

Zeitschrift: Schweizerisches Jahrbuch für Wirtschafts- und Sozialgeschichte =
Annuaire Suisse d'histoire économique et sociale

Herausgeber: Schweizerische Gesellschaft für Wirtschafts- und Sozialgeschichte

Band: 33 (2019)

Artikel: Wörter statt Autos zählen? : Potenziale von "Culturomics" in der
Verkehrs- und Mobilitätsgeschichte

Autor: Haefeli, Ueli

DOI: <https://doi.org/10.5169/seals-842474>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

Download PDF: 01.06.2025

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Ueli Haefeli

Wörter statt Autos zählen?

Potenziale von «Culturomics» in der Verkehrs- und Mobilitätsgeschichte¹

Count words instead of cars?

Possibilities for 'culturomics' in transportation and mobility history

Based on Google's Ngrams and the example of traffic history, this paper examines the potential of new historiographic approaches to large databases. Hypotheses on the history of mobility are examined on the basis of the frequencies of key terms in the source corpus. As a preliminary conclusion – for historiography is certainly only in its infancy in this respect – the transport and mobility historical potential of Ngrams has been confirmed. It has been shown that this methodological approach can be used firstly to better support existing hypotheses, secondly to produce new, original hypotheses, and thirdly to partly test these using Ngrams. Fourthly, the simple form of visualization using Ngrams is important. However, the exemplary analyses also point to more or less fundamental challenges, which must be taken into account when using Ngrams. A distinction has to be made between problems that primarily concern Google's Ngrams and those that generally have to be taken into account in text mining in large databases.

Die grundlegenden Innovationen in der jüngeren Geschichte der Informations- und Kommunikationstechnologien der letzten Jahrzehnte stellen auch die Wissenschaft vor neue Herausforderungen. Ob damit paradigmatische Weichenstellungen im kuhn'schen Sinn verbunden waren, sei hier dahingestellt; einiges spricht jedoch dafür.²

1 Diese Publikation entstand im Rahmen des Nationalfondsprojekts «Mobilität im schweizerischen Bundesstaat. Ein neuer Blick auf die Verkehrsgeschichte der Schweiz nach 1848». Für wertvolle Hinweise danke ich meinen Kollegen Christian Rohr, Hans-Ulrich Schiedt, Markus Sieber und Benjamin Spielmann. Siehe www.hist.unibe.ch/forschung/forschungsprojekte/mobilitaet_im_schweizerischen_bundesstaat/index_ges.html (26. 7. 2016). Mein Dank gilt auch der Herausgeberschaft dieses Bandes und den beiden unbekannten Reviewern.

2 Thomas S. Kuhn, *The Structure of Scientific Revolutions*. Chicago 1962.

Sicher haben sich neue «Buzzwords» eingebürgert, was als untrügliches Zeichen für einen hohen Bedarf an Selbstvergewisserung innerhalb der Scientific Community verstanden werden kann. Und zweifellos ist «Big Data» ein solches Buzzword, das umschreiben soll, wie grosse, komplexe, schwach strukturierte und sich mit hoher Dynamik verändernde Datenbestände traditionelle Methoden der Datenanalyse überfordern und nach neuen Herangehensweisen rufen.³ Wie die Geisteswissenschaften ganz allgemein schwanken auch Historikerinnen und Historiker angesichts der neuen Datenberge zwischen spontaner Faszination und verschrecktem Zurückweichen.⁴ Die Mutigeren der Zunft versuchen die quellentechischen Potenziale möglichst stringent auszuschöpfen und haben dazu mit «Digital Humanities»⁵ ihr eigenes Buzzword kreiert, während die Vorsichtigeren gerne auf die methodischen Fallstricke und die ungelösten theoretischen Fragen hinweisen: so etwa auf die mangelnde Transparenz der Daten, aber auch auf die Gefahr unreflektierter, positivistischer und technizistischer Ansätze.⁶

Im vorliegenden Beitrag sollen ausgehend von Google Ngrams⁷ (Deutsch: N-Gramme) und am Beispiel der Verkehrsgeschichte die Potenziale neuer historiografischer Herangehensweisen an grosse Datenbestände ausgelotet werden, welche andernorts auch unter dem Stichwort «Culturomics» – verstanden als eine Art quantitativer Inhaltsanalyse – zusammengefasst worden sind.⁸

3 Vgl. etwa den Bericht zur Tagung «Datenspuren: Big Data im Kontext von Kultur und Gesellschaft» vom 3./4. Juli 2015 in Basel: www.infoclio.ch/de/node/138244. Vgl. auch die 7. Jahrestagung von infoclio.ch zum Nationalen Forschungsprogramm 75 «Big Data» vom 16. Oktober 2015, siehe www.infoclio.ch/de/node/138612 (26. 7. 2016). Eine methodische Übersicht bieten: Shawn Graham, Ian Milligan, Scott Milligan, *Exploring Big Historical Data. The Historians Macroscope*, Hackensack 2016.

4 Peter Haber, *Digital Past. Geschichtswissenschaft im digitalen Zeitalter*, München 2011.

5 Vgl. als erste Annäherung: Manfred Thaller, *Controversies around the Digital Humanities*, in: *Historical Social Research* 37/3, 2012, S. 7–23. Sowie: Matthew K. Gold (Hg.), *Debates in the Digital Humanities*, Minneapolis, 2012. Und: David M. Berry, *Understanding Digital Humanities*, Basingstoke, 2012. Eine aktuelle Einführung in methodische Aspekte geben Fotis Jannidis, Hubertus Kohle, Malte Rehbein (Hg.), *Digital Humanities – eine Einführung*, Berlin 2017. Auffallend ist auch die Absenz der Geschichtswissenschaft im NFP 75 «Big Data», siehe www.nfp75.ch/de (26. 7. 2016).

6 Vgl. dazu: Philipp Sarasin, *Sozialgeschichte vs. Foucault im Google Books Ngram Viewer. Ein alter Streitfall in einem neuen Tool*, in: Pascal Maeder, Barbara Lüthi, Thomas Mergel (Hg.), *Wozu noch Sozialgeschichte? Eine Disziplin im Umbruch*, Göttingen 2012, S. 152–174, hier S. 154–157. Für einen aktuellen Überblick über die Argumente: Chris Gratien, Daniel Pontillo, *Google Ngram. An Introduction for Historians HAZINE*, 2014, siehe www.academia.edu/7787608/GOOGLE_NGRAM_AN_INTRO_FOR_HISTORIANS (28. 7. 2016).

7 Google Books Ngram Viewer, <https://books.google.com/ngrams>; vgl. dazu auch LitRe-Wiki, Google Ngram Viewer, http://litre.uni-goettingen.de/index.php/Google_Ngram_Viewer, alle Seiten besucht am 15. 1. 2016.

8 Jean-Baptiste Michel et al., *Quantitative Analysis of Culture Using Millions of Digitized Books*, in: *Science* 331, 2011, S. 176–182; Kalev H. Leetaru, *Culturomics 2.0. Forecasting Large-Scale Human Behavior using Global News Media Tone in Time and Space*, in: *First Monday* 16 (9), 2011,

Das Beispiel der Verkehrsgeschichte eignet sich aus verschiedenen Gründen für N-Gramm-Analysen. Zunächst weil es sich um eine Thematik von herausragendem gesellschaftlichem und wirtschaftlichem Stellenwert handelt, welcher in vielen Diskursen der letzten beiden Jahrhunderte kontinuierlich eine wichtige Rolle gespielt hat und damit in der Literatur breit präsent sein dürfte. Darüber hinaus ist auch die historiografische Ausgangslage interessant: Quantitative Ansätze haben in der Verkehrsgeschichte seit jeher eine wichtige Rolle gespielt, sei es durch einfache Kennziffern wie Motorisierungsraten, Verkehrsleistungen, Unfallzahlen usw. oder sei es bei komplexeren, oft multivariaten Analysen wie etwa Fogels klassische Studie zur wirtschaftsgeschichtlichen Bedeutung der Eisenbahn in den USA⁹ oder in der Schweiz die Arbeiten von Schiedt und Frey zur Entwicklung der Erreichbarkeit.¹⁰ Das wird auch so bleiben. Neuere, eher kultur- und mobilitätsgeschichtlich orientierte Fragestellungen ermuntern jedoch dazu, das Methodenspektrum zu erweitern.¹¹ Im Beitrag soll darüber reflektiert werden, inwieweit Analysen der blossen Häufigkeit, aber auch der Kombinationen verschiedener Schlüsselbegriffe einer Mobilitätsanalyse in deutschsprachigen Texten des N-Gramm-Quellenkorpus qualitative Befunde ergänzen und vertiefen, aber auch korrigieren können. Mit anderen Worten, es gilt, die Tauglichkeit von N-Grammen zur Exploration, zur Hypothesengenerierung, aber auch zur Hypothesenprüfung zu diskutieren.

