

»Google Books Ngram Viewer« und historische Computerlexikologie: Der Sprachgebrauch der NS-Zeit

Von KLAAS WILLEMS

Abstract

»Ngram Viewer« (<http://books.google.com/ngrams>) ist ein neues sprachtechnologisches Tool, das im Internet frei zur Verfügung steht und auf einem digitalen Korpus von mehr als 5 Millionen optisch eingescannten Büchern in verschiedenen Sprachen basiert. Das Tool verspricht, ein wichtiges Instrument eines neuen Forschungsfeldes, »culturomics«, zu werden, das einen Zweig der Computerlexikologie darstellt und aufgrund von gedruckten Quellen die Geschichte menschlicher Kultur im weitesten Sinne zum Gegenstand hat. Der vorliegende Beitrag untersucht die Möglichkeiten und den Nutzen von »Ngram Viewer« für die historische Wortschatzanalyse des Deutschen. Das Korpus, aufgrund dessen statistische Daten über die deutsche Sprache erhoben werden können, beträgt zurzeit 37 Milliarden Wörter (zum Vergleich: das englische Korpus enthält 361 Milliarden Wörter). Anhand von mehr als 100 Ausdrücken, die vornehmlich Victor Klemperers Buch *LTI. Notizbuch eines Philologen* (1947; ³1957), daneben aber auch anderen Publikationen über den Sprachgebrauch der NS-Zeit (u. a. Cornelia Schmitz-Berning, *Vokabular des Nationalsozialismus* 2000) entnommen sind, werden spezifische Thesen über die Gebrauchsgeschichte einzelner Ausdrücke vor, während und nach der NS-Zeit (1918–1945) überprüft, wobei auch die Zuverlässigkeit von »Ngram Viewer« ein besonderes Augenmerk des Beitrags bildet.

»Ngram Viewer« (<http://books.google.com/ngrams>) is a new search tool which is freely available through the internet and created to browse over 5 million books in different languages digitized by Google (this is roughly a third of all the books that have been digitized so far and approximately 4 % of all books ever published). »Ngram Viewer« promises to be a powerful tool in an emerging new discipline called »culturomics«. »Culturomics« is a form of computational lexicology focusing on the history of human culture as it manifests itself in published material. This article investigates the possibilities and problems of »Ngram Viewer« with regard to the analysis of the history of the German lexicon. The available German corpus currently contains 37 billion words (compared to 361 billion words in the English corpus). The study draws on a sample of more than 100 expressions (»n-grams«) which are commonly regarded as typical of the language of the Third Reich. The sample is taken from publications such as Victor Klemperer's book *LTI. Notizbuch eines Philologen* (1947; ³1957) and Cornelia Schmitz-Berning's *Vokabular des Nationalsozialismus* (2000). The usage frequency of the expressions is examined by means of »Ngram Viewer«, with a focus on the historical data before, during and after the Third Reich period (1918–1945), and with special attention to some of the scanning errors arising from difficulties when the optical character recognition program (OCR) is applied to older German texts.

1 Einführung

Dass die Korpuslinguistik die Sprachwissenschaft seit den 1990er Jahren maßgeblich geprägt hat, bezweifelt heute niemand mehr. Die Beschäftigung mit authentischen, sowohl zeitgenössischen als auch historischen, Sprachdaten aufgrund immer umfangreicherer und differenzierterer Korpora hat darüber hinaus eine Fülle neuer und wertvoller Erkenntnisse zu Tage gefördert. Gebrauchsfrequenzen von Wörtern, deren kollokationelle Beschränkungen und semantische Orientierungen in bestimmten Kontexten und Textsorten, die Häufigkeit von Konstruktionen und deren Präferenzen für bestimmte Wörter (und umgekehrt) – all das sind Facetten des Sprachgebrauchs, die durch die Digitalisierung der

sprachwissenschaftlichen Ressourcen allenthalben in die Reichweite der Forscher gerückt sind, und zwar in einem Maße, das vor einigen Jahrzehnten noch unvorstellbar war. Entsprechend ist die Anzahl an interessanten Aufsätzen und Monographien, die oft detaillierte Korpusanalysen enthalten, heute kaum noch überschaubar – auch wenn das Verhältnis zwischen der Fülle an empirischen Beobachtungen und dem jeweiligen systemlinguistischen Stellenwert einzelner Sprachzeichen und Konstruktionen manchmal nur wenig Beachtung findet.

Neben sprachwissenschaftlichen Korpora haben auch Internetsuchmaschinen wie Google die wissenschaftliche Praxis von Linguisten in jüngster Zeit zunehmend beeinflusst. Die enorme Fülle an Sprachmaterial, die den Internetbenutzern zugänglich ist, kann oft hilfreich sein, bestimmte korpuslinguistische Beobachtungen entweder zu ergänzen oder zu nuancieren, unter der Bedingung freilich, dass die Daten, die im Netz erhoben werden, sorgfältig kontrolliert und abgegrenzt werden (für ein ausführliches Beispiel s. u. a. Willems 2011). Neuerdings stellt die Firma Google im Internet ein sprachtechnologisches Tool namens »Ngram Viewer« (<http://books.google.com/ngrams>) zur Verfügung, das die Möglichkeiten einer quantitativen Analyse von gedrucktem Sprachmaterial noch einmal erweitert. »Ngram Viewer« basiert auf einem digitalen Korpus von etwa 5,2 Millionen optisch eingescannten Büchern in verschiedenen Sprachen. Das Korpus umfasst ein gutes Drittel aller bisher eingescannten Bücher, die in »Google Books« (<http://books.google.com/>) durchsucht werden können; 15 Millionen Bücher entsprechen ungefähr 12 % aller seit der Erfindung des Buchdrucks erschienenen Bücher, deren Anzahl auf etwa 129 Millionen geschätzt wird. Für eine ausführliche Erläuterung von »Ngram Viewer« und der zugrunde gelegten Korpora s. Michel/Shen/Aiden u. a. (2011). Die Autoren betrachten »Ngram Viewer« als ein wichtiges Instrument eines neuen Forschungsfeldes, das sie »culturomics« nennen und das als Teil der Computerlexikologie menschliche Kultur im weitesten Sinne unter dem Aspekt quantitativer Analyseverfahren zum Gegenstand hat.

Die jeweilige Häufigkeit von N-Grammen¹ wird mit »Ngram Viewer« auf eine benutzerfreundliche Weise statistisch berechnet und präsentiert. Die Ergebnisse werden in Form von übersichtlichen Liniendiagrammen dargestellt, deren Abszissenachse die geschichtliche Entwicklung und deren Ordinatenachse die Häufigkeit des gewählten N-Gramms im Vergleich zu allen Wörtern des Korpus in Prozenten angeben. Die Datenerhebung kann in sprachlich unterschiedlichen Korpora vorgenommen werden, außer in verschiedenen Teilkorpora mit englischen (u. a. britischen und amerikanischen) Texten auch in französischen, deutschen, hebräischen, russischen und spanischen Korpora. Zum Vergleich: Das englische Korpus enthält 361 Milliarden Wörter, das deutsche Korpus 37 Milliarden Wörter (Michel/Shen/Aiden u. a. 2011: 176). Der Zeitraum der erfassten Buchpublikationen reicht vom 16. Jahrhundert bis zum Anfang des 21. Jahrhunderts. Weil es aber erhebliche Unterschiede gibt in der Erfassung von Texten aus den unterschiedlichen Perioden, insofern deren Anzahl in den frühen Jahrhunderten natürlich viel kleiner ist als seit dem 19. Jahrhundert, müssen im Hinblick auf frühere Perioden statistische Normalisierungen durchgeführt werden, die im Prinzip die Ratio zwischen der Belegdichte eines Ausdrucks und der Anzahl

¹ Als N-Gramme bezeichnet man in der Computerlinguistik und Quantitativen Linguistik Fragmente eines Textes, die aus einer beliebigen Anzahl von (aufeinander folgenden) Zeichen bestehen, z. B. Monogramme, Bigramme, Trigramme usw. Die Zeichen können Buchstaben, Syllaben, Zahlen usw. sein, aber auch ganze Wörter, wie die N-Gramme (Multigramme), mit denen wir uns im Folgenden beschäftigen.

der erfassten Bücher in einem gegebenen Zeitraum berücksichtigen. Diese Verfahrensweise wirkt sich allerdings nur ausnahmsweise in verlässliche Graphiken für die frühen Jahrhunderte (16., 17., zum Teil 18. Jh.) aus, was für eine Analyse von Worthäufigkeiten ab dem 19. Jahrhundert aber weiter nicht schlimm ist.

Die Zeitspanne für die angepeilte Erhebung kann beliebig variiert werden, etwa von 1800 bis 2000 oder gezielt von 1933 bis 1945. Lässt man »Ngram Viewer« zum Beispiel die Häufigkeit der bestimmten Artikelformen von 1800 bis 2000 in deutschen Texten errechnen, dann sieht das Ergebnis aus wie in Diagramm 1. Die erfassten Wörter sind zufällig allesamt Trigramme. Ein so genanntes »Smoothing« im Wert von 3 wurde durchgeführt; für jedes Jahr wurde also ein Durchschnittswert von insgesamt 7 Jahren errechnet (3 Jahre vor und 3 Jahre nach dem jeweiligen Bezugsjahr), mit entsprechendem Effekt auf die graphische Darstellung.

