

Digital Dylan – Computergestützte Analyse der Liedtexte von Bob Dylan (1962 – 2016)

Sippl, Colin

Colin.Sippl@stud.uni-regensburg.de
Lehrstuhl für Medieninformatik, Universität Regensburg

Fuchs, Florian

Florian.Fuchs@stud.uni-regensburg.de
Lehrstuhl für Medieninformatik, Universität Regensburg

Burghardt, Manuel

manuel.burghardt@ur.de
Lehrstuhl für Medieninformatik, Universität Regensburg

Hintergrund und Zielsetzung

Am 13. Oktober 2016 gab die Schwedische Akademie bekannt, dass sie den Nobelpreis in Literatur an Bob Dylan „für seine poetischen Neuschöpfungen in der großen amerikanischen Songtradition“ verleihen werde. Dass Dylan als Musiker und Songwriter den Literaturnobelpreis erhielt wurde mitunter sehr kontrovers diskutiert. Auf Kritik stieß z.B. die unzulässige Herauslösung von Dylans Texten aus der Musik und die Deutung seiner Lieder als Gedichte. Unbestritten ist nichtsdestotrotz Bob Dylans Rolle als einer der einflussreichsten Musiker des 20. Jahrhunderts.

Die (welt-)politischen Entwicklungen, die das Schaffen Dylans inspirierten, sind im Kontext seines Wirkens umfassend diskutiert worden, unter anderem in „Bob Dylan und die sechziger Jahre: Aufbruch und Abkehr“, und liefern noch immer Diskussionsstoff, wie etwa eine in jüngerer Zeit erschienene Arbeit von Taylor & Israelson (2015) über Dylans politische Einflüsse zeigt. Erfolgte die Beschäftigung mit Dylans Werk bislang allenfalls episodisch, so muss eine systematische Analyse des Gesamtwerks als Desiderat gelten, welches in gewisser Weise bereits von Bob Dylan selbst formuliert wurde:

All these people who say whatever it is I'm supposed to be doing – that's all gonna pass, because, obviously, I'm not gonna be around forever. That day's gonna come when there aren't gonna be any more records, and then people won't be able to say 'Well, this one's not as good as the last one.' **They're gonna have to look at it all (eigene Hervorhebung).** And I don't know what the picture will be, what people's judgement will be at that time. I can't help you in that area. – *Bob Dylan*

Dieser Beitrag erprobt, inwiefern mithilfe digitaler Methoden im Sinne des *Distant Reading*-Paradigmas ein neuer Zugang zu Dylans Gesamtwerk geschaffen werden kann. Bezugnehmend auf das Konferenzmotto einer „Kritik der Digitalen Vernunft“ soll untersucht werden, wo die Grenzen und Möglichkeiten eines solchen digitalen Analyseansatzes liegen, indem überprüft wird, ob sich bestehende qualitative Einteilungen von Dylans Werk in unterschiedliche Schaffensperioden auch anhand statistisch signifikanter Wörter (Rayson, Berridge & Francis 2004) und N-Gramme (Evert 2005) belegen lassen.

Stand der Forschung

Bereits vor der Auszeichnung Dylans mit dem Nobelpreis in Literatur, waren seine Texte Gegenstand wissenschaftlicher Betrachtungen im Sinne des *Close Reading* (vgl. etwa Brown 2014). Brown unterteilt Dylans Werk in einzelne Phasen und verknüpft diese jeweils mit allgemeinen, zeitgeschichtlichen Ereignissen sowie biographischen Meilensteinen des Künstlers. Taylor & Israelson (2015) gehen einen ähnlichen Weg, versuchen jedoch Dylans Werk abseits verbreiteter politischer Einordnungen zu betrachten. Etwas anders ausgerichtet ist die Arbeit von Wissolik et al. (1994): Hierbei handelt es sich um eine Art Wörterbuch, in dem Namen und Gegenstände, die in Dylans Texten auftauchen, erläutert werden.

