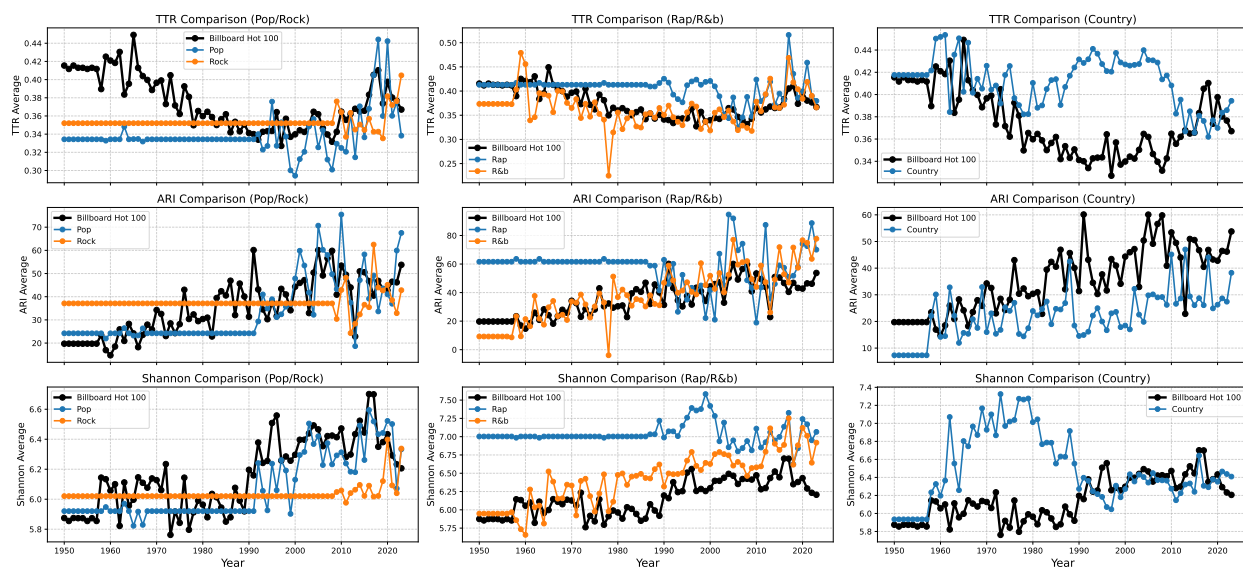


Abbildung 4: Die Grafik vergleicht die Entwicklung der durchschnittlichen Werte von TTR (orange), ARI (blau) und Shannon-Index (schwarz) über die Jahre hinweg für die Billboard Hot 100 Charts mit verschiedenen Musikgenres (Pop, Rock, Rap, R&B und Country). Sie zeigt, wie sich die lexikalische Vielfalt (TTR), die strukturelle Komplexität (ARI) und die Gleichmäßigkeit der Wortverteilung (Shannon-Index) in den Songtexten von 1950 bis 2023 entwickelt haben und wie diese Entwicklungen im Vergleich zu den genrespezifischen Top-Songs stehen.

In der folgenden Analyse werden die Unterschiede und Gemeinsamkeiten zwischen den einzelnen Genres im Vergleich zu den allgemeinen Trends der Billboard Hot 100 Songs untersucht. Der Fokus liegt dabei auf der Frage, ob und wie stark die spezifischen Trends der Genres die allgemeinen Entwicklungen in der lexikalischen Vielfalt (TTR), der strukturellen Komplexität (ARI) und der Wortverteilung (Shannon-Index) beeinflussen haben. Ziel ist es, zu untersuchen, ob die einzelnen Genres den Gesamtdurchschnitt geprägt oder ob sich signifikante Abweichungen von den Billboard Hot 100 Trends in den genrespezifischen Entwicklungen zeigen.

TTR

Pop-Songs lagen überwiegend unter dem TTR-Durchschnitt der Billboard 100, was auf ihren Beitrag zum Rückgang der lexikalischen Vielfalt deutet. Ab 2018 verzeichnet sie einen kurzzeitigen Anstieg über den Billboard-Durchschnitt, der sich bis 2023 wieder umkehrte. Pop folgte somit insgesamt dem Abwärtstrend des Billboard-TTR und trug zur negativen Entwicklung bei.

Rap startete mit einem höheren TTR als die Billboard-Charts und blieb bis 2005 über dem Durchschnitt. Danach näherten sich die Werte dem allgemeinen Billboard-TTR an, blieben aber zumeist darüber. Der sinkende TTR des Rap-Genres bis 2023 spiegelt den allgemeinen Abwärtstrend wider, was zeigt, dass Rap ähnlich wie Pop einen signifikanten Einfluss auf die allgemeine Entwicklung des Billboard-TTR hatte.