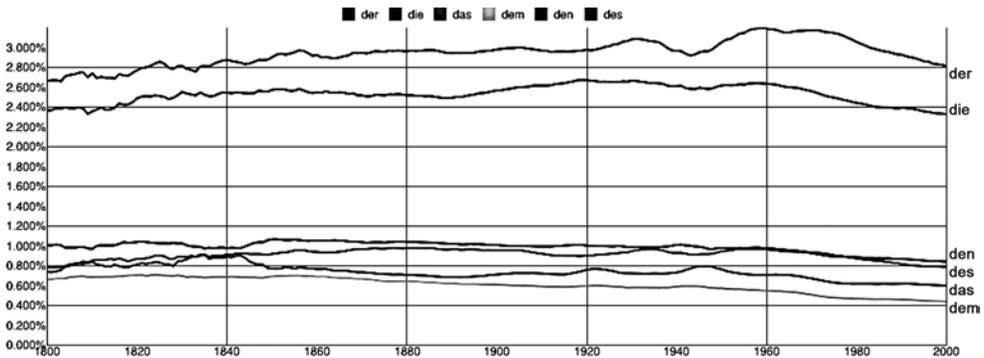


Diagramm 1: Die Häufigkeit der bestimmten Artikelformen *der*, *die*, *den*, *des*, *das* und *dem* zwischen 1800 und 2000 in deutschsprachigen Texten nach »Ngram Viewer« (Google)

Aus Diagramm 1 geht hervor, dass die Häufigkeit der bestimmten Artikelformen in den ins Auge gefassten zwei Jahrhunderten relativ konstant ist (wobei die Formen *der*, *die*, *das* und *dem* die größten Schwankungen aufweisen). Die Gesamtfrequenz des bestimmten Artikels liegt 2000 bei 7,8 %, was ungefähr mit dem Ergebnis traditioneller Frequenzstatistiken übereinstimmt und somit auch als ein Indiz für die generelle Zuverlässigkeit von »Ngram Viewer« angesehen werden kann.² Wenn man bedenkt, dass Google unablässig Bücher aus mehreren Jahrhunderten einscann und (wenn auch, aus autorenrechtlichen Gründen, in der Regel nur teilweise) im Netz zur Verfügung stellt, dann dürfte die Hoffnung berechtigt sein, dass »Ngram Viewer« ein leistungsfähiges Instrument zu sein verspricht, das manchen Sprachwissenschaftlern in ihren historischen Analysen von Ausdrücken (Wörtern, Phrasen, Konstruktionen) gute Dienste leisten kann.

Vor dem soeben skizzierten Hintergrund verfolgt der vorliegende Beitrag ein doppeltes Ziel. Einerseits erscheint es interessant zu untersuchen, ob bzw. in welchem Umfang bestimmte vorhandene Thesen über die Geschichte des Wortschatzes einer bestimmten Sprache mit den Ergebnissen von »Ngram Viewer« übereinstimmen. Andererseits stellt der Beitrag einen ersten Versuch dar, den Nutzen von »Ngram Viewer« und dessen Zuverlässigkeit

² S. aber Abschnitt 4 unten. Bekanntlich gilt der bestimmte Artikel in einschlägigen Sprachen als das Sprachzeichen (»expression unit«), das in der Regel die höchste Frequenz aufweist (s. Zipf 1935, Baayen 2001 sowie die ständig aktualisierten Frequenzlisten zum deutschen Wortschatz unter <http://wortschatz.uni-leipzig.de/>).

an einem relativ gut abgrenzbaren Teilbereich des Lexikons zu überprüfen. Diese beiden Fragen sollen im Folgenden anhand einer Auswahl von Ausdrücken aus dem Sprachgebrauch der NS-Zeit beantwortet werden, die wir im Anschluss an Schmitz-Berning (2000: vii) als die Zeit zwischen 1918 und 1945 definieren.³ Als Leitfaden für die Untersuchung dient uns dabei vornehmlich Victor Klemperers berühmtes, kurz nach dem Zweiten Weltkrieg erschienenes Buch *LTI. Notizbuch eines Philologen* (1947; ³1957).⁴

2 LTI

LTI, oder »Lingua Tertii Imperii«, ist die Bezeichnung, die Victor Klemperer in seinem Buch der Sprache des Dritten Reiches gegeben hat, die, wie Klemperer öfter hervorhebt, in kürzester Zeit in den Sprachgebrauch sehr vieler Menschen in Deutschland eingegangen sei (Klemperer ³1957: 100 spricht von einer »nazistischen Infektion«). Das Buch ist geradezu am Leitfaden von Ausdrücken abgefasst, denen Klemperer seit den 1920er Jahren, und besonders seit der Machtübernahme der Nationalsozialisten im Jahre 1933, sowohl in schriftlicher als auch in mündlicher Form ständig begegnete. In den insgesamt 36 Kapiteln des Buches konzentriert sich Klemperer auf einzelne Wörter und Ausdrücke wie *aufziehen*, *gleichschalten*, *betreuen*, *fanatisch*, *kämpferisch*, *total*, *völkisch*, *charakterlich*, *Gefolgschaft*, *Untermensch* und *Übermensch*, *Umbruch*, *aufnorden*, *Gau*, *Sippe*, *Volkskanzler*, *Großkundgebung*, *deutsche Pflicht*, *gesundes Rechtsempfinden* usw.⁵ Auch bestimmten syntaktischen Konstruktionen sowie der Interpunktion schenkt Klemperer seine Aufmerksamkeit.⁶

»Ngram Viewer« bietet sich geradezu von selbst an, Klemperers Beobachtungen und Thesen über den Sprachgebrauch des Nationalsozialismus zu überprüfen. Dabei sind drei Bemerkungen Klemperers vorneweg als wichtig hervorzuheben. Sie stellen ebenso viele Hypothesen dar, die von einem EDV-Standpunkt wie dem durch »Ngram Viewer« ermöglichten besonders untersuchenswert erscheinen. Erstens weist Klemperer (³1957: 24–25)

³ Diese Zeitspanne wird von Schmitz-Berning (2000: vii) in die sog. »Kampfzeit 1918–1933« und das »Dritte Reich 1933–1945« unterteilt.

⁴ Wir legen unserer Analyse die im VEB Max Niemeyer Verlag (Halle/Saale) erschienene Ausgabe des zunächst 1947 im Berliner Aufbau-Verlag veröffentlichten Buches zugrunde. Die von uns herangezogene Ausgabe ist 1957 in 3. Auflage erschienen (Klemperer ³1957). Das Buch hat seitdem als Band 278 verschiedene Neuauflagen in Reclams Universal-Bibliothek erfahren. Vgl. zum Thema auch Klemperer (1995). Unter den Buchpublikationen, die man bei Bedarf komplementär zu Klemperers Buch ebenfalls konsultieren kann, sei u. a. hingewiesen auf Sternberger/Storz/Süskind (³1968), Seidel/Seidel-Sloty (1961), Bork (1970), Ehlich (Hg., 1989), Schmitz-Berning (2000) und Michael/Doerr (2002). Daneben lassen sich auch Publikationen zur Geschichte des Nationalsozialismus, wie z. B. Zentner/Bedürftig (1985) und Wildt (2008), mit Gewinn für historische lexikologische Studien heranziehen. Für einen knappen, aber informativen geschichtlichen Abriss über die Sprache des Nationalsozialismus s. Polenz (1999: 538–555).

⁵ Ein so gut wie vollständiges, über 400 Seiten starkes »Nazi-Deutsches Lexikon« haben Michael/Doerr (2002) erarbeitet.

⁶ Beispielsweise weist Klemperer (³1957: 276) darauf hin, dass ehemals intransitive Verben wie *fliegen* und *frieren* unter den Nationalsozialisten zu transitiven Verben »aktiviert« worden seien (*man fliegt eine schwere Maschine*, *man friert Gemüse im neuen Verfahren der Tiefenkühlung*) und dass es sowohl in Zeitungsartikeln als auch in »ruhiger gehaltenen ausführlichen Studien« zur NS-Zeit von Anführungszeichen geradezu gewimmelt habe; mit Ausrufezeichen sei die LTI dagegen sparsam umgegangen (Klemperer ³1957: 90–91).

darauf hin, dass das Dritte Reich »die wenigsten Worte seiner Sprache selbstschöpferisch geprägt [hat], vielleicht, wahrscheinlich sogar, überhaupt keines«. Die LTI nutzt dem Autor zufolge in der Regel »nur, was sie schon vorfindet« (Klemperer ³1957: 144; vgl. auch 62 sowie Sternberger/Storz/Süskind ³1968: 15 und Frind 1966: 130). Nur ausnahmsweise, so vermutet Klemperer, könne von einer »Neuprägung der Nazis« die Rede sein, so etwa im Falle von *charakterlich* (Klemperer ³1957: 236). Zweitens habe die LTI die bereits vorhandenen Wörter und Ausdrücke »in ihrem Wert verändert« und viel häufiger gebraucht als zuvor, sie habe »Wertwerte und Worthäufigkeiten« geändert und zum Allgemeingut gemacht, »was früher einem einzelnen oder einer winzigen Gruppe gehörte«; »in alledem durchtränkt sie Worte und Wortgruppen und Satzformen mit ihrem Gift« (Klemperer ³1957: 24–25). Drittens scheint die Sprache des Dritten Reiches Klemperer zufolge »in manchen charakteristischen Ausdrücken überleben zu sollen; sie haben sich so tief eingefressen, daß sie ein dauernder Besitz der deutschen Sprache zu werden scheinen« (Klemperer ³1957: 23; vgl. auch Sternberger/Storz/Süskind ³1968 und Doerr 2002 sowie Polenz 1999: 551). Diese drei Beobachtungen lassen sich vor dem Hintergrund unserer Beschäftigung mit »Ngram Viewer« als Forschungsfragen, und damit auch als Ausgangspunkt für die vorliegende Untersuchung, wie folgt neu formulieren: Stimmt es, dass

- die traditionell als »Nazi-Deutsch« (Michael/Doerr 2002) eingestuft Ausdrücke in der Regel keine Neuprägungen in der Geschichte der deutschen Sprache darstellen?
- sich die Sprache des Nationalsozialismus der vorangehenden Periode gegenüber in erster Linie durch geänderte Worthäufigkeiten auszeichnet?
- viele Ausdrücke des Nationalsozialismus ein fester Bestandteil der deutschen Sprache geworden sind und seit 1945 weiter verwendet werden?