Eine umfassende Untersuchung Dylans Werks mithilfe computerbasierter Methoden fand sich bis zum Abfassungszeitpunkt des vorliegenden Texts nicht. Allerdings sind quantitative Verfahren zur stilistischen und inhaltlichen Analyse von Liedtexten in den Digital Humanities durchaus verbreitet. So beschreiben etwa, wie mithilfe von N-Gramm-Modellen ein Liedtext-Korpus anhand der Merkmale Textlänge, Textstruktur, Wortschatz und Semantik analysiert werden kann, um eine automatisierte Genrezuordnung vornehmen zu können. Daneben existieren stilometrische Untersuchungen von Liedern oder Gedichten, die sich mit der Berechnung von autoren- und genrespezifischen Merkmalen befassen, wie z.B. Suzuki & Hosoya (2014), die japanische Pop-Songs analysieren.

Forschungsmethodik

Bezugsrahmen der Analyse: Schaffensphasen Bob Dylans

Den analytischen Bezugsrahmen dieser Studie stellt die phasenweise Einteilung von Dylans Schaffen nach Brown (2014) dar. Brown unterscheidet dabei neun unterschiedliche Phasen, die mit „Becoming Bob Dylan“ (1960-1964) beginnen und vorläufig mit „Bob Dylan Revisited“ (2000-2012) enden. Diese Stilphasen

umfassen z.B. Dylans Hinwendung zum Christentum oder seine elektronische „Folk Rock“-Phase (vgl. Brown, 2014).

Korpus und Datenaufbereitung

In dieser Arbeit wurde ein Korpus bestehend aus 452 Liedtexten mit einem Umfang von 133.045 Tokens untersucht, die Bob Dylan zwischen den Jahren 1962 und 2016 auf Studio-Alben veröffentlicht hat. Die Liedtexte und Metainformationen wie etwa Titel, Album und Jahr stammen von der Plattform *LyricsWikia*¹. Da es sich bei *LyricsWikia* um ein community-gestütztes Projekt handelt, erfolgte vorab ein stichprobenartiger Abgleich einzelner Lieder mit den offiziellen Texten nach, wobei keinerlei Abweichungen festgestellt werden konnten.

Das Korpus wurde weiterhin mit Methoden der Computerlinguistik aufbereitet, insbesondere unter Verwendung des *Python Natural Language Toolkits* (NLTK²). Die Verarbeitung des Korpus umfasst die grundlegende Lemmatisierung mit dem *WordNet-Lemmatizer* (Teil des NLTK) und eine Stoppwortbereinigung (NLTK-Stoppwortliste für Englisch mit eigener Erweiterung³) sowie die Wortartenannotation mithilfe des *Stanford Log-linear Part-of-Speech-Taggers*. Da Dylan in seinen Texten häufig umgangssprachliche Formulierungen wie etwa verkürzte Grundformen (bspw. „savin“, „swimmin“) verwendet, wurde für den POS-Tagger ein Modell verwendet, welches auf der Grundlage von Twitter-Texten trainiert wurde und gute Ergebnisse für Texte mit nicht-standardisiertem Vokabular und Slang liefert.

Korpusvergleich – Assoziationsmaße und Referenzkorpus

Assoziationsmaße

Ein etabliertes Verfahren, um aus einem Korpus spezifische Wörter zu extrahieren, ist ein direkter Korpusvergleich mit dem *Log-Likelihood-Test*, der sich zum Vergleich von Korpora unterschiedlicher Größe besonders eignet (Rayson, Berridge, & Francis, 2004). Damit können Wörter, die im untersuchten Korpus mit einem signifikanten Frequenzunterschied zum Referenzkorpus auftreten, als Schlagworte betrachtet werden. Dies kann besonders aussagekräftige Ergebnisse liefern, wenn zusätzlich eine POS-Filterung erfolgt, womit sich beispielsweise signifikante Nomen oder Verben eines Korpus berechnen lassen. Darüber hinaus wurde eine Berechnung von N-Grammen in Form von Bi- und Trigrammen umgesetzt. Die berechneten N-Gramme lassen sich in der Web-App unter der Wahl eines Assoziationsmaßes, wie dem *Chi Quadrat-Test*, dem *Jaccard-Test*, dem *Poisson-Stirling-Test*, dem *Likelihood Ratio-Test* sowie dem *Pointwise Mutual*