Rock zeigte ab 2008 zunächst TTR-Werte unter dem Billboard-Durchschnitt. Doch in den letzten Jahren, insbesondere ab 2023, verzeichnete Rock einen steigenden Trend und übertraf sowohl Pop als auch den allgemeinen Billboard-TTR. Rock hat sich somit entgegen dem allgemeinen Abwärtstrend entwickelt, was auf einen positiven Einfluss auf die lexikalische Vielfalt hinweist, der sich jedoch weniger stark auf die Billboard Hot 100 insgesamt auswirkte.

R&B begann mit einem höheren TTR als die Billboard-Charts, folgte jedoch bald den allgemeinen Auf- und Abwärtstrends der Billboard Hot 100. Diese parallele Entwicklung verdeutlicht, dass R&B einen starken Einfluss auf den Verlauf der Billboard-Charts hatte, da die Trends des Genres eng mit den allgemeinen Bewegungen der Billboard Hot 100 verknüpft waren und somit maßgeblich zu deren Entwicklung beitrugen.

Country verlief zunächst parallel zu den Billboard-Werten, bevor es in den 1970er Jahren einen eigenständigen, steigenden Trend entwickelte. Dieser positive Trend setzte sich bis in die 2020er Jahre fort, während der Billboard-TTR weiter fiel. Dies zeigt, dass Country sich zunehmend von den allgemeinen Billboard-Trends abkoppelte und daher weniger Einfluss auf deren Verlauf hatte.

ARI

Rap und R&B zeigen ebenfalls eine enge Verbindung zu den allgemeinen Trends der Billboard Hot 100, weichen jedoch häufig positiv ab. Rap zeigt dabei stärkere Abweichungen als R&B. Seit 2013 haben sich beide Genres zunehmend von den Billboard-Trends entfernt und zeigen einen positiven Verlauf, was darauf hindeutet, dass sie möglicherweise einen positiven Einfluss auf die allgemeine Entwicklung der Billboard Hot 100 hatten. Gegen 2023 zeigt Rap einen Abwärtstrend, während R&B weiterhin ansteigt. Beide Genres bleiben jedoch nahe beieinander, während die Billboard Hot 100 ebenfalls einen leichten Aufwärtstrend verzeichnet, allerdings mit etwas Abstand.

Country folgt grundsätzlich den Auf- und Abwärtstrends der Billboard Hot 100, bleibt jedoch durchgängig mit einem deutlichen negativen Abstand zurück. Dies deutet darauf hin, dass die Informationsdichte und Textkomplexität in Country-Songs zwar den allgemeinen Trends der Billboard Hot 100 folgen, aber deutlich weniger ausgeprägt sind.

SHANON

Seit etwa 1990 folgt der Shannon-Index von Pop dem Verlauf der Billboard Hot 100, bleibt jedoch stets leicht darunter, was auf eine geringere Wortvielfalt in Pop-Songs im Vergleich zur allgemeinen Musikindustrie hindeutet. Ab 2003 nähert sich der Index stärker an den der Hot 100 an. Im Jahr 2023 fällt der Shannon-Index sowohl für Pop als auch für die Hot 100, wobei der Rückgang bei Pop deutlicher ausfällt.

Rock lag über viele Jahre hinweg deutlich unter den Billboard Hot 100, verzeichnete jedoch 2023 einen Anstieg, bei dem der Shannon-Index sowohl Pop als auch die Hot 100 übertraf. Dies deutet auf eine gestiegene Wortvielfalt in Rock-Songs hin, die nun die allgemeinen Charts übertroffen hat.

Rap behielt über viele Jahre einen größeren Abstand zu den Hot 100 bei, wobei der Shannon-Index durchgängig höher war. Auch 2023 zeigt Rap einen steigenden Trend, was auf eine zunehmende Wortvielfalt hindeutet und möglicherweise positiv auf die Hot 100 gewirkt hat. R&B hingegen spiegelte lange Zeit die Trends der Hot 100 wider, was auf eine enge Verbindung zwischen R&B und den allgemeinen Entwicklungen hinweist. Ab 1980 begann sich jedoch ein Abstand aufzubauen, der bis 1990 andauerte. Danach verlief der R&B-Index wieder parallel zu den Hot 100, diesmal jedoch mit einem positiven Abstand.

Country folgte zunächst nahezu deckungsgleich den Hot 100, entfernte sich jedoch ab den 1970er Jahren deutlich. Ab 1990 näherten sich beide erneut an, und Country spiegelte danach fast vollständig die Trends der Hot 100 wider, was darauf hindeutet, dass die allgemeine Musikindustrie einen stärkeren Einfluss auf die Entwicklung von Country hatte als umgekehrt.