In einem ersten Teil des Beitrags wird das Instrument der N-Gramme mit seinen verschiedenen Recherchefunktionen vorgestellt und einige Überlegungen zur Plausibilität der zu erwartenden Ergebnisse und zu methodischen Herausforderungen angestellt. Daran anschliessend folgt eine Skizze von Fragestellungen der Verkehrs- und Mobilitätsgeschichte, welche mit N-Grammen oder selbstverständlich auch mit ähnlich aufbereiteten grossen Quellenkorpussen, angegangen werden könnten. Dies in der Überzeugung, dass auch im Bereich von «Big Data» die konkrete Fragestellung die Analyse leiten muss und nicht ein diffuses «Text Mining» in unerschöpflichen Datenarchiven. Der dritte Teil beinhaltet einige beispielhafte Anwendungen von N-Grammen für verkehrs- und mobilitätsgeschichtliche Ansätze. Im letzten Teil werden die Erkenntnisse diskutiert und Überlegungen zu weiterführenden Forschungsarbeiten angestellt.

siehe <http://firstmonday.org/article/view/3663/3040> (26. 7. 2016); Jean-Paul Delahaye, Nicolas Gauvrit, *Culturomics le numérique et la culture*, Paris 2013.

9 Robert W. Fogel, *Railways and American Economic Growth*, Baltimore 1964.

10 Thomas Frey, *Die Beschleunigung des Schweizer Verkehrssystems 1850–1910*, in: *Schweizerische Zeitschrift für Geschichte* 1, 2006, S. 38–45; Thomas Frey, Hans-Ulrich Schiedt, *Die internationale Erreichbarkeit von alpinen Schweizer Tourismuszentren 1850–1930 am Beispiel der Rigi*, in: David Gugerli et al. (Hg.), *Die Internationalität der Eisenbahn 1850–1970*, Zürich 2003, S. 219–235.

11 Gijs Mom, *The Crisis of Transport History. A Critique and a Vista*, in: *Mobility in History* 6/1, 2015, S. 7–19; vgl. auch: Mathieu Flonneau, Vincent Guigueno (Hg.), *De l'histoire des transports à l'histoire de la mobilité*, Rennes 2009.

Was sind Google Ngrams?

N-Gramme sind ein Konzept der quantitativen Linguistik.¹² Sie sind zu verstehen als das Ergebnis einer Zerlegung von Texten in Fragmente wie Buchstabengruppen, Wörter oder Wortkombinationen. Als 1-Gramm wird also beispielsweise «Europa» bezeichnet, als 2-Gramm «Apollo 11», als 5-Gramm «das sind genau fünf Worte» usw. Die Analysemöglichkeiten auf dieser Basis sind vielfältig, wobei in der Regel die Häufigkeit des Auftretens eines Wortes im Vordergrund steht. Besonders interessant sind solche N-Gramme aus naheliegenden Gründen bei der Analyse sehr grosser Datenbestände. Google Ngrams lassen sich mithilfe des frei zugänglichen Ngram Viewer auf recht einfache Art und Weise online generieren. Die zugrunde liegenden Quellenkorpusse können bei Bedarf und bei der entsprechenden Speicherkapazität aber auch lokal heruntergeladen werden. Der Ngram Viewer wurde nicht etwa von Geisteswissenschaftlern, sondern von Epidemiologen an der Harvard University entwickelt und 2010 öffentlich zugänglich gemacht.¹³ Als Quellenbasis dienen 5,2 Millionen, zwischen 1500 und 2008 erschienene, mit einer Texterkennung (OCR) gescannte und mit bibliografischen Metadaten versehene Bücher, womit immerhin 4% aller in diesem Zeitraum gedruckten Bücher (exklusive Zeitschriften) enthalten sein sollen. Der deutschsprachige Datensatz umfasst rund 483 000 Bücher mit etwa 37 Milliarden Wörtern beziehungsweise mit Leerzeichen abgegrenzten Wortketten.¹⁴ Die bekannten urheberrechtlichen Probleme der Google Library werden angesichts der Beschränkung auf einfache Wortketten nicht tangiert, indirekt beeinflussen diese den Forschungsprozess aber durchaus, weil der für eine verlässliche Quellenkritik oft notwendige Kontext der Wortverwendung im betroffenen Buch häufig nicht online rekonstruiert werden kann.

Der Ngram Viewer erlaubt über die reine Häufigkeit von Ngrams hinaus weitere vordefinierte Recherchemöglichkeiten. Mit Operatoren lassen sich beispielsweise mehrere Begriffe in einer Grafik vergleichen (+/–), ein Multiplikator für sehr unterschiedlich vorkommende Begriffe einbauen (*), oder das Verhältnis zwischen zwei Begriffen darstellen. Weiter ist es möglich, «Wildcards»¹⁵ zu setzen, es kann nach

12 Reinhard Köhler, Gabriel Altmann, Rajmund G. Piotrowski (Hg.), *Quantitative Linguistik – Quantitative Linguistics. Ein internationales Handbuch*, Berlin 2005. Siehe auch Haber (wie Anm. 4), S. 113–115.

13 Michel (wie Anm. 8).

14 Die Datenbasis, welche Google aus weltweit über 40 Bibliotheken zusammengetragen hat, umfasst sogar über 15 Mio. Bücher und damit etwa 12% aller in diesem Zeitraum verfassten Bücher. Für den Ngram Viewer wurde die Zahl aufgrund der Qualität der OCR-Daten und der Metadaten reduziert. Die N-Gramm-Bildung geschah über Satzgrenzen hinweg, aber nicht über Seitengrenzen. Es wurden nur Wörter aufgenommen, die in mindestens 40 Büchern vorkamen. Vgl. dazu: Michel (wie Anm. 8), S. 176.

15 Wird ein * anstelle eines Worts gesetzt, zeigt der Ngram Viewer die häufigsten zehn Ergänzungen.

Flexionen gesucht oder die Klein-/Grossschreibung ausgeschaltet werden usw. Es sind also recht komplexe Suchstrategien denkbar. Im Grundsatz kann der Ngram Viewer aber sehr rasch und ohne besondere Vorkenntnisse eingesetzt werden, eine Niederschwelligkeit, die manchmal von der Chance zum Risiko mutieren kann. Eigentlich lapidare methodische Herausforderungen können das Ergebnis einer Ngram-Suche enorm verzerren, ohne dass dies schnell offensichtlich wird. So ist zu bedenken, dass das scharfe «S» in der Schweiz nicht verwendet wird, dass Begriffe wie «Pendler» in der Einzahl gleich lauten wie im Plural, viele andere Begriffe, die man vielleicht vergleichen möchte, jedoch nicht. «Pendler» kann zudem auch Wüschelrutengänger bedeuten usw. Ganz zu schweigen von der Vieldeutigkeit von Begriffen wie «Verkehr» oder «Mobilität» und deren sich auch im Zeitverlauf wandelnden Bedeutung.¹⁶ Der Umgang mit dem Ngram Viewer bedarf also bei Lichte besehen doch einiger Umsicht.¹⁷