Information-Test anzeigen. Dabei liefert jedes Verfahren zur N-Gramm-Berechnung eigene spezifische Ergebnisse. Dieser Freiraum wird ganz bewusst erhalten, um die verschiedenen Facetten eines Texts, die ein Assoziationsmaße jeweils anzeigt, für die spätere Datenanalyse nutzen zu können.

Referenzkorpus

Als Referenzkorpus dient das mündliche Subkorpus des *Open American National Corpus* (OANC; American National Corpus Project), welches insgesamt 3.862.172 Tokens umfasst. Das Korpus enthält viele Belege aus der mündlichen Kommunikation und eignet sich dadurch in besonderer Weise als Vergleichskorpus für Dylans Texte, die wie bereits beschrieben einen hohen Anteil umgangssprachlicher Formulierungen und Slang-Ausdrücke enthalten.

Beim Korpusvergleich kann entweder das gesamte Dylan-Korpus mit dem Referenzkorpus verglichen werden, oder mit den jeweiligen Dylan-Subkorpora, also bspw. all seinen Texten aus den 1970er-Jahren oder aus der ersten Schaffensperiode „*Becoming Bob Dylan*“ (1960-1964). Ein Vergleich der einzelnen Dylan-Subkorpora zum Gesamtwerk ist ebenso möglich. Letztere Option wird z.B. genutzt, um anhand jeweils signifikanter Wörter die einzelnen Schaffensperioden nach Brown (2014) zu überprüfen und damit die grundsätzliche Eignung solch quantitativer Verfahren zur Identifikation thematischer Verschiebungen zu untersuchen. Die Ergebnisse dieses Korpusvergleichs sind, zusammen mit allen anderen Ergebnissen der angewandten Analyseverfahren, in einer interaktiven Webanwendung über unterschiedliche Visualisierungen (Balkendiagramm, *treemap*, *wordcloud*, *Tabelle*) für weitere Interpretationen zugänglich⁴. Wie schon bei den Assoziationsmaßen, so gilt auch hier, dass jede Visualisierungsform eine bestimmte Perspektive auf die Berechnungsergebnisse eröffnet.

Ergebnisse

Im direkten Vergleich des gesamten Dylan-Korpus (1962-2016) mit dem OANC-Referenzkorpus treten einige interessante, signifikant-häufige Wörter im Werk Dylans hervor. Die von Bob Dylan verwendeten Adjektive erzeugen in der Gesamtschau tendenziell eher eine bedrückende Stimmung (*blind*, *weary*, *lonely*, *drunken*, *scared*, *restless*, *ragged*). Bei den Substantiven mischen sich unter viele Personen- und Ortsnamen auch religiöse Begriffe (*soul*, *heaven*, *devil*, *eden*, *prayer*, *paradise*). Viele der übrigen Begriffe sind erwartungsgemäß typisch für Folk-Musik (*levee*, *rooster*, *train*), was sich wiederum durch die Wahl des Referenzkorpus, das verschiedenartige mündliche Textquellen enthält, erklären lässt (Rayson, Berridge, & Francis, 2004: 8).

Die Analyse signifikant-häufiger Wörter für die einzelnen Schaffensphasen Dylans liefert Ergebnisse mit hoher Aussagekraft. So fällt etwa für die Phase „The Changing of the Guard“ (1978-1981), in der sich Dylan dem Christentum hinwendet, auf, dass das Vokabular tatsächlich viele christliche Motive aufweist (*lord, Jesus, devil, altar, faith, confession, grace, power, serve*). Insgesamt nimmt der Anteil an „düsterem“ Vokabular in dieser Phase ab, verschwindet jedoch nicht komplett (bspw. *shot, destruction*). Der Anteil an hoffnungsvollen Wörtern nimmt hingegen zu (bspw. *beginning, ready, arise, wake, thank*). Bei den übrigen Schaffensphasen fallen die Ergebnisse jedoch mitunter wesentlich weniger deutlich aus.