3 »Ngram Viewer«

3.1 Die Methodologie

Die von Klemperer durchgeführten Analysen, die sich durch Informationen aus vielen anderen Quellen ergänzen lassen, können zugleich zum Anlass genommen werden zu untersuchen, ob die durch »Ngram Viewer« generierten Ergebnisse zuverlässig sind, bietet es doch den Benutzern im Prinzip die Möglichkeit, alle eruierten N-Gramme separat abzurufen und zu kontrollieren (»im Prinzip«, weil die Berechnungen auf den Daten in »Google Books« aufbauen, die aber nicht immer von einseharen Textbelegen begleitet werden). In der Regel aber kann man über die Publikationen, auf denen jede N-Gramm-Statistik jeweils beruht, eine gute Übersicht erhalten.⁷ Nicht nur Autor und Erscheinungsjahr kann man überprüfen, sondern auch, ob die Erhebung korrekt verlaufen ist und z. B. keine formal ähnlichen, aber dennoch nicht identischen Wortformen miterfasst wurden und ob die Quellen richtig datiert sind (wir werden weiter unten in Abschnitt 4 sehen, dass solche Fehler tatsächlich auftreten).

Bevor wir uns die Ergebnisse der Erhebungen genauer anschauen, sind noch einige allgemeine Beobachtungen über die Daten, die wir aufgrund von »Ngram Viewer« vorführen werden, angebracht:

⁷ Zu beachten ist aber, dass die abrufbaren Belegstellen nicht nur aus dem selektierten Teilkorpus (z. B. deutscher Bücher), sondern aus allen von »Google Books« eingescannten Publikationen stammen (s. <http://books.google.com/ngrams/info>). Vgl. auch unten Abschnitt 4.

- Die Belegdichte der untersuchten Ausdrücke bewegt sich in der in Betracht gezogenen Periode 1850–2000 zwischen $2.1e-2$ (*Philosophie*) und $8.0e-6$ (*niedermachen*); ein N-Gramm wird statistisch als gebräuchlich (»common«, Michel/Shen/Aiden u. a. 2011: 176) eingestuft, sobald es häufiger ist als 1×10^{-9} (d. h. also $1.0e-9$ oder 0,000000001); alle von uns analysierten Ausdrücke liegen klar über dieser Untergrenze;
- aus Platzgründen können im Folgenden nur einige Liniendiagramme zur Illustration aufgeführt werden; je nach N-Gramm erweist es sich als sinnvoll, entweder eine längere oder eine kürzere Periode ins Auge zu fassen; für *HJ* oder *Hitlerjugend* ist es natürlich nicht sinnvoll, weit in die Zeit zurückzugehen, für *charakterlich* ist das aber durchaus der Fall (s. Abschnitt 4);
- weil das Deutsche eine komplexere Morphologie hat als das Englische, sind in der Regel mehr Sucheingaben erforderlich, um in den beiden Sprachen eine vergleichbare Vollständigkeit in den Ergebnissen zu erzielen (z. B. insgesamt 6 Adjektivformen, etliche Verbformen, auch mehrere Substantivformen, um alle Kasus und Numerus zu erfassen, usw.); jedoch ist es nicht das Ziel unserer Analyse, die Vorkommenshäufigkeit einzelner Ausdrücke im absoluten Sinne zu bestimmen, sondern wir sind vielmehr an deren Aufkommen, relativer Frequenz und den Schwankungen im Laufe der Zeit interessiert, so dass partielle Korpusabfragen durch die Bank ausreichen.

Wir gehen jetzt zur Besprechung einzelner Ausdrücke über. Weil »Ngram Viewer« nur solche Ausdrücke erfasst, die im gewählten Korpus mindestens 40-mal vorkommen (Michel/Shen/Aiden u. a. 2011: 176) und darüber hinaus nur in gedruckten Quellen, eignen sich einige der von Klemperer als LTI identifizierte Ausdrücke nicht für eine Analyse mittels »Ngram Viewer«. Für Komposita wie *Hitler-Eiche* (Klemperer ³1957: 23), *Fahrjuden*, *Lauffjuden*, *Waschjuden* (Klemperer ³1957: 237) oder umgangssprachliche und auf den gesprochenen Sprachgebrauch beschränkte Formen wie *Knif* (»Kommt nicht in Frage«), *Kakfif* (»Kommt auf keinen Fall in Frage«) (Klemperer ³1957: 110) gibt es im deutschen »Google Books«-Korpus keine Belege. Über deren Entstehungszeit (beispielsweise weist Klemperer darauf hin, dass er das Akronym *Knif* zwei Jahre vor dem Krieg zum ersten Mal gehört habe) lassen sich anhand von »Ngram Viewer« somit keine Aussagen machen. Über weitaus die meisten der von Klemperer und anderen Autoren angeführten Wörter und Phrasen aber gibt »Ngram Viewer« Auskünfte, und diese belegen zwar nicht immer, aber doch mehrheitlich (s. weiter unten), dass Klemperers Einschätzungen nicht nur zutreffen, sondern oft auch sehr präzise sind.

Wir nehmen zunächst solche Ausdrücke in Augenschein, die Klemperer (³1957: 23) zufolge bereits vor der Zeit der LTI vorhanden waren, denen aber gleichsam über Nacht in der NS-Zeit ein Häufigkeits- und Wertewandel zuwuchs, wobei auch auffällige Unterschiede im Häufigkeitwandel nach 1945 interessant sein können. Wir wenden uns daraufhin Fremdwörtern und zwei Eigennamen sowie einigen Wortbildungen (Zusammensetzungen und Ableitungen) zu; letztere Kategorie scheint darüber hinaus besonders geeignet, bestimmte Thesen über (vermeintliche) »Neuprägungen« aus der NS-Zeit zu überprüfen. Abschließend gehen wir auf einige Probleme im Hinblick auf Quellendatierung und Wort- bzw. Formerkennung durch das Programm ein, auf die wir bei der Verwendung von »Ngram Viewer« gestoßen sind.

3.2 Alte Ausdrücke, andere Häufigkeiten, neue semantische Werte

Klemperer weist darauf hin, dass bereits bestehende Ausdrücke wie *kämpferisch* (Klemperer ³1957: 23), *fanatisch* (24, 72–77, 233–235), *Staatsakt* (57), *aufziehen* (57, 60–63), *Umbruch* (234) usw. in der NS-Zeit in kurzer Zeit viel häufiger verwendet werden als zuvor und zugleich einen neuen semantischen Wert annehmen. Beispielsweise erhält ein Verb wie *aufziehen*, das bis dahin pejorativ konnotiert war, in der LTI entweder eine neutrale oder eine ausgesprochen positive Konnotation (Meliorisierung). Ähnliche Änderungen im Bedeutungswert (sog. »Konnotation«) gelten für Adjektive wie *fanatisch*, *rücksichtslos* und viele andere Wörter. Über die semantischen Verschiebungen selbst gibt »Ngram Viewer« keine Auskünfte, Änderungen in der Belegdichte der Ausdrücke über die Jahrzehnte treten aber klar hervor. Diagramm 2 gibt die Frequenz des Nomens *Staatsakt* von 1850 bis 2000 wieder. Das Diagramm kann hier als Modellbeispiel für alle Wörter aus dieser Gruppe stehen, die sowohl einen Häufigkeits- als auch einen Wertewandel durchmachen und im Großen und Ganzen (allerdings mit unterschiedlichen Vorgeschichten in den früheren Jahrhunderten) vergleichbare Entwicklungslinien aufweisen.⁸



Diagramm 2: Die Häufigkeit des Nomens *Staatsakt* zwischen 1850 und 2000 (Smoothing: 3)

Die erste unserer Forschungsfragen (s. Abschnitt 2), ob die Ausdrücke der LTI Neuprägungen sind oder nicht, lässt sich aufgrund einer Überprüfung von mehr als 100 nicht sonderlich »technischen« Ausdrücken unzweideutig beantworten: Korpusabfragen anhand von »Ngram Viewer« ergeben in weitaus den meisten Fällen ähnliche Liniendiagramme, die bezeugen, dass viele Ausdrücke von den Nationalsozialisten nicht neu geprägt, sondern aus dem vorhandenen Wortschatz übernommen wurden. Außerdem stellt sich heraus, dass