Grundsätzlicherer Natur ist die Kritik am Quellenkorpus, die von verschiedener Seite geäußert wurde: Letztlich gibt es keine bekannten Kriterien, welche die Auswahl der Bücher geleitet haben, sodass die Repräsentativität des Quellenkorpus trotz seiner Grösse nicht abschliessend belegt werden kann.¹⁸ Ein Stück weit lässt sich die Plausibilität der Ergebnisse allerdings mit Testauswertungen belegen. So müsste beispielsweise die Verwendung des Begriffs «Tschernobyl» nach der Reaktorkatastrophe von 1986 sprunghaft ansteigen, was sich – vergleiche Abb. 1 – auch bestätigt. Ähnlich plausible Resultate liefert auch eine Suche nach dem 2-Gramm «Berliner Mauer» (Abb. 2): Die zu erwartenden Peaks nach dem Mauerbau 1961 und dem späteren Fall der Mauer 1989 zeigen sich eindrücklich – gerade auch, wenn man bedenkt, dass die Tagespresse im Quellenkorpus nicht enthalten ist. Darüber hinaus gibt das zweite Beispiel auch bereits Hinweise auf das Potenzial solcher Auswertungen: Erstens zeigt sich, dass der Fall der Mauer mehr Resonanz gefunden hat als der Bau, und zweitens, dass beide Ereignisse lange nachgewirkt haben, denn nach den beiden Peaks sinken die Kurven viel langsamer ab als im Fall von Tschernobyl. Als vorläufiges Fazit bleibt meines Erachtens, dass es durchaus Sinn ergibt, die historiografischen Potenziale von N-Grammen vertieft zu untersuchen.¹⁹ Denn es

16 Vgl. dazu: David Bamman, Gregory Crane, Measuring Historical Word Sense Variation, in: *Proceedings of the 11th ACM/IEEE Joint Conference on Digital Libraries, 2011*, siehe www.perseus.tufts.edu/publications/bamman-11.pdf (17. 4. 2017).

17 Vgl. dazu aus sozialwissenschaftlicher Sicht: Brendan O'Connor, David Bamman, Noah A. Smith, Computational Text Analysis for Social Science. Model Assumptions and Complexity, NIPS Workshop on Computational Social Science and the Wisdom of Crowds, 2011, siehe <http://people.cs.umass.edu/~wallach/workshops/nips2011css/papers/OConnor.pdf> [17. 4. 2017]. Siehe auch Justin Grimmer, Brandon M. Stewart, Text as Data. The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts, in: *Political Analysis* 21/3, 2013, S. 267–297.

18 Vgl. dazu: Sarah Zhang, The Pitfalls of Using Google Ngram to Study Language, siehe www.wired.com/2015/10/pitfalls-of-studying-language-with-google-ngram (26. 7. 2016).

19 Zu einem ähnlichen Befund kam Philipp Sarasin bereits 2012 (wie Anm. 6).

gilt immer im Blick zu behalten, dass auch die überkommenen Quellenbestände in den Archiven und Bibliotheken keinesfalls zufällig zustande gekommen sind und nicht interessenfrei gesammelt wurden. Niemand wird behaupten wollen, dass sie ein unverzerrtes Bild der Vergangenheit anbieten.²⁰

Die Y-Achse zeigt den Prozentsatz der Nennungen im betreffenden Jahr, die X-Achse zeigt das Jahr.

Die verkehrs- und mobilitätsgeschichtlichen Potenziale der N-Gramme

Verkehrsgeschichte ist längst mehr als eine Ansammlung verstaubter Hagiografien über geniale Tüftler und tollkühne Rennfahrer, Piloten oder Astronauten. Über den Status eines unbedeutenden Randgebiets der Wirtschafts- und Technikgeschichte hinaus hat sie sich zu einer eigenständigen historiografischen Subdisziplin entwickelt.²¹ Der Schwerpunkt dieser Verkehrsgeschichte lag und liegt beim Verkehrsangebot, so insbesondere bei der Geschichte einzelner Verkehrsmittel und Verkehrsinfrastrukturen. Stärker von der Nachfrage nach Verkehr und Mobilität aus geht die sich seit gut einem Jahrzehnt etablierende Mobilitätsgeschichte, welche sich vom sozialwissenschaftlichen Feld der Mobility Studies und dem darin geprägten «New Mobilities Paradigm» inspirieren liess.²² Auch wenn die Abgrenzung zwischen verkehrs- und mobilitätsgeschichtlichen Ansätzen in der Praxis nicht immer ganz einfach ist, lassen sich doch einige Charakteristika einer Mobilitätsgeschichte herauschälen, basierend auf dem erwähnten erkenntnisleitenden Interesse für die Nachfrage: So interessiert sich die Mobilitätsgeschichte ganz besonders für schicht- und genderspezifische Unterschiede des Zugangs zum Verkehrssystem (Access) und sie strebt eine akteursbezogene Analyse von Mobilitätsverhalten mit seinen komplexen inter- und multimodalen Bezügen an.

Die neuere mobilitätsgeschichtliche Herangehensweise macht die traditionelle verkehrsgeschichtliche keineswegs obsolet, weshalb im Folgenden die Potenziale von N-Grammen für beide Ansätze an einigen Beispielen ausgeleuchtet werden. Eine verkehrsgeschichtliche Frage betrifft die Häufigkeit, in welcher die Begriffe «Auto»,

20 Dazu aus literaturwissenschaftlicher Sicht: Mark Algee-Hewitt et al., Canon/Archive. Large-scale Dynamics in the Literary Field, Pamphlets of the Stanford Literary Lab 11, 2016, siehe <https://litlab.stanford.edu/LiteraryLabPamphlet11.pdf> (28. 12. 2017).

21 Hans-Ulrich Schiedt, Swiss Traffic History. A Research Report, in: Gijs Mom, Gordon Pirie, Laurent Tissot (Hg.), Mobility in History. The State of the Art in the History of Transport, Traffic and Mobility, Neuchâtel 2009, S. 187–192; Hans-Ulrich Schiedt et al. (Hg.), Verkehrsgeschichte (Schweizerische Gesellschaft für Wirtschafts- und Sozialgeschichte – Société suisse d'histoire économique et sociale 25), Zürich 2010.

22 Mimi Sheller, John Urry, The New Mobilities Paradigm, in: Environment and Planning A, 38/2 (2006), S. 207–226.

«Eisenbahn» und «Flugzeug» (mit ihren verschiedenen Ableitungen) im Quellenkorpus auftauchen. Damit verbunden lässt sich folgende Hypothese formulieren: Begriffsgeschichten verlaufen anders als das mit diesen Begriffen verknüpfte «reale» Verkehrsgeschehen, und zwar in der Regel so, dass die begrifflichen Verfestigungen von neuen Mobilitätspraktiken vorausgehen und die «reale» Inszenierung im Sinne eines Handelns im Verkehrsgeschehen tendenziell nachgelagert erfolgt.²³

Welche Aussagen zu dieser Hypothese lassen sich aus dem N-Gramm zum deutschen Quellenkorpus ziehen (Abb. 3)? Zunächst zeigt sich, dass die Häufigkeit der Nennungen der Eisenbahn im 19. Jahrhundert rasch anschwillt und hohe Werte erreicht²⁴ und dann tatsächlich gemäss der Hypothese nach dem Höhepunkt um 1880 relativ rasch an Bedeutung verliert. Dies steht in einem ganz offensichtlichen Kontrast zum Verkehrsaufkommen der Bahnen, welches sich ganz anders entwickelte. In der Schweiz beispielsweise kamen pro Kopf der Bevölkerung 1870 auf dem Netz der Normalspurbahnen vier Fahrten, 1880 waren es dann acht. Von 19 Fahrten (1900) stieg die Zahl kontinuierlich weiter auf 57 (1950). Danach kam es im Zeichen der Massenmotorisierung zu einer Stagnation, ja sogar zu einem leichten Rückgang der Fahrten pro Kopf bis zu einem neuerlichen Anstieg in den ersten Jahren des 21. Jahrhunderts auf heute über 70 Fahrten.²⁵ Diese Zahlen können somit als erster deutlicher Beleg für die Hypothese dienen, wonach begriffliche Entwicklungen und das mit diesen Begriffen bezeichnete Verkehrsgeschehen keineswegs parallel verlaufen müssen. In einem nächsten Schritt – dazu fehlt hier der Raum – liessen sich die Erkenntnisse fruchtbar mit neueren wirtschaftsgeschichtlichen Ergebnissen zur Entwicklung der Eisenbahnen verknüpfen. So beispielsweise mit der Frage, zu welcher Zeit die Investitionen in die Eisenbahnen und danach der eigentliche Betrieb des neuen Verkehrsmittels welchen Beitrag zur Entwicklung der jeweiligen Volkswirtschaft geleistet haben.²⁶