Ein differenziertes Bild ergibt sich für die N-Gramm-Analyse, was einerseits der Vielfalt an verfügbaren Methoden zur Berechnung und andererseits den unterschiedlichen N-Gramm-Längen geschuldet ist. Die Ergebnisse für Bigramme mit Hilfe des *Pointwise-Mutual-Information-Tests* (PMI) erschienen dabei am geeignetsten, um die thematischen Schwerpunkte von Dylans Schaffensphasen nach nachzuvollziehen. So findet das PMI-Verfahren im Subkorpus der Phase „The Changing of the Guard“ (1978-1981) Bigramme wie *close prayer, name lucifer, jesus good, jesus bone* oder *arise upon*, die eindeutig religiöse Bezüge in Dylans Texten dieser Phase veranschaulichen. Generell fällt jedoch die Dominanz von Refrain-Versen in den Liedern bedeutend ins Gewicht (z.B. *knock heaven door*), was die Qualität der Ergebnisse insbesondere bei den Trigrammen beeinflusst.

Diskussion

Im Sinne einer Kritik der Digitalen Vernunft bleibt demnach festzuhalten, dass sich Methoden der computergestützten Textanalyse und des statistischen Korpusvergleichs grundsätzlich dafür eignen, einen inhaltlichen Gesamtüberblick zu einem Liedtext-Korpus zu erhalten. Es können damit diachrone Entwicklungen des Wortschatzes und Verlagerungen thematischer Schwerpunkte als grobe Tendenzen aufgezeigt werden, um das Bild des Gesamtwerks zu ergänzen. Ein solcher Ansatz eignet sich demnach gut für die initiale Thesengenerierung und kann in gewisser Weise die Funktion eines Empfehlungs- bzw. Hinweissystems für erklärungsbedürftige Stellen⁵ in den Geisteswissenschaften übernehmen.

Die Identifikation konkreter Schaffensperioden, ausschließlich auf Basis signifikant häufiger Wörter ist aber – zumindest für das Werk Dylans – nicht ohne Weiteres erfassbar. Bei den N-Grammen zeigt sich, dass im Falle von Dylans Texten methodenübergreifend und mit zunehmender N-Gramm-Länge meist keine brauchbaren Ergebnisse erzielt werden konnten. Dies ist ein Hinweis darauf, dass die hier präsentierten Analysemethoden, die für andere Textsorten wie bspw. Parlamentsprotokolle bereits erfolgreich eingesetzt werden

konnten (vgl. Sippl et al. 2016), auf Liedtexte nur eingeschränkt anwendbar sind. Ein möglicher Kritikpunkt am hier beschriebenen Vorgehen mag zudem das verwendete OANC-Referenzkorpus sein, welches trotz hoher Anteile mündlicher Kommunikation doch nur beschränkt vergleichbar mit der Textsorte „Liedtext“ ist. Für künftige Vergleichsstudien böte sich ggf. ein Vergleich mehrerer unterschiedlicher Künstler und deren Liedtexte an, also bspw. Bob Dylan vs. Johnny Cash.

Fußnoten

1. <http://lyrics.wikia.com>, alle Hyperlinks dieses Dokuments wurden zuletzt abgerufen am 10.01.2018
2. Verfügbar unter <http://www.nltk.org/>
3. Filterung von Stoppwörtern, wie „hey“, „ah“, „yeah“, und Verkürzungen, wie „‘ve“, „‘s“ etc.
4. <https://www.colin-sippl.de/dylan> (Klick auf den Analyse-Button rechts oben)
5. Diesen Gedanken äußerte Hubertus Kohle auf der #DigiCampus-Tagung im Juni 2017 in München, vgl. <https://twitter.com/8urghardt/status/876725916487036928>.