⁸ Hierunter folgt eine Auswahl von etwa 65 weiteren Ausdrücken, die in diese (und daneben eventuell auch in eine oder mehrere andere Gruppen) gehören und die wir aufgrund von Klemperers Buch (³1957) ausgewählt und mittels »Ngram Viewer« überprüft haben: *abwandern/abgewandert*, *Amtswaltung* und *Amtswalter*, *Angriff*, *arisch*, *aufgenordet*, *aufgezogen*, *aufzwingen/aufgezwungen*, *auslöschen*, *ausrotten*, *befrieden* und *Befriedung*, *Brauchtum*, *Dritte(s) Reich*, *Garant*, *Durchbruch*, *Einbruch*, *flink*, *Führer*, *Gau*, *Gefolgschaft*, *gemeinschaftsunfähig*, *Generalgouverneur* und *Generalgouvernement*, *gigantisch*, *Großdeutschland*, *heldenhaft* und *heroisch*, *herrisch*, *Herzland*, *historisch*, *Jude* und *jüdisch*, *Kadaververwertung*, *Kameraden*, *Korridor*, *lebensfroh*, *Menschenmaterial*, *Menschentum*, *Menschlichkeit* und *Humanität*, *Mischling*, *organisch*, *planmäßig*, *rassisch*, *Reich*, *Reichsbürger*, *Sippe*, *Strafexpedition*, *Terror*, *Treck*, *Übermensch* und *Untermensch*, *unermüdlich*, *völkisch* und *volkhaft*, *Volksempfinden*, *Volkskörper*, *Volkskrieg*, *vollbringen*, *Wehrmacht*, *Weltanschauung*, *Wesensmitte*, *zwangsweise*.

sie nach der NS-Zeit in der Regel wieder weniger häufig gebraucht werden. Auch auf die zweite Forschungsfrage, inwieweit die LTI in erster Linie durch geänderte Worthäufigkeiten gekennzeichnet ist, lautet die Antwort positiv: Entwicklungen wie in Diagramm 2, wobei die Belegdichte eines LTI-Ausdrucks ab der ersten Hälfte der 1920er Jahre in kurzer Zeit um ein Vielfaches (z. B. zehnfach) zunimmt und dann nach einer bestimmten Zeit (die allerdings variabel ist, s. weiter unten) wieder abnimmt, stellen sich als durchaus häufig heraus.

Die Antwort auf die dritte Forschungsfrage, die das »Nachleben« der LTI in der deutschen Sprache nach der NS-Zeit zum Gegenstand hat, erfordert eine differenzierte Betrachtung der Daten. Wie bereits erwähnt, klärt uns »Ngram Viewer« nicht unmittelbar über semantische Werte auf, aber ob ein Ausdruck mit einer spezifischen Konnotation oder auch nur metasprachlich (insbesondere in Texten *über* die NS-Zeit und ihren Sprachgebrauch) gebraucht wird, lässt sich meist ohne weiteres ermitteln, weil die Belege in der Regel einzeln abrufbar sind. Das ist natürlich zeit- und arbeitsaufwändig, aber ein solches Vorgehen hat den Vorteil, dass die rein quantitativen Daten, die »Ngram Viewer« generiert, in einen interpretativen Rahmen gestellt werden, der auch tiefere Schlüsse qualitativer Art erlaubt. Schauen wir uns deshalb einige Diagramme etwas näher an.

Obwohl die Liniendiagramme der ersten Gruppe von Wörtern entweder eine charakteristische Spitzenform oder, bei entsprechendem Smoothing, eine Glockenform (einer Gaußkurve ähnlich) aufweisen, die das Ansteigen und Abklingen einer Worthäufigkeit im Zeitraum von etwa 2 Jahrzehnten visualisiert, treten zwischen den einzelnen Ausdrücken doch interessante Unterschiede zu Tage, die mitunter historisch präzise gedeutet werden können.⁹ Das in Diagramm 2 gezeigte Ergebnis für den Ausdruck *Staatsakt* z. B. fängt mit Belegen aus dem 19. Jahrhundert an (der älteste Beleg stammt aus einem Buch aus dem Jahre 1729), die auf »feierliche« Staatsakte in unterschiedlichen Nationen und zu den verschiedensten Anlässen verweisen. Die Belegdichte nimmt von 1910 bis 1934 um mehr als das 14-fache zu (von 2.0e-5 auf 29.0e-5), der steilste Anstieg aber ist exakt 1933 zu verzeichnen, also im Jahre der Machtübernahme durch die Nationalsozialisten (Ergebnis ohne Smoothing). Auffällig ist ferner, dass die Frequenz von *Staatsakt* bereits 1934 rasch abnimmt und im letzten Kriegsjahr wieder den Stand von 1922 und 1926 (3.0e-5) erreicht. Danach lässt sich das Wort bis zum Ende der 50er Jahre aber wieder häufiger im Korpus belegen. Aus einer gezielten Musterung des Korpus im Zeitraum von 1948 bis 1960 geht hervor, dass darin nicht nur die historische Beschäftigung mit älteren Zeitabschnitten (u. a. dem nationalsozialistischen Deutschland), sondern auch die Gründungen der Bundesrepublik Deutschland und der Deutschen Demokratischen Republik (1949) eine Rolle spielen. In Bezug auf einen Ausdruck wie *Staatsakt* erhält Klemperer also insofern Recht, als der Ausdruck auch nach der NS-Zeit weiterhin gebräuchlich ist, wobei dessen Vorkommen allerdings auch an den Sprachgebrauch vor der NS-Zeit anschließen dürfte.

Adjektive wie *fanatisch* und *rücksichtslos*, die in der LTI eine auffallende (wenn auch nicht durchgehende) Meliorisierung erfahren (s. Klemperer³ 1957: 232–235 und 285), weisen insofern andere Entwicklungslinien auf, als die entsprechenden N-Gramm-Diagramme im 20. Jahrhundert bereits vor der Zeit des Dritten Reiches markante Frequenzänderungen erkennen lassen, und zwar sowohl während des Ersten Weltkriegs als auch kurz danach.

⁹ »Ngram Viewer« zeigt übrigens auch, dass manche der Ausdrücke in früheren Jahrhunderten (insbesondere im 18. oder 19. Jh.) nicht nur belegt werden können, sondern mitunter häufig verwendet wurden (so etwa *heroisch*, *historisch*, *organisch* usw.).

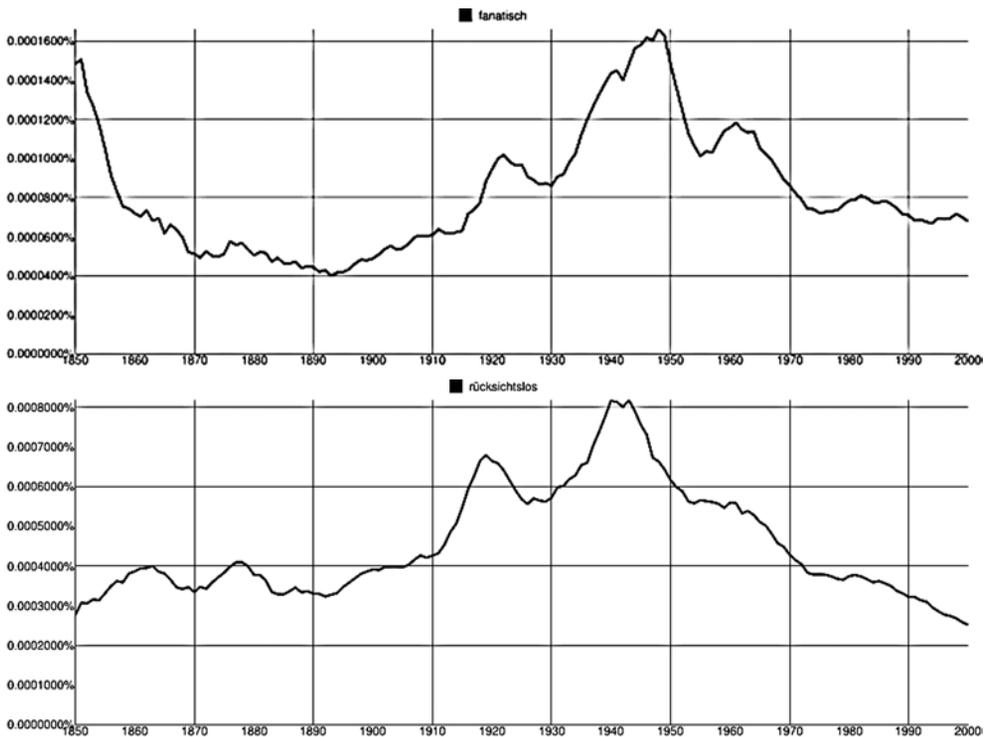


Diagramme 3 und 4: Die Häufigkeit der Adjektive *fanatisch* (3) und *rücksichtslos* (4) zwischen 1850 und 2000 (Smoothing: 3)