Die Kurven für das «Automobil» und das «Flugzeug» belegen das bei den Eisenbahnen beobachtete Phänomen eines Time-Lags zwischen Begriffsgeschichte und

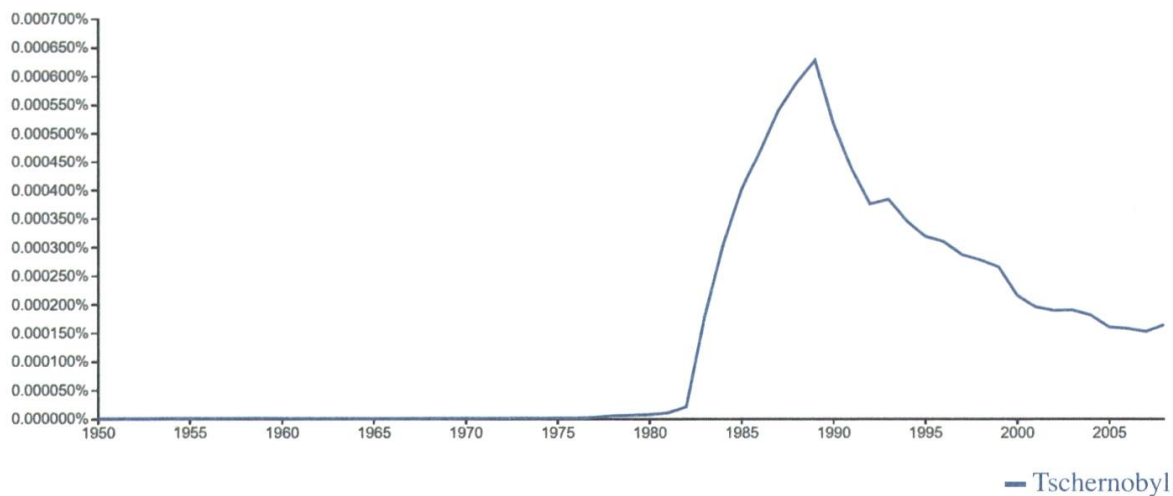
23 So ist beispielsweise Christoph Maria Merki in seiner grundlegenden Studie zum Aufstieg des Automobils in Frankreich, Deutschland und der Schweiz zum Schluss gekommen, dass die Motorisierung des Strassenverkehrs um 1930 abgeschlossen war: Christoph Maria Merki, *Der holprige Siegeszug des Automobils 1895–1930. Zur Motorisierung des Strassenverkehrs in Frankreich, Deutschland und der Schweiz*, Wien 2002; vgl. dazu auch: Gijs Mom, *Atlantic Automobilm. Emergence and Persistence of the Car, 1895–1940*, New York 2015. Mom setzt die Genese seines «Automobilismus» nicht nur für die USA für die Zeit bis 1940 an.

24 Der Wert von 0,01% heisst, dass in den 1880 publizierten und in den N-Grammen erfassten Publikationen jedes zehntausendste N-Gramm für die Begriffe Eisenbahn oder Eisenbahnen steht.

25 Quellen: Bis 1929: Datenbank Bahndaten von Viastoria, siehe <http://bahndaten.ch>, von 1921 bis 1985: Historische Statistik der Schweiz Online, siehe www.histecon.uzh.ch/de/hssso.html, danach bis 2014: Bundesamt für Statistik, Personenverkehr Eisenbahn, siehe www.bfs.admin.ch/bfs/portal/de/index/themen/11/01/new/nip_detail.html?gnpID=2015-406, alle Seiten besucht am 28. Juli 2016.

26 Vgl. als Überblick zu diesem Forschungsstrang: Alfonso Herranz-Loncan, *Railroad Impact in Backward Economies. Spain 1850–1913*, in: *The Journal of Economic History* 66/4, 2006, S. 853–881.

Abb. 1: N-Gramm zum Begriff «Tschernobyl» für 1950–2000 in deutschsprachigen Publikationen.



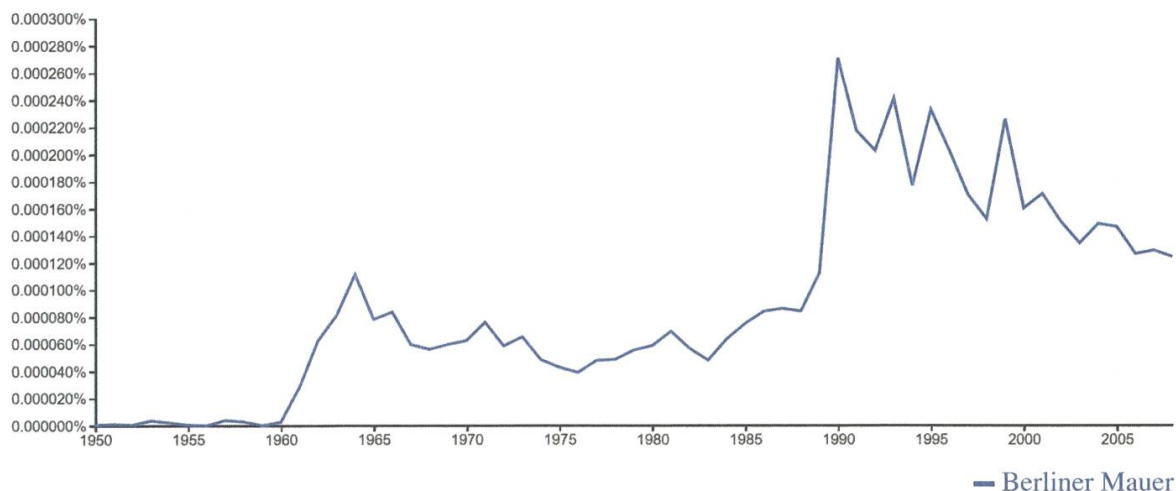
Quelle: <http://books.google.com/ngrams>.

Verkehrspraxis. Das N-Gramm bringt aber über die Hypothese hinaus einige weitere interessante Ergebnisse: So zeigt sich, dass den Begriffen «Automobil» und «Flugzeug» nie der überragende relative Stellenwert im Quellenkorpus zukam, den die «Eisenbahn» im 19. Jahrhundert hatte. Kann dies als Beleg dafür dienen, dass die Eisenbahn als erstes (zusammen mit dem hier ausgeblendeten Dampfschiff) motorisiertes und gegenüber dem Pferd deutlich schnelleres Fortbewegungsmittel das Mobilitätssystem stärker revolutioniert hat, als dies später beim Auto und beim Flugzeug der Fall war? Wird diese Frage abschliessend bejaht, wäre dies für die Verkehrs- und Mobilitätsgeschichte enorm bedeutsam.

Erstaunlich ist ein weiterer Befund, welcher auf das Potenzial von N-Grammen auf die Generierung von neuen Hypothesen hinweist, die auch mit anderen Mitteln als der N-Gramm-Analyse geprüft werden können: Der Begriff «Flugzeug» wird ganz offensichtlich zeitlich parallel mit dem Ersten und dem Zweiten Weltkrieg populär. Es wäre demnach interessant, zu untersuchen, inwieweit die militärische Verwendung von Flugzeugen, die mentalen Repräsentationen dieses Verkehrsmittels generell und vor allem im zivilen Bereich beeinflusst hat.