6 Fazit

Zusammenfassend lässt sich sagen, dass die Hypothese, die von einer Verschlechterung der Songtextkomplexität ausgeht, nicht bestätigt werden kann. Zwar hat die lexikalische Vielfalt (gemessen durch die TTR) im Laufe der Jahrzehnte abgenommen, was darauf hinweist, dass sich die Songtexte durch eine stärkere Wiederholung von Wörtern und eine eingeschränkte sprachliche Vielfalt auszeichnen. Allerdings zeigt der Shannon-Index, dass die Verteilung der Wörter gleichmäßiger geworden ist. Dies bedeutet, dass die wiederholten Wörter besser im Text verteilt sind, was auf eine gleichmäßigere Verteilung der Begriffe hindeutet. Parallel dazu hat die strukturelle Komplexität der Texte (gemessen durch den ARI) erheblich zugenommen. Die Songtexte sind über die Jahre hinweg insgesamt komplexer geworden.

Insgesamt lässt sich festhalten, dass die Songtexte zwar in Bezug auf die lexikalische Vielfalt etwas verloren haben, jedoch an Wortverteilung und struktureller Komplexität gewonnen haben. Die Qualität der Songtexte hat sich in Bezug auf ihre Komplexität somit nicht verschlechtert.

Im nächsten Schritt wird diese Analyse auf genrespezifische Entwicklungen ausgeweitet, um zu untersuchen, ob bestimmte Musikrichtungen ähnliche oder abweichende Trends aufweisen.

Die genrespezifische Analyse zeigt deutliche Unterschiede: Pop hat mit seiner geringeren lexikalischen Vielfalt die negative Tendenz der Billboard Hot 100 verstärkt. Rap und R&B hingegen weisen eine höhere Vielfalt und Komplexität auf und haben somit positiv auf die allgemeinen Trends eingewirkt. Rock, obwohl es sich in den letzten Jahren positiv entwickelt hat, zeigt aktuell wenig Einfluss auf die Billboard Hot 100. Country folgt zwar den Billboard-Trends, bleibt jedoch durchgängig weniger komplex.

Diese genrespezifischen Entwicklungen verdeutlichen, dass einige Genres, wie Rap und R&B, die Qualität der Songtexte positiv beeinflusst haben, während andere, wie Pop und Country, eher zur Vereinfachung beigetragen haben.

6.1 Einschränkungen

Eine der wichtigsten Einschränkungen dieser Arbeit lag bei den verfügbaren Daten. Die Informationen zu den einzelnen Genres waren nicht immer vollständig verfügbar, besonders in den frühen Jahrzehnten, als einige Musikrichtungen noch nicht so stark ausgeprägt waren oder schlecht dokumentiert wurden. Dadurch war es schwer, genau zu beurteilen, wie viel Einfluss die Genres in bestimmten Jahrzehnten hatten.

Gerade für Genres, die erst später populär wurden, fehlen in manchen Zeiträumen verlässliche Daten, was die Aussagekraft der Analyse in diesen Jahren beeinträchtigen könnte. Diese Unvollständigkeit sollte bei der Interpretation der genrespezifischen Ergebnisse immer mit bedacht werden.

6.2 Zukünftige Forschung

Eine vielversprechende Erweiterung der Forschung wäre die Analyse der Komplexität von Songtexten in anderen Sprachen, insbesondere in der deutschen Musik. Durch die Untersuchung, ob sich die lyrische Komplexität in deutschsprachigen Songs ähnlich wie in englischsprachigen Songs entwickelt hat, könnten wertvolle Einblicke in sprachliche und kulturelle Unterschiede in der Musik gewonnen werden.

Darüber hinaus könnte der Einfluss technologischer Entwicklungen, wie Streaming-Plattformen, AI-generierter Musik oder sozialer Medien, auf die lyrische Komplexität weiter erforscht werden. Diese Innovationen haben die Art und Weise, wie Musik produziert und konsumiert wird, grundlegend verändert und könnten die Vielfalt und Komplexität der Songtexte signifikant beeinflusst haben.

Referenzen

1. L. Reinhardt, [LyricComplexity: https://github.com/TheGim/LyricComplexity](https://github.com/TheGim/LyricComplexity). *GitHub repository*, (2024).
2. M. E. W. Varnum, J. A. Krems, C. Morris, A. S. Wormley, & I. Grossmann, Why are song lyrics becoming simpler? a time series analysis of lyrical complexity in six decades of American popular music. *PLoS ONE*, **16** (2021). <https://doi.org/10.1371/journal.pone.0244576>.
3. K. Choi & J. S. Downie, A Trend Analysis on Concreteness of Popular Song Lyrics. *Proceedings of the 6th International Conference on Digital Libraries for Musicology*, (2019). <https://doi.org/10.1145/3358664.3358673>.
4. [Billboard: https://www.billboard.com/](https://www.billboard.com/). *Website*, (Eingesehen am 14.08.2024).
5. [Genius: https://genius.com/](https://genius.com/). *Website*, (Eingesehen am 18.08.2024).
6. [MusicBrainz: https://musicbrainz.org/](https://musicbrainz.org/). *Website*, (Eingesehen am 20.08.2024).