Bei *fanatisch* lässt sich zwischen 1915 und 1919, bei *rücksichtslos* zwischen 1914 und 1918 jeweils ein markanter Anstieg der Belegdichte verzeichnen, also in Zeitabschnitten, die größtenteils den Ersten Weltkrieg umfassen. Die Häufigkeit der beiden Wörter nimmt dann zunächst wieder ab, wonach zur Zeit des Zweiten Weltkriegs ein neuer Höhepunkt erreicht wird (*fanatisch*: $2.0e-4$ im Jahre 1945; *rücksichtslos*: $9.5e-4$ im Jahre 1943).¹⁰ Diagramm 3 deutet darüber hinaus darauf hin, dass *fanatisch* in deutschen Texten der Mitte des 19. Jahrhunderts bereits häufig vorkommt; anhand von »Ngram Viewer« lassen sich zwischen 1830 und 1860 z. B. 1661 Belege für *fanatisch* (alle flektierten Formen zusammen) verzeichnen (höchster Wert 1852: $1.8e-4$). Klemperers Behauptung im Jahre 1947, dass *fanatisch* als einer der wenigen nicht technischen Ausdrücke fast ganz aus der Gegenwortsprache »verschwunden« sei (³1957: 77), wird aber nicht ganz bestätigt. Im Vergleich zu 1945 hat die Zahl der Belege im Jahre 2000 zwar abgenommen, und das Wort wird oft im Zusammenhang des Dritten Reiches verwendet, daneben kommt es aber auch in anderen

¹⁰ Zum Vergleich: Im französischen Korpus taucht das entsprechende Wort *fanatique* (>in religiöser Verzückung, in ekstatischen Kampfständen befindlich<, Klemperer ³1957: 22) seit den 1860er Jahren immer weniger oft auf, und in der Zeitspanne zwischen 1914 und 1945 wird es im Schnitt nicht einmal halb so oft verwendet wie zur Zeit der französischen Revolution (vgl. etwa $3.5e-4$ 1922 gegenüber $1.1e-3$ 1792). Weshalb übrigens der Höhepunkt von *fanatisch* im deutschen Korpus gerade nach Kriegsende und derjenige von *rücksichtslos* gerade davor liegt, wäre Gegenstand weiterer Analysen. Eine erste Sichtung des Korpus weist aus, dass die intensive Beschäftigung mit der NS-Zeit in Texten unmittelbar nach dem Krieg zumindest zum Teil für die andauernde Belegdichte von *fanatisch* im Zeitraum 1945–1951 verantwortlich ist.

Texten noch vor (Stand 2000: 6.5e-5), allerdings ausschließlich mit einem negativen Wert. Die jeweilige Halbwertszeit (»half-life«, Michel/Shen/Aiden u. a. 2011: 179) gegenüber 1945 bzw. 1943 erreichen die Adjektive *fanatisch* und *rücksichtslos* um 1970 (Stand 2000 für *rücksichtslos*: 2.5e-4).

3.3 Schwankende Häufigkeiten in der Nachkriegszeit

Nicht bei allen Ausdrücken weist die Graphik die charakteristische Form auf, die außer einer Zunahme der Häufigkeit in der NS-Zeit auch einen relativ konstanten Rückgang der Häufigkeit in den Nachkriegsjahren impliziert. Aus den Diagrammen 5 (*Großoffensive*), 6 (*Eintopf*), 7 (*überrollen*) und 8 (*kämpferisch*) geht hervor, dass bestimmte Ausdrücke, deren häufige Verwendung in der NS-Zeit einsetzt, ohne größere Unterbrechungen bis heute gebräuchlich geblieben sind oder sogar noch häufiger geworden sind.

Aus diesen vier Diagrammen geht hervor, dass die Geschichte von Wörtern wie *Großoffensive* und *überrollen*, die erst in der NS-Zeit häufiger werden, relativ bruchlos in die Nachkriegszeit übergeht und bis zum heutigen Tag andauert (höchste Belegdichte *Großoffensive*: 5.2e-5 1994 und *überrollen*: 2.9e-5 1993). Bei Wörtern wie *Eintopf* und *kämpferisch* ist die Entwicklung zwar weniger geradlinig verlaufen, aber dass diese Ausdrücke auch in die Gegenwartssprache Eingang gefunden haben, ist allemal klar, und zwar nicht nur nach der Belegdichte zu urteilen (höchster Wert *Eintopf*: 4.3e-5 1988 und *kämpferisch*: 1.2e-4 1943, vgl. 1.0e-4 1983), sondern auch, wenn man sich die Belege unter semantischem Gesichtspunkt ansieht, weil nämlich gar kein Bezug zur NS-Zeit vorhanden zu sein braucht.¹¹ Das unterscheidet diese Ausdrücke einerseits von einem Wort wie *Volkskanzler*, andererseits von *Ostarbeiter* (Diagramme 9 und 10).

Was diese beiden Graphiken (S. 92) gemeinsam haben, ist, was man als eine Latenzzeit im Gebrauch der Lexeme *Volkskanzler* und *Ostarbeiter* nach der NS-Zeit bezeichnen könnte. Nach ihrem sprunghaften Aufkommen in der historisch einschlägigen Phase der Vorkriegsjahre und ihrem fast ebenso jähem Rückgang im Sprachgebrauch nach dem Krieg folgt eine Periode unterschiedlicher Länge (in diesem Fall 3 bzw. 5 Jahrzehnte), in der die Ausdrücke relativ selten sind. 1963 enthält das Korpus für *Volkskanzler* aber wieder genauso viele Belege wie 1934 (4.5e-5), und auch die Belegdichte von *Ostarbeiter* steigt nach einer Zeit wieder bedeutend an: 1943: 5.5e-4, 1974: 2.0e-5, 1999: 2.1e-4.¹² Der Grund für die erneute Hausse der beiden Ausdrücke *Volkskanzler* und *Ostarbeiter* ist aber zweimal ganz verschieden. Einerseits ergab es sich so, dass nicht nur Hitler in den 1930er Jahren, sondern auch Ludwig Erhard in den 1960er Jahren »Volkskanzler« genannt wurde, offenbar ohne dass diese Bezeichnung in den 60er Jahren als ein NS-besudeltes Epitheton empfunden wurde. Andererseits geht aus dem Korpus hervor, dass die steigende Belegdichte von *Ostarbeiter* in den 1990er Jahren so gut wie ausschließlich auf das Konto einer Fülle

¹¹ Zu den Ausdrücken, die vergleichbare Entwicklungslinien aufweisen, gehören u. a. *schlagartig*, *Sportstudent*, *Umbruch*, *Vorfeld*.

¹² Ähnliche Ergebnisse zeitigt »Ngram Viewer« für *Engpässe*, *Schneckentempo*, *sichergestellt* und *volksnah* sowie, wenn auch in weniger ausgesprochenem Maße, *ausradieren* und die ehemals viel diskutierten Wörter *betreuen/betreut/Betreuung* (s. u. a. Klemperer ³1957: 290, Sternberger/Storz/Süskind ³1968: 31–36, Schmitz-Berning 2000: 89–94 und Polenz 1999: 551). Die Zeit wird zeigen, ob ein Ausdruck wie *lebensfroh* (Klemperer ³1957: 150), dessen Häufigkeit seit dem Anfang der 1990er Jahre anzusteigen scheint, eine ähnliche Entwicklung durchmachen wird.

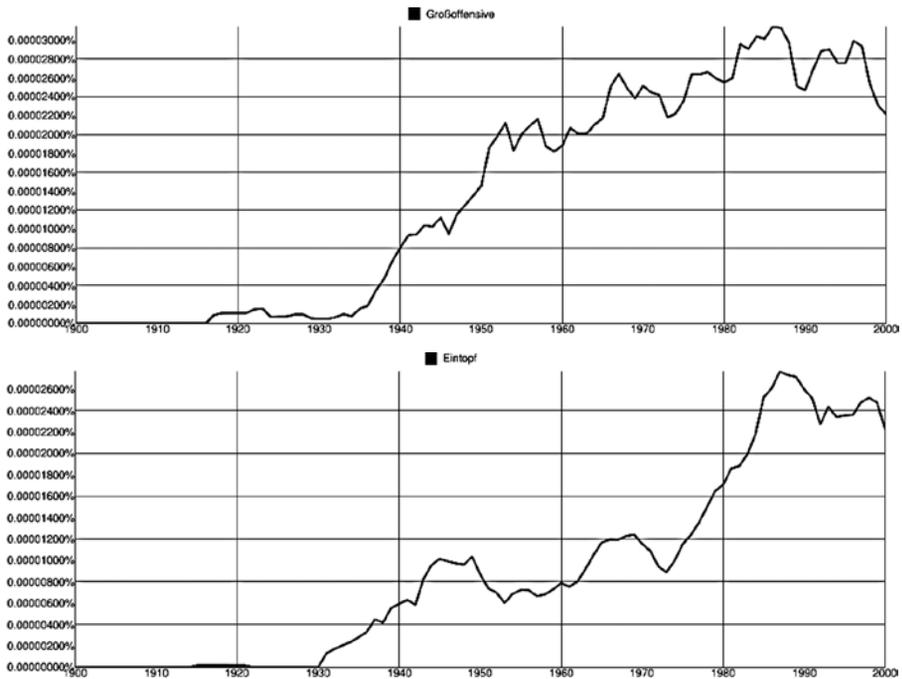


Diagramme 5 und 6: Die Häufigkeit der Nomen *Großoffensive* (5) und *Eintopf* (6) zwischen 1900 und 2000 (Smoothing: 3)