Selbstverständlich wären zur Erhärtung dieser Hypothese weitergehende Analysen notwendig. Trotzdem zeigt das Beispiel, dass mit N-Grammen anregende und durchaus originäre Beiträge zu verkehrswissenschaftlichen Fragen geleistet werden können. Wenden wir uns nun zwei mobilitätsgeschichtlichen Anwendungen von N-Grammen zu: Zunächst einmal soll nach der grundlegenden Unterscheidung zwischen den absoluten Schlüsselbegriffen «Verkehr» und «Mobilität» gefragt werden. Dazu werden die folgenden Hypothesen formuliert: Der Begriff «Mobilität» bürgert sich im deutschsprachigen Raum seit 1970 zunehmend als Ersatz für den Begriff «Ver-

Abb. 2: N-Gramm zum Begriff «Berliner Mauer» für 1950–2000 in deutschsprachigen Publikationen.



Quelle: <http://books.google.com/ngrams>.

kehr» ein (Hypothese 1). Das Aufkommen des Begriffs steht dabei in einem engen Zusammenhang mit der Umweltwende²⁷ um 1970. Er wurde zunächst vor allem von Personen und Institutionen verwendet, welche den negativen Begleiterscheinungen der Massenmotorisierung kritisch gegenüberstanden.²⁸ In den folgenden Jahrzehnten – so lautet die Hypothese 2 – hat sich die Begriffsverwendung von «Mobilität» ein Stück weit von der Konnotation Umweltschutz emanzipiert. Der Begriff verlor gleichzeitig an Eindeutigkeit und Klarheit: Mobilität gilt nun beispielsweise auch in der Autowerbung als attraktivere Alternative zum Begriff Verkehr.²⁹

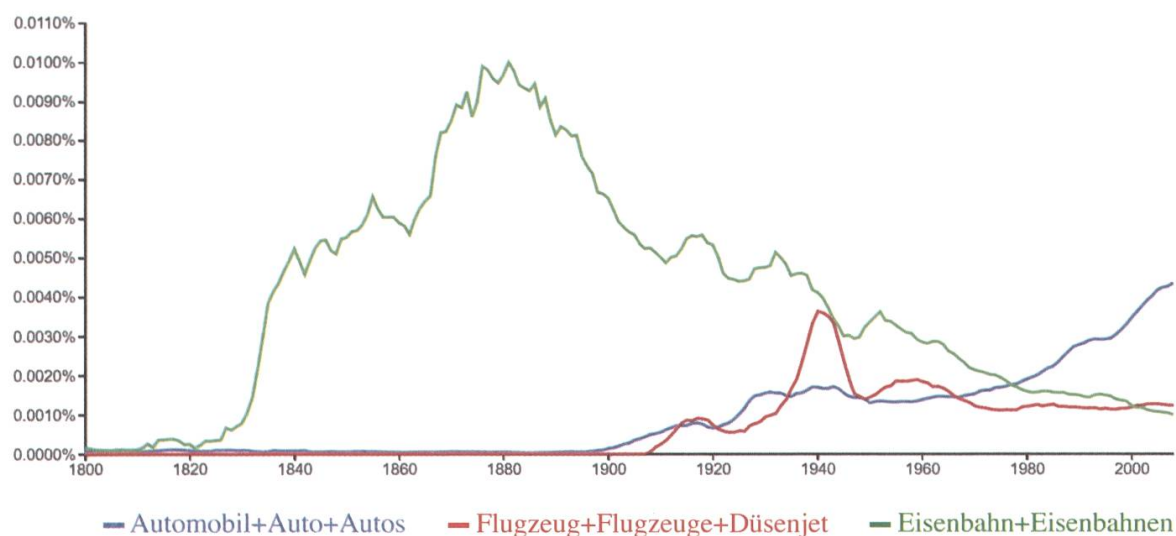
Wenden wir uns zunächst Hypothese 1 zu: Tatsächlich zeigt Abb. 4 eine stark gegenläufige Entwicklung der beiden Begriffe zwischen 1950 und 2000. Und tatsächlich nähern sich die beiden Kurven auch im Zeitraum 1965 bis 1975 am schnellsten an. Die Hypothese wird grundsätzlich gestützt, es gilt aber auch, wie oben bereits erwähnt, die Vielschichtigkeit dieser Begriffe im Auge zu behalten. «Mobilität» kann auch im Sinne von sozialem Auf- oder Abstieg verstanden werden oder als medi-

27 Patrick Kupper, Die «1970er Diagnose». Grundsätzliche Überlegungen zu einem Wendepunkt der Umweltgeschichte, in: Archiv für Sozialgeschichte 43, 2003, S. 325–348. Ueli Haefeli, Der lange Weg zum Umweltschutzgesetz. Die Antwort des politischen Systems auf das neue gesellschaftliche Leitbild «Umweltschutz», in: Allgemeine Geschichtsforschende Gesellschaft Schweiz (AGGS) (Hg.), 1798–1848–1998 Bd. 3, Zürich 1998, S. 241–249.

28 Vgl. ausführlicher dazu: Ueli Haefeli, The Notion of Mobility as a Product of the Environmental Turn of the 1970ies. The Discourses among Professional Actors in Germany and Switzerland, Fourth International Conference on the History of Transport, Traffic and Mobility, Paris September 2006, siehe www.academia.edu/15027404/The_notion_of_mobility_as_a_product_of_the_environmental_turn_of_the_1970s_A_paradigm_shift_among_professional_actors_in_Germany_and_Switzerland.

29 Peter Cerwenka, Mobilität kontra Verkehr? Aufklärung einer Dialektik zu ihrem zehnjährigen Bestehen, in: Der Nahverkehr 6, 2004, S. 28–30.

Abb. 3: N-Gramm zu den Begriffen «Automobil+Auto+Autos», «Eisenbahn+Eisenbahnen» sowie «Flugzeug+Flugzeuge+Düsenjet» für 1950–2008 in deutschsprachigen Publikationen (gleitender 3-Jahres-Durchschnitt).



Quelle: <http://books.google.com/ngrams>.

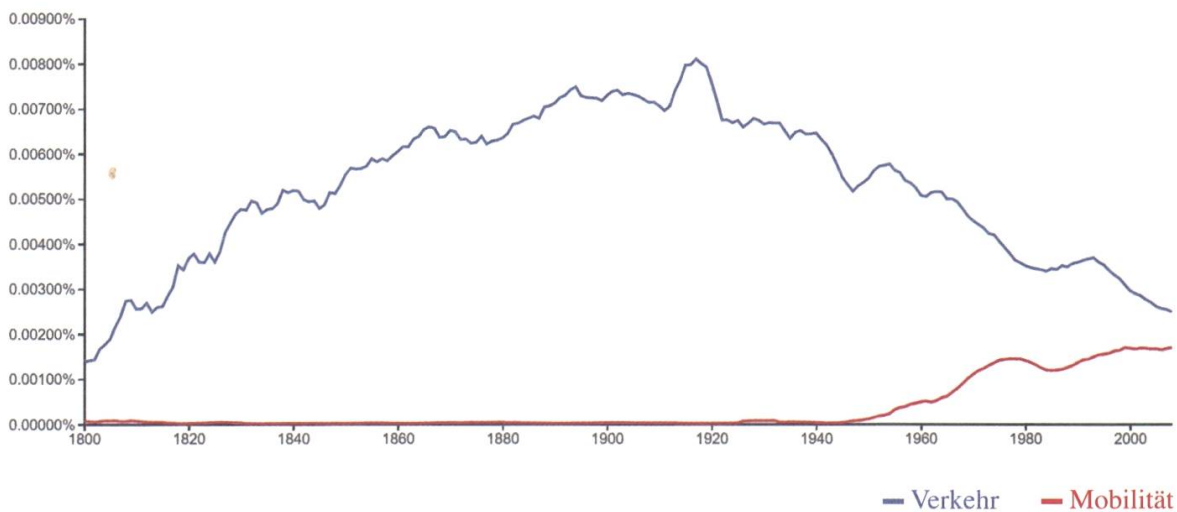
zinischer Begriff und «Verkehr» kann auch gesellschaftlichen Umgang bedeuten. Zudem ist davon auszugehen, dass sich das Verständnis beider Begriffe langsam und für Zeitgenossen fast unmerklich verschiebt; dies kann im Laufe der Jahrzehnte zu beträchtlichen semantischen Differenzen führen, welche sich in der Darstellungsform von Abb. 4 nicht nachweisen lassen.³⁰