Bibliographie

- American National Corpus Project** (2015a): *American National Corpus. Frequency Data*. <http://www.anc.org/data/anc-second-release/frequency-data/> [Letzter Zugriff 10. März 2017].
- American National Corpus Project** (2015b): *The Open American National Corpus (OANC)*. <http://www.anc.org/> [Letzter Zugriff 10. März 2017].
- Brown, Donald** (2014): *Bob Dylan: American troubadour*. Lanham, Md. [u.a.]: Rowman & Littlefield.
- Cott, Jonathan** (2006): *Bob Dylan, the essential interviews*. New York: Wenner Books.
- Derczynski, Leon et al.** (2013): "Twitter part-of-speech tagging for all: Overcoming sparse and noisy data", in: *Proceedings of the Recent Advances in Natural Language Processing* September, 198–206. http://www.derczynski.com/sheffield/papers/twitter_pos.pdf [Letzter Zugriff 9. März 2017].
- Dylan, Bob** (2016): *The lyrics: 1961-2012*. New York: Simon & Schuster.
- Evert, Stefan** (2005): "The Statistics of Word Cooccurrences, Word Pairs and Collocations", in: *Unpublished doctoral dissertation Institut für maschinelle Sprachverarbeitung Universität Stuttgart* 98: August 2004, 353. <http://en.scientificcommons.org/19948039> [Letzter Zugriff 3. März 2017].
- Fell, Michael / Sporleder, Caroline** (2014): "Lyrics-based Analysis and Classification of Music", in: *International Conference on Computational Linguistics* 25: 23–29, 620–631.

Geisel, Sieglinde (2016): *Bob Dylan - Literaturnobelpreisträger wider Willen*. Deutschlandradio Kultur. http://www.deutschlandradiokultur.de/bob-dylan-literaturnobelpreistraeger-wider-willen.1005.de.html?dram:article_id=373494 [Letzter Zugriff 7. März 2017].

Rayson, Paul / Garside, Roger (2000): "Comparing corpora using frequency profiling", in: *Proceedings of the workshop on Comparing Corpora* 1–6.

Rayson, Paul / Berridge, Damon / Francis, Brian (2004): "Extending the Cochran rule for the comparison of word frequencies between corpora", in: *JADT 2004: 7es Journées internationales d'Analyse statistique des Données Textuelles*: 1–12.

Schmidt, Mathias R. (1983): *Bob Dylan und die sechziger Jahre: Aufbruch und Abkehr*. Frankfurt am Main: Fischer Taschenbuch Verlag.

Sippl, Colin / Burghardt, Manuel / Wolff, Christian / Mielke, Bettina (2016): Korpusbasierte Analyse österreichischer Parlamentsreden. In: *Netzwerke: Tagungsband des 19. Int. Rechtsinformatik Symposions IRIS 2016: 25.- 7. Feb. 2016, Univ. Salzburg, S. 139-148*.

Suzuki, Takafumi / Hosoya, Mai (2014): "Computational Stylistic Analysis of Popular Songs of Japanese Female Singer-songwriters", in: *Digital Humanities Quarterly* 8: 1, .

Svenska Akademien (2016): *Der Nobelpreis in Literatur des Jahres 2016*.

Taylor, Jeff / Israelson, Chad (2015): *The Political World of Bob Dylan: Freedom and Justice, Power and Sin*. New York: Palgrave Macmillan.

Toutanova, Kristina / Klein, Dan / Manning, Christopher D (2003): "Feature-rich part-of-speech tagging with a cyclic dependency network", in: *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1 (NAACL '03)*, 252–259. <http://nlp.stanford.edu/~manning/papers/tagging.pdf> [Letzter Zugriff 3. März 2017].

Wissolik, Richard David / McGrath, Scott. / Colaianne, A. J. (1994): *Bob Dylan's words: a critical dictionary and commentary*. Greensburg, PA: Eadmer Press.