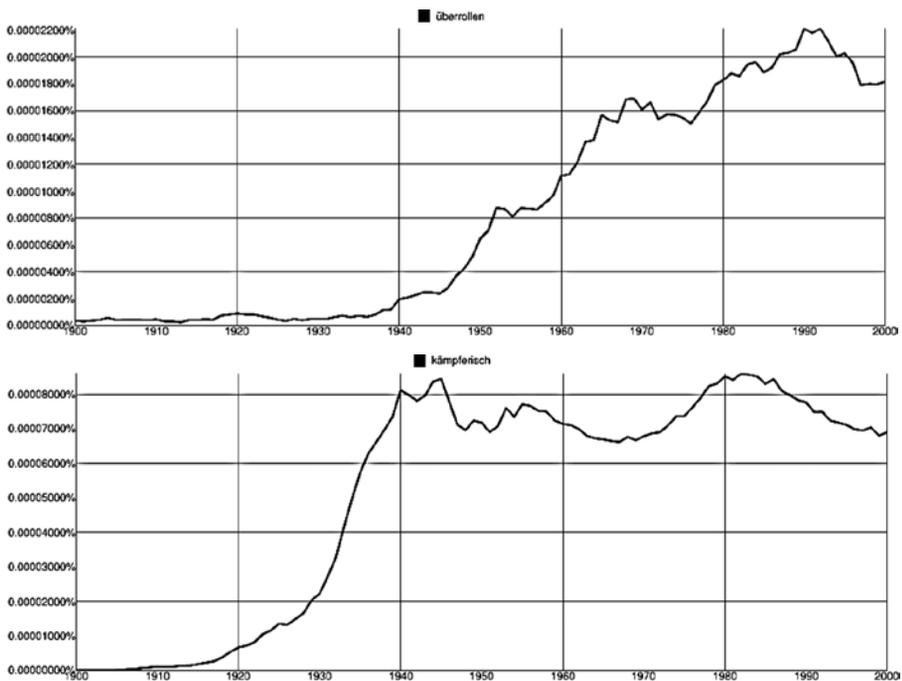


Diagramme 7 und 8: Die Häufigkeit des Verbs *überrollen* (7) und des Adjektivs *kämpferisch* (8) zwischen 1900 und 2000 (Smoothing: 3)

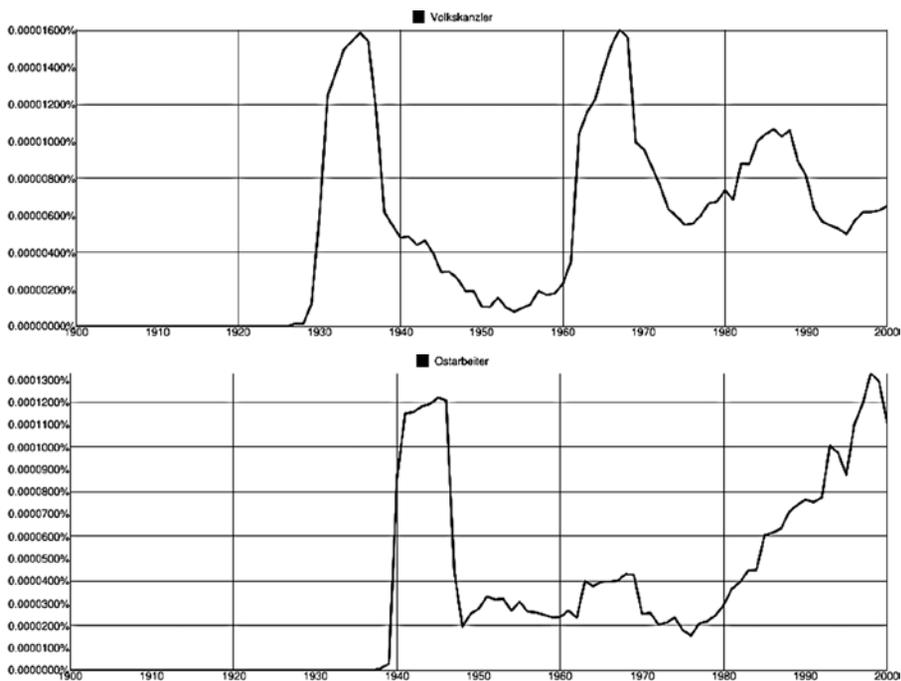


Diagramme 9 und 10: Die Häufigkeit der Nomen *Volkskanzler* (9) und *Ostarbeiter* (10) zwischen 1900 und 2000 (Smoothing: 3)

von historischen Betrachtungen über die Arbeiterpolitik der NS-Zeit geht.¹³ Wir lernen aus diesem Vergleich vor allem, dass »Ngram Viewer« aufgrund von ganz unterschiedlichen Wortverwendungen sehr ähnliche Graphiken hervorbringen kann und dass eine zusätzliche, genaue qualitative Analyse der Belege erforderlich ist, damit die quantitativen Ergebnisse auch inhaltlich richtig interpretiert werden.

3.4 Favorisierte und disfavorisierte Ausdrücke in der NS-Zeit

Genauso interessant wie die geläufigen Ausdrücke der LTI, die in den Texten eindeutig bevorzugt werden, sind natürlich diejenigen Ausdrücke (Wörter, Phrasen, Eigennamen usw.), die in der nationalsozialistischen Zeit absichtlich nicht oder so wenig wie möglich gebraucht wurden. So weist Klemperer (³1957: 122–123) darauf hin, dass das Wort *Philosophie* von den Nationalsozialisten geradezu »totgeschwiegen« und durchgängig durch *Weltanschauung* ersetzt worden sei (vgl. auch Klemperer ³1957: 175–182). Auch gegenüber dem Wort *System* habe eine tiefe Abneigung bestanden, ähnlich wie gegenüber *Intelligenz* und *Objektivität*; das Wort *System* wurde Klemperer zufolge aber dennoch häufig gebraucht, wenn auch im negativen Sinne (Klemperer ³1957: 122). Eine Korpusabfrage anhand von »Ngram Viewer« im Hinblick auf den Zeitraum 1900–1980 ergibt für die fünf gerade erwähnten Ausdrücke die folgenden Ergebnisse:

¹³ Ähnliches gilt für Ausdrücke wie *Volksempfänger* und *gleichschalten/Gleichschaltung* (s. Klemperer ³1957: 190), während ein Ausdruck wie *Heimatfront* in der zweiten Hälfte des 20. Jahrhunderts zwar vorrangig, aber nicht ausschließlich mit Referenz auf den Zweiten Weltkrieg verwendet wird.

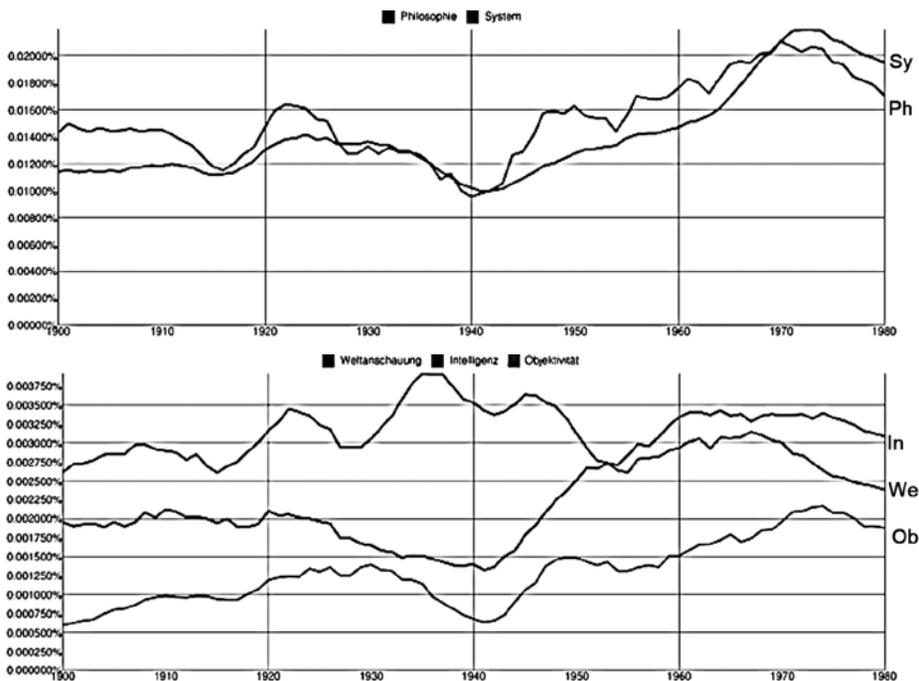


Diagramme 11 und 12: Die Häufigkeit der Nomen *System* (Sy) und *Philosophie* (Ph) (11) bzw. *Intelligenz* (In), *Weltanschauung* (We) und *Objektivität* (Ob) (12) zwischen 1900 und 1980 (Smoothing: 3)

Dass *System* zwischen 1921 ($1.6e-2$) und 1942 ($1.0e-2$) eine rückläufige Häufigkeit aufweist, stellt Diagramm 11 klar heraus, aber die Abnahme ist nur geringfügig. Auch die Häufigkeit von *Philosophie* nimmt im gleichen Zeitraum ab (1923: $2.1e-2$, 1942: $7.0e-3$). Noch bemerkenswerter aber ist die Feststellung, dass beide Belege schon kurz danach, sogar noch vor Kriegsende, wieder häufiger werden, um dann gegen 1960 aufs Neue den Stand zu Beginn der 1920er Jahre zu erreichen.