Diesen Problemen lässt sich ein Stück weit Rechnung tragen: Der N-Gramm-Viewer ermöglicht es, verschiedene Begriffsverwendungen mit einer sogenannten Wildcard-Suche zu untersuchen. So lässt sich nach dem Wort suchen, welches dem Begriff Mobilität am häufigsten vorangeht. Die Abb. 5 zeigt für die Begriffskombination «soziale Mobilität» und deren Flexionen eine nach Mitte der 1970er-Jahre abnehmende Häufigkeit, während für die Kombination «räumliche Mobilität» genau das Gegenteil der Fall ist. Diese Aussage stützt die Hypothese 1 zweifellos weiter. Trotzdem zeigt sich auch hier, dass erst weitergehende, vor allem auch qualitative Untersuchungen im Quellenkorpus eine sicherere Beurteilung erlauben.

Erst recht gilt dies für die Hypothese 2. Auf den ersten Blick bestätigt sich die Hypothese voll und ganz. Die Kurven von «Mobilität» und «Umweltschutz» verlaufen bis Anfang der 1990er-Jahre einigermassen parallel und entkoppeln sich danach deutlich (Abb. 6). Wird aber berücksichtigt, dass nach der epochalen Konferenz der Vereinten Nationen über Umwelt und Entwicklung von 1992 in Rio de Janeiro die Begriffe «Nachhaltigkeit» und «nachhaltige Entwicklung» sehr häufig anstelle des Begriffs

30 Vgl. dazu: Bamman, Crane (wie Anm. 16).

Abb. 4: N-Gramm zu den Begriffen «Verkehr» und «Mobilität» für 1950–2008 in deutschsprachigen Publikationen (gleitender 3-Jahres-Durchschnitt).



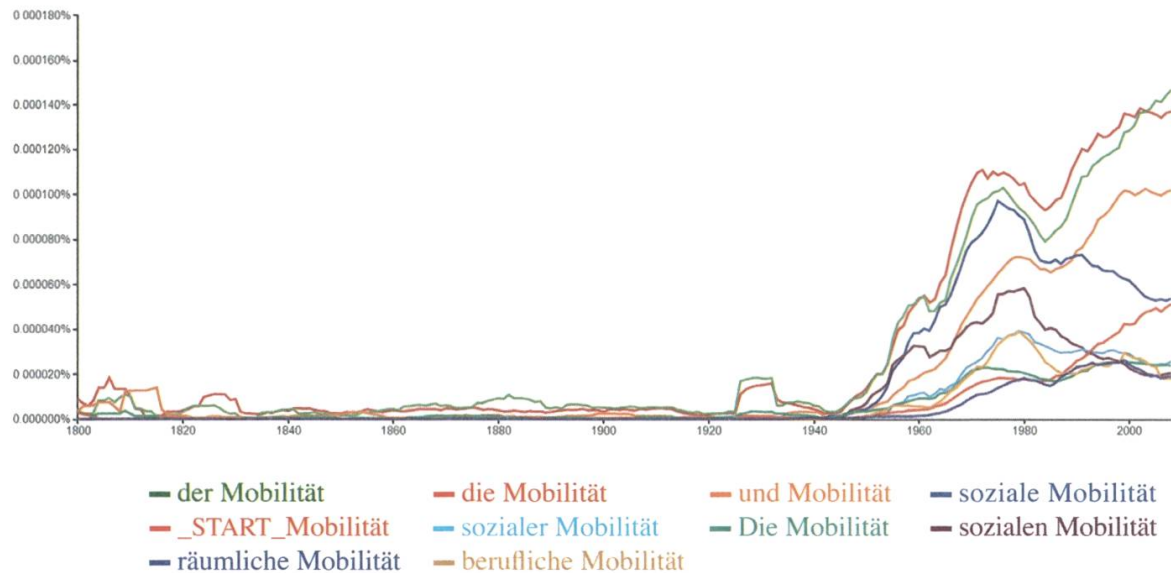
Quelle: <http://books.google.com/ngrams>.

«Umweltschutz» verwendet wurden, relativiert sich dieser Befund (Abb. 7.): Auch hier bleiben Fragen offen, die sich nicht mittels N-Grammen beantworten lassen. Beispielsweise liesse sich weiterführend fragen, ob nicht nur der Begriff «Mobilität» eine gewisse Beliebtheit erlangt hat, sondern auch der Begriff «Nachhaltigkeit», was dann wieder als Beleg für die Hypothese 2 zu verstehen wäre.

Als drittes und letztes Anwendungsbeispiel wollen wir uns nun der Frage von Geschwindigkeitsbegrenzungen – einem zentralen Diskursstrang der Verkehr- und Mobilitätspolitik – zuwenden. Es werden die vier Begriffe «Geschwindigkeitsbegrenzung», «Verkehrsberuhigung», «Tempolimit» und «Tempo 30» einander gegenübergestellt. Als Hypothese sei ähnlich wie beim Begriff Mobilität die Erwartung formuliert, dass sich die Verwendung dieser Begriffe nach der Umweltwende – die nun eben auch als Mobilitätswende verstanden werden kann – um 1970 grundsätzlich verändert hat. Bereits ein erster Blick auf das Ergebnis der N-Gramm-Analyse deutet auf eine eindruckliche Bestätigung dieser Hypothese hin: Einzig der Begriff «Geschwindigkeitsbegrenzung» taucht im Korpus vor 1970 auf (Abb. 8). Im Lauf der 1950er-Jahre wurde vermehrt über Geschwindigkeitsbegrenzung geschrieben. Dies kann niemanden überraschen, der sich mit der verkehrspolitischen Debatte um die Höchstgeschwindigkeit in dieser Zeit beschäftigt hat.³¹ Insbesondere der Verzicht auf eine Höchstgeschwindigkeit innerorts von 1953 in Deutschland hatte hohe Wellen geschlagen und wurde dann 1957 wieder

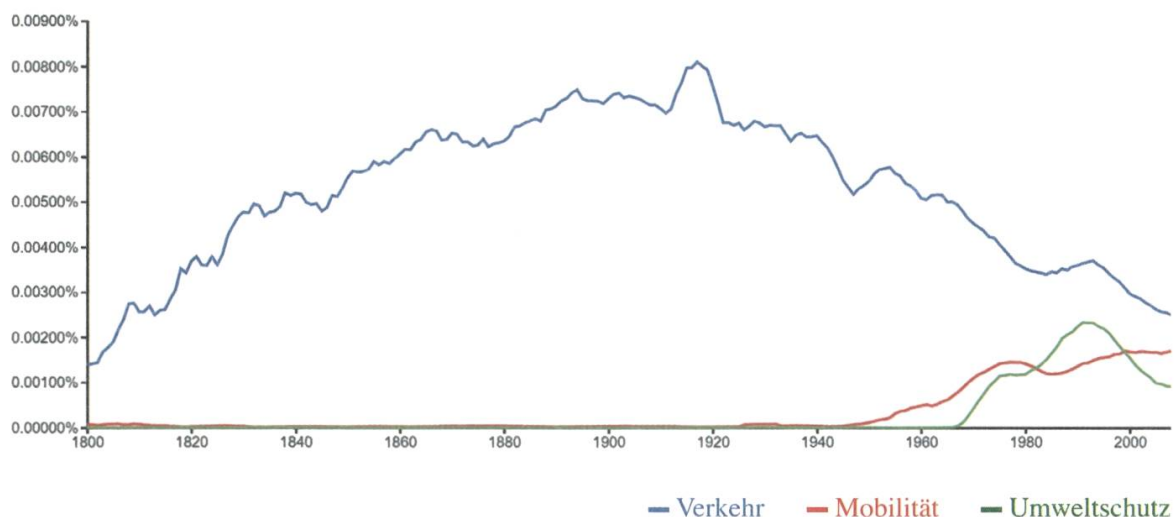
31 Dietmar Klenke, «Freier Stau für freie Bürger». Die Geschichte der bundesdeutschen Verkehrspolitik 1949–1994, Darmstadt 1994.

Abb. 5: N-Gramm-Wildcard-Suche «* Mobilität» (_START_Mobilität meint, dass der Begriff Mobilität nach einem Punkt oder am Anfang eines Absatzes steht) für 1950–2008 in deutschsprachigen Publikationen (gleitender 3-Jahres-Durchschnitt).



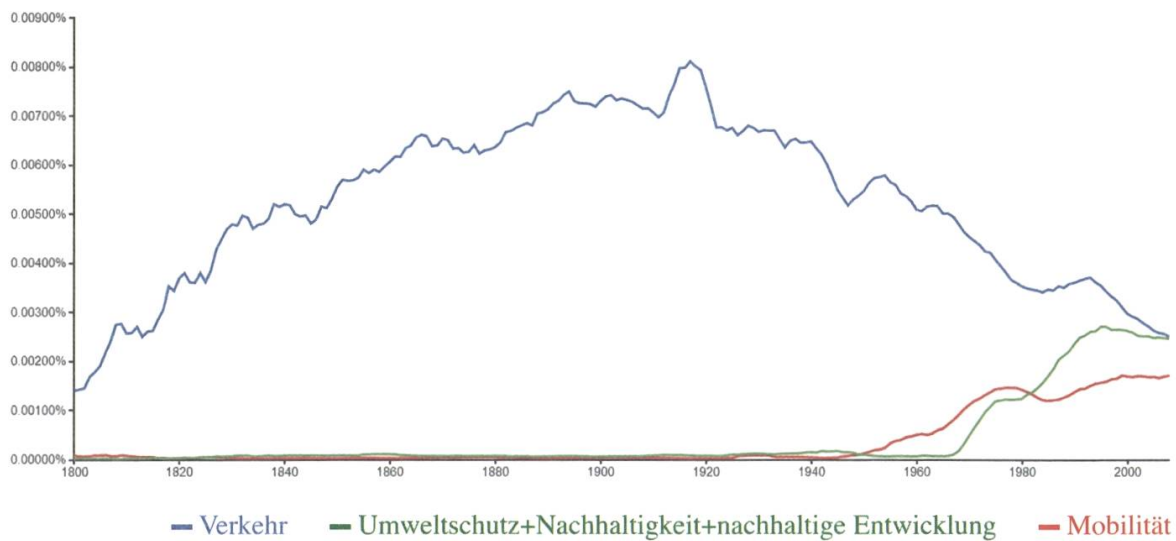
Quelle: <http://books.google.com/ngrams>.

Abb. 6: N-Gramm zu den Begriffen «Verkehr», «Mobilität» und «Umweltschutz» für 1950–2008 in deutschsprachigen Publikationen (gleitender 3-Jahresdurchschnitt).



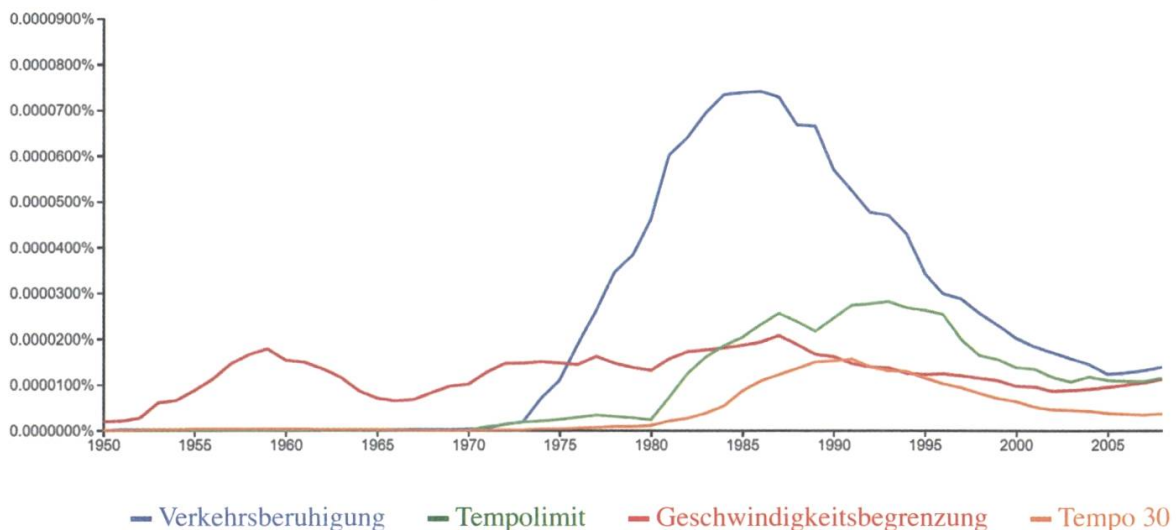
Quelle: <http://books.google.com/ngrams>.

Abb. 7: N-Gramm zu den Begriffen «Verkehr», «Mobilität» und «Umweltschutz+Nachhaltigkeit+nachhaltige Entwicklung» für 1950–2008 in deutschsprachigen Publikationen (gleitender 3-Jahres-Durchschnitt).



Quelle: <http://books.google.com/ngrams>.

Abb. 8: N-Gramm zu den Begriffen «Verkehrsberuhigung», «Tempolimit», «Geschwindigkeitsbegrenzung» und «Tempo 30» für 1950–2008 in deutschsprachigen Publikationen (gleitender 3-Jahres-Durchschnitt).



Quelle: <http://books.google.com/ngrams>.

zurückgenommen, nachdem die Zahl der Verkehrstoten in die Höhe geschossen war. Das neue Tempolimit von 50 km/h innerorts entschärfte auch die Debatte um Geschwindigkeitsbegrenzungen ganz generell, wie nicht zuletzt das N-Gramm bestätigt. Damit erhärtet unsere Analyse einen in der vorliegenden Literatur bereits recht gut abgesicherten Befund, ohne grundlegende neue Erkenntnisse hervorzubringen. In der bisherigen Literatur kaum behandelt wird hingegen, dass die nach 1970 einsetzende Debatte um die Verkehrsberuhigung offensichtlich eine sehr viel höhere Intensität erreichte als die Frage der Geschwindigkeitsbegrenzung. Ebenso interessant scheint der klare Relevanzverlust dieser Debatte ab Mitte der 1980er-Jahre, was mit einer Phasenverschiebung von fünf bis acht Jahren auch für die weniger häufig gebrauchten Begriffe «Tempolimit» und «Tempo 30» gilt. Diese Befunde sind in dieser Klarheit neu und sie verdienen es zweifellos, künftig vertieft untersucht zu werden. Denn es ist keineswegs so, dass die Begriffe Verkehrsberuhigung und Tempo-30-Zonen im deutschsprachigen Raum Mitte der 1980er-Jahre unumstritten gewesen wären, sie setzten sich vielmehr, wenn überhaupt, erst mehr als ein Jahrzehnt später durch. Mit diesem dritten Fallbeispiel, so ein weiteres Zwischenfazit, kann die formulierte Hypothese hinlänglich bestätigt werden und es können darüber hinaus neue Fragestellungen abgeleitet werden, welche der Forschung zur Geschwindigkeitsbegrenzung zweifellos wertvolle neue Impulse verleihen können.