Auch Klemperers Beobachtungen über *Weltanschauung*, *Intelligenz* und *Objektivität* werden in Diagramm 12 weitgehend bestätigt, wobei aber auf ähnliche zeitliche Verschiebungen wie bei *Philosophie* und *System* hinzuweisen ist. Nachdem die Frequenz von *Weltanschauung* zwischen 1921 und 1930 ($2.5e-3$) sinkt, steigt sie ab 1930 rasch wieder an (die Spitze wird 1935 erreicht: $4.5e-3$), wonach das Wort erhebliche Häufigkeitsschwankungen aufweist; nach dem Krieg wird es wieder weniger gebräuchlich, bis es Anfang der 1970er Jahre auf den Stand der 1910er Jahre zurückfällt. Die Frequenz von *Intelligenz* nimmt von 1923 ($2.75e-3$) bis 1944 ($1.25e-3$) tatsächlich so gut wie ständig ab (ähnlich übrigens *Intellekt*, vgl. Schmitz-Berning 2000: 315–317), aber bemerkenswerterweise werden die Belege bereits ab 1944 wieder häufiger, eine Tendenz, die ohne Unterbrechungen bis weit über die letzten Kriegsjahre hinaus andauert (1961: $3.75e-3$). Eine vergleichbare Entwicklung lässt sich für das Wort *Objektivität* feststellen. Die genauen Gründe dafür, dass die im Diagramm 12 vorgeführten Daten Klemperer nur zum Teil Recht geben, müssten eigens erforscht werden. Klartext sprechen dann wieder die Diagramme 13 und 14, die darauf hindeuten, dass Klemperers Behauptung, zur NS-Zeit seien Namen wichtiger jüdischer Gelehrter wie *Einstein* und *Hertz* im Unterricht absichtlich so viel wie möglich verschwiegen worden (Klem-

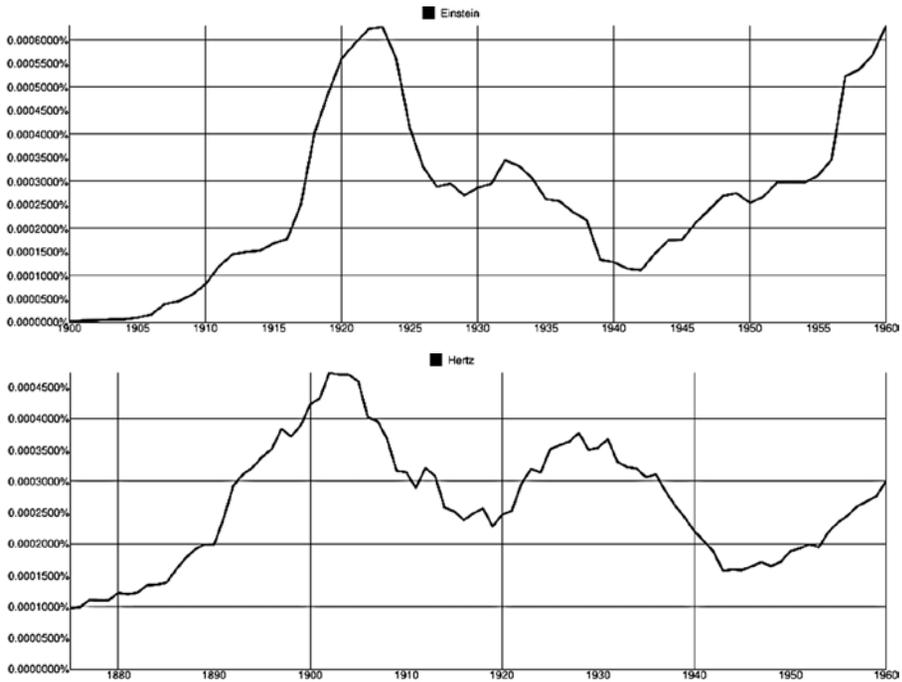


Diagramme 13 und 14: Die Häufigkeit von *Einstein* (13) zwischen 1900 und 1960 und *Hertz* (14) zwischen 1875 und 1960 (Smoothing: 3)

perer ³1957: 97), ohne weiteres zutrifft (vgl. *Einstein* 1921: 1.3e-3, 1945: 1.0e-4 und *Hertz* 1925: 5.0e-4, 1945: 1.2e-4):

Bei der Überprüfung von Klemperers Behauptungen muss man natürlich berücksichtigen, dass die erste Auflage seines Buches aus dem Jahre 1947 stammt, so dass verlässliche Aussagen über die Nachkriegsjahre noch kaum möglich waren. Immerhin kann man die LTI mittels »Ngram Viewer« gut mit dem Wortschatz im 19. Jahrhundert und den ersten Jahrzehnten des 20. Jahrhunderts vergleichen. Unter anderem hebt Klemperer (³1957: 67, 291) hervor, dass *spontan* eines der Lieblingswörter der LTI gewesen sei, aber Diagramm 15 vermittelt ein etwas nuancierteres Bild: Nachdem die Häufigkeit von *spontan* in der zweiten Hälfte des 19. Jahrhunderts stetig ansteigt (von 1.0e-4 um 1850 auf 8.5e-4 um 1900), nimmt die Zahl der Belege zwischen 1914 (9.5e-4) und 1942 (4.0e-4) beständig ab. Zwar tritt in den letzten Kriegsjahren eine Kehrtwendung ein, aber eine mit dem ausgehenden 19. Jahrhundert vergleichbare Häufigkeit des Ausdrucks wird erst ab Ende der 1950er Jahre wieder erreicht.

Interessanter als absolute Häufigkeiten sind freilich relative Häufigkeiten. So weist Klemperer (³1957: 216) z. B. darauf hin, dass im Sprachgebrauch der Nationalsozialisten das Adjektiv *jüdisch* häufiger gewesen sei als das Nomen *Jude*, was unsere Korpusanalyse aber nicht bestätigt. Auch hat man laut Klemperer (³1957: 278) in der NS-Zeit statt *Niederlage* lieber *Rückschlag* gesagt (»das klingt weniger definitiv«, Klemperer ³1957: 278), aus Diagramm 16 geht aber hervor, dass auch im Dritten Reich *Niederlage* häufiger verwendet wird als *Rückschlag*, zumindest in Büchern:

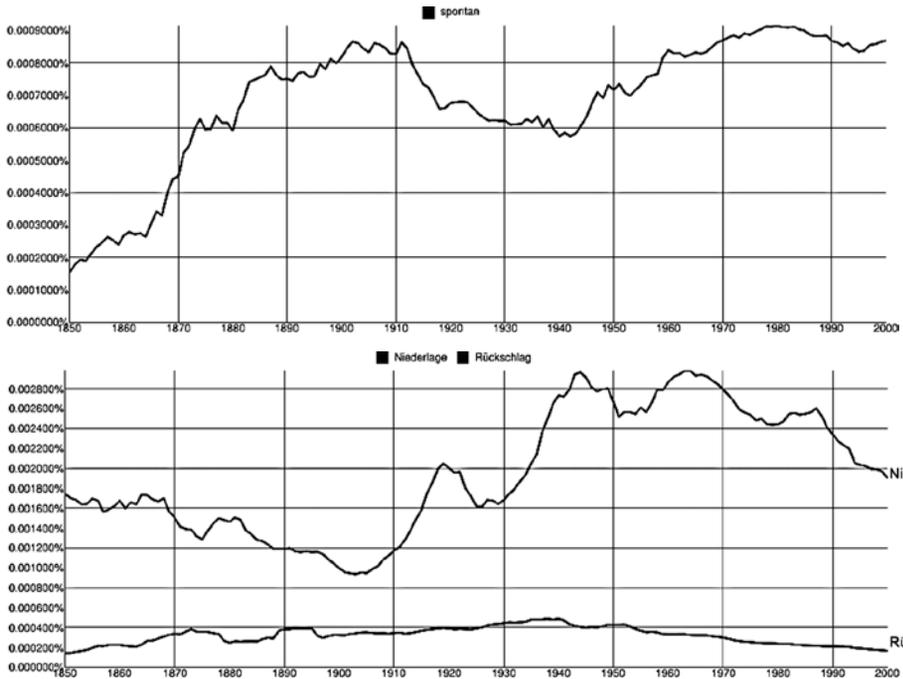


Diagramme 15 und 16: Die Häufigkeit des Adjektivs *spontan* (15) und der Nomen *Niederlage* (Ni) und *Rückschlag* (Rü) (16) zwischen 1850 und 2000 (Smoothing: 3)

Die Graphik in Diagramm 16 bedeutet freilich nicht unbedingt, dass Klemperers Einschätzung unzutreffend wäre. Detailanalysen sind erforderlich, um die genaue Distribution der beiden Ausdrücke zu bestimmen, denn es könnte z. B. sein, dass *Rückschlag* häufiger für Niederlagen auf deutscher Seite verwendet wurde und *Niederlage* für das, was den Gegnern zugefügt wurde. Auch das wäre eingehender an den Texten selbst zu erforschen.

3.5 Fremdwörter und Wortbildungen

An mehreren Stellen in seinem Buch weist Klemperer darauf hin, dass sich die LTI durch ein durchaus ambivalentes Verhältnis Fremdwörtern gegenüber auszeichne. Bald bediene sie sich »approximativer Verdeutschungen« (Klemperer ³1957: 308), so bevorzuge sie etwa *Bestallung* statt *Approbation*, *Entpflichtung* statt *Emeritierung*, *Belange* statt *Interessen*, *Menschlichkeit* statt *Humanität* (Klemperer ³1957: 306); bald greife sie »ohne Not zum Fremdwort« (Klemperer ³1957: 308). Klemperer stellt z. B. fest, dass Hitler »in jeder Rede, jedem Bulletin in zwei durchaus unnötigen und keineswegs überall verbreiteten und verstandenen Fremdwörtern« geradezu schwelge, nämlich: *diskriminieren* (»er sagt regelmäßig diskriminieren«) und *diffamieren* (Klemperer ³1957: 307).¹⁴ Daneben finde man ebenfalls sowohl *niedermachen* als auch *liquidieren* (Klemperer ³1957: 185).