Fazit

Als vorläufiges Fazit – denn sicherlich steht die Historiografie diesbezüglich erst am Anfang – hat sich das verkehrs- und mobilitätsgeschichtliche Potenzial von N-Grammen bestätigt. Es konnte gezeigt werden, dass sich mit diesem methodischen Ansatz erstens bestehende Hypothesen besser abstützen lassen, dass sich zweitens neue, originäre Hypothesen ergeben und dass sich diese drittens teilweise auch mittels N-Grammen überprüfen lassen. Viertens ist die einfache Form der Visualisierung durch N-Gramme von Bedeutung. Die beispielhaften Analysen weisen aber auch auf mehr oder weniger grundsätzliche Herausforderungen hin, denen es beim Einsatz von N-Grammen Rechnung zu tragen gilt. Dabei ist zu unterscheiden zwischen Problemen, die in erster Linie Google Ngrams betreffen, und solchen, die es generell beim Text Mining in grossen Datenbeständen zu berücksichtigen gilt.

Wenn wir zunächst die spezifischen Unzulänglichkeiten des Google-Quellenkorpus in den Blick nehmen, so sei erstens auf die Selektivität des Korpus hingewiesen: Bücher sind nicht dasselbe wie Broschüren, Zeitungsartikel, Werbeunterlagen, Protokolle von Parlamenten und Parlamentskommissionen usw. In solchen Quellenbeständen sind spezifische und von der generellen Entwicklung abweichende, aber letztlich doch sehr relevante Begriffsgeschichten zu erwarten, hinter denen sich auch eigenständige Diskursstränge verbergen. Daran anknüpfend kann zweitens auch moniert werden,

dass der Google-Quellenkorpus nur eine Art «Generaldiskurs» abbildet und somit die entscheidenden Schnittstellen der Begriffsgeschichte zu wenig präzise erfasst. Drittens ist sehr zu bedauern, dass nach 2008 publizierte Bücher bisher nicht enthalten sind. Viertens lassen die zulässigen Suchstrategien bei aller Vielfalt doch auch grosse Wünsche offen. Erwähnt sei hier nur, dass es bisher nicht möglich ist, zu überprüfen, ob zwei oder mehrere Begriffe in ein und derselben Publikation vorkommen. Fünftens wäre selbstredend mehr Transparenz bezüglich des Quellenkorpus dringend wünschbar. Unbeantwortet bleiben bisher Fragen nach der Auswahl der Bibliotheken und den Auswahlkriterien dieser Bibliotheken bei der Selektion der ausgewählten Bücher. Sechstens werden Bücher mit sehr kleiner und sehr grosser Auflage undifferenziert gleichgestellt, was die publizistische Strahlkraft einer Publikation natürlich überhaupt nicht abbildet. Siebtens und abschliessend sei nochmals auf die konkreten methodischen Herausforderungen hingewiesen, die weiter oben schon ausführlicher angesprochen wurden: Der einfache Zugang zur N-Gramm-Analyse verleitet schnell zu einem fahrlässigen und unreflektierten Umgang mit dieser Methodik.

Mit der Diskussion der grundsätzlichen Herausforderungen von N-Grammen verbindet sich auch die Hoffnung und letztlich auch die Erwartung, dass in Zukunft auch andere grosse Quellenbestände in einer ähnlich aufgearbeiteten Form vorliegen werden und dass diese mit einfachen, aber gegenüber heute deutlich verbesserten Analyseinstrumenten ausgewertet werden können. Attraktiv wären beispielsweise verwandte Untersuchungen im schweizerischen «Bundesblatt» oder in den Protokollen des eidgenössischen Parlaments. Darüber hinaus sind aber auch Analysen von ganz anderen Beständen denkbar. So wurden im Rahmen des Nationalfondsprojekts «Mobilität im schweizerischen Bundesstaat. Ein neuer Blick auf die Verkehrsgeschichte der Schweiz nach 1848» an der Universität Bern beispielsweise die gesamten Publikationen (es sind seit 1932 über 1000 Hefte) des Schweizerischen Jugendschriftwerks digitalisiert. Hannes Theinhardt hat diesen Textkorpus bezüglich mobilitätsgeschichtlicher Fragestellungen ausgewertet.³² Es ist offensichtlich, dass sich dieser Korpus auch für die Untersuchung ganz anderer Fragestellungen anbieten würde. Gelingt es, in diesem Sinne weitere Quellenkorpusse aufzuarbeiten, dann dient dies nicht zuletzt dem Anliegen, wirkliche Transparenz und Forschungsfreiheit herstellen zu können und die Abhängigkeit historiografischen Schaffens von einem einzigen Grosskonzern abzuschwächen.

Bezüglich der genannten Herausforderungen bei N-Gramm-Analysen oder von verwandten Ansätzen der Digital Humanities sei erstens darauf hingewiesen, dass wie in allen einschlägigen Quelleninterpretationen die Kenntnis des Forschungsstandes sowie grundlegende quellenkritische Verfahren eine unabdingbare Voraussetzung bleiben. Ein einfaches und schnelles Scannen von grossen Datenmengen mag verfüh-

32 Hannes Theinhardt, Mobilität in den SJW-Heften. Eine quantitativ-qualitative Analyse, Masterarbeit Historisches Institut Universität Bern, 2017.

rerisch wirken, es als zielführend zu erachten, wäre aber wohl naiv. Zweitens braucht es in den allermeisten N-Gramm-Analysen nachgelagerte, in der Regel recht aufwändige Untersuchungsschritte, dies haben auch die beispielhaften Analysen in diesem Beitrag deutlich gemacht. Meist dürfte es sich dabei um interpretative, qualitativ angelegte Verfahren handeln: Es muss beispielsweise nach subtilen Veränderungen in der Verwendung von Schlüsselbegriffen gefragt werden.

In diesem Sinn können N-Gramme hauptsächlich als wertvolle Ergänzung des bestehenden Methodenkanons gesehen werden und weniger als dessen Ersatz. Inwieweit auch elaborierte statistische Verfahren angewendet werden könnten, sei hier dahingestellt. In jedem Fall gälte es dabei zu beachten, dass die damit verbundenen Operationalisierungsschritte nicht zu unzulänglichen Verzerrungen führen, vor allem dann, wenn es sich um lange Analysezeiträume handelt.

Unter der Voraussetzung, dass ein derart skizzierter Umgang mit Google Ngrams, aber auch mit anderen grossen Quellenbeständen und mit verwandten Analyseansätzen im Feld der Culturomics gelingt, sollten die Geschichtswissenschaften diese Chance vermehrt nutzen. Dieses Statement ist nicht zuletzt auch als forschungspolitische Strategie zu verstehen. Es scheint klar, dass sich im Bereich von Big Data und damit verbundener Forschung in Zukunft viel bewegen wird, mit oder ohne die historischen Wissenschaften. In diesem Sinn sind die hier geäusserten Gedanken auch als Beitrag zu einer fachinternen Debatte zu verstehen.

Die in diesem Beitrag angedachten Beispiele von N-Gramm-Analysen haben das grosse Potenzial solcher Ansätze für die Verkehrs- und Mobilitätsgeschichte deutlich gemacht. Es lassen sich nicht nur bestehende Befunde erhärten, wie etwa zur Begriffsgeschichte der Mobilität, sondern es öffnen sich auch neue Zugänge zu alten Fragen wie etwa derjenigen nach der gesellschaftlichen Bedeutung der Schlüsselinnovationen Eisenbahn, Auto und Flugzeug im direkten Vergleich. Und es entstehen auch neue spannende Forschungsfragen, wie etwa zum kurzen Hype des Begriffs Verkehrsberuhigung in den 1980er-Jahren. Und nicht zuletzt: Der Verkehrs- und Mobilitätsgeschichte werden neue Quellengattungen erschlossen, deren Verwendung in der Forschungsdebatte immer wieder angemahnt wurde, beispielsweise belletristische Werke oder – über den googleschen Korpus hinaus – auch Tageszeitungen und andere Periodika. Der oft problematischen Verengung von verkehrs- und mobilitätsgeschichtlichen Analysen auf technische, ökonomische oder rein politische Diskurse lässt sich damit Erfolg versprechend entgegenwirken.³³ In diesem Sinne soll dieser Beitrag zu einem lustvollen und gleichzeitig stets gut reflektierten Experimentieren mit den neuen Möglichkeiten der Digital Humanities ermuntern.

33 Mom (wie Anm. 11).