¹⁴ Klemperer (³1957: 308–309) spricht in diesem Zusammenhang vom »tönenden Fremdwort«, das imponiere: »es imponiert um so mehr, je weniger es verstanden wird; in seinem Nichtbegriffenwerden beirrt und betäubt es, übertönt es eben das Denken. Schlechtmachen würde jedermann verstehen; diffamieren verstehen wenige«. Zu *diffamieren* s. auch Klemperer (³1957: 99, 238).

In den beiden folgenden Diagrammen werden die N-Gramm-Ergebnisse für zwei Wortpaare gezeigt. Diagramm 17 gibt die Entwicklungen der Klemperer zufolge miteinander in Konkurrenz stehenden Ausdrücke *Menschlichkeit* und *Humanität* wieder, Diagramm 18 kontrastiert das Fremdwort *liquidieren* mit *niedermachen*.

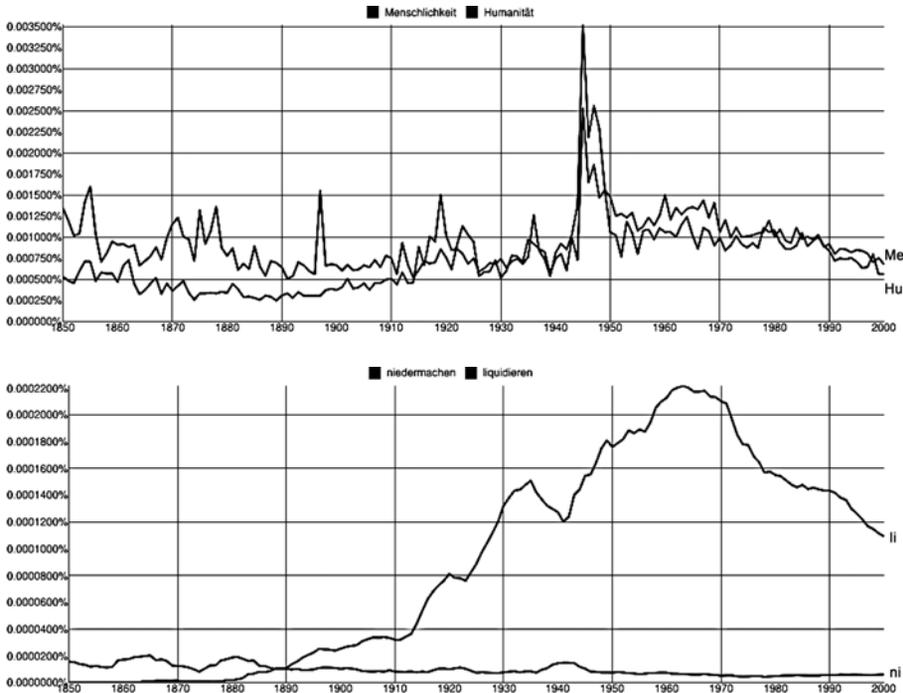


Diagramme 17 und 18: Die Häufigkeit der Nomen *Menschlichkeit* (Me) und *Humanität* (Hu) (17; Smoothing: 0) bzw. der Verben *liquidieren* (li) und *niedermachen* (ni) (18; Smoothing: 3) zwischen 1850 und 2000

Tatsächlich sind *Menschlichkeit* und *Humanität* zwischen 1915 und 1943 im Korpus etwa gleich häufig vertreten, nachdem *Humanität* im gesamten 19. Jahrhundert der leicht bevorzugte Ausdruck gewesen war. 1944 erleben beide Ausdrücke gleichzeitig einen ansehnlichen, aber nur sehr kurz andauernden Frequenzschub (1945 *Menschlichkeit*: $3.5e-3$, *Humanität*: $2.5e-3$). Bis Ende der 1970er Jahre ist *Menschlichkeit* dann der häufigere Ausdruck geblieben, allerdings ist die Differenz nur geringfügig; seit den 1980er Jahren sind beide Ausdrücke wiederum so gut wie gleich häufig. Aus Diagramm 18 geht hervor, dass *liquidieren* im Deutschen bereits seit 1890 gebräuchlicher ist als *niedermachen*. Ebenso ist klar, dass *niedermachen* zwischen 1939 und 1942 zeitweilig zunimmt (1939: $8.0e-6$, 1942: $2.2e-5$), während die Häufigkeit von *liquidieren* zwischen 1937 und 1944 zwar leicht abnimmt (von $1.6e-4$ auf $1.0e-4$), aber weiterhin weit über dem Wert von *niedermachen* liegt. Erst ab 1969 nimmt die Häufigkeit von *liquidieren* merklich ab, jedoch ohne, dass dies auf die Häufigkeit von *niedermachen* einen erkennbaren Effekt hätte.

Wie steht es nun um die Ausdrücke, die traditionell als Neuprägungen der Nationalsozialisten gelten? Wie oben erläutert, spricht Klemperer (³1957: 25) die Vermutung aus, dass so gut wie kein Ausdruck der LTI »selbstschöpferisch geprägt« worden sei, weil die Nationalsozialisten nur genutzt hätten, was schon gegeben gewesen sei (³1957: 144). Unter dem

Begriff »Neuprägung« sind in diesem Zusammenhang natürlich nicht so sehr Neologismen im strengen Sinne zu verstehen, die ohnehin im Wortschatz einer Sprache eine untergeordnete Rolle spielen, sondern Wortbildungen, die in der Regel nach produktiven Mustern gebildet werden. Einerseits soll Klemperers Einschätzung dahingehend nuanciert werden, dass immerhin eine erhebliche Anzahl von Ausdrücken, zumal diejenigen, die eine mehr oder weniger »technische« Bedeutung haben, durchaus als NS-Ausdrücke bezeichnet werden können (vgl. Schmitz-Berning 2000: xi und Polenz 1999: 549 »Organisations- und Verwaltungswörter«), so etwa *Atlantikwall*, *Blutschutzgesetz*, *gemeinschaftsunfähig*, *Hitlerjugend* (sowie die entsprechende Abkürzung *HJ*), *Ostarbeiter*, *Richterbriefe*, *Reichsautozug*, *Reichsfrauenführerin*, *Reichskulturkammer*. Andererseits ist es erstaunlich, in wie vielen Fällen Klemperer Recht behält. Nach »Ngram Viewer« zu urteilen, sind u. a. die folgenden Ausdrücke, die manch einer intuitiv bis heute zum NS-Vokabular rechnen dürfte, bereits vor 1918 in gedruckten Quellen belegbar: *Arbeitsschlacht*, *Blutfahne*, *blutmäßig*, *Jungvolk*, *Konzentrationslager*, *Landhilfe*, *Rassenbewusstsein*, *Reichsführer*, *Schutzhaft*, *Sondergericht* usw. (zu all diesen Ausdrücken s. die einzelnen Lemmata in Schmitz-Berning 2000; vgl. auch Zentner/Bedürftig 1985 und Michael/Doerr 2002).

4 Die Zuverlässigkeit von »Ngram Viewer«

Abschließend wenden wir uns der Frage zu, wie verlässlich die Korpusdaten sind, auf denen die statistischen Ergebnisse von »Ngram Viewer« beruhen. Dabei sind wir an zwei Aspekten vorrangig interessiert: Einerseits möchten wir herausfinden, ob das von »Ngram Viewer« verwendete optische Scanprogramm OCR (»optical character recognition«, Michel/Shen/Aiden u. a. 2011: 176) immer zuverlässige Ergebnisse liefert; andererseits soll stichprobenartig überprüft werden, ob die Erscheinungsjahre der eingescannten Bücher richtig erfasst bzw. dargestellt werden.

Zunächst zum Erscheinungsjahr. Wie »Ngram Viewer« genau mit Neuauflagen von Büchern verfährt, ist nicht ganz klar, aber es kann angenommen werden, dass zwischen Neuerscheinungen und Neuauflagen keine Unterscheidung getroffen wird (was im Hinblick auf die Zielsetzungen von »Ngram Viewer«, nämlich die Geschichte von Sprachzeichen in gedruckten Quellen wiederzugeben, auch ohne weiteres legitim sein dürfte). In anderer Hinsicht sind wir aber auf Fehler gestoßen, die die Ergebnisse der Korpusabfragen in einzelnen Fällen beeinträchtigen. Um die Erscheinungsjahre der eingescannten Publikationen zu kontrollieren, haben wir 20 Begriffe eingegeben, die wir gezielt Schmitz-Bernings *Vokabular des National-Sozialismus* (2000) entnommen haben und die als relativ genau datierbar gelten, um sie daraufhin sorgfältig auf ihre chronologische Erfassung durch das Programm hin zu überprüfen.¹⁵ Erwartungsgemäß stimmen die Resultate von »Ngram Viewer« in den meisten Fällen mit der historischen Wirklichkeit überein. Wider Erwarten signalisierte das Programm für die folgenden Ausdrücke jedoch Belege vor der NS-Zeit (wir erwähnen jeweils das früheste Belegjahr nach »Ngram Viewer«):

¹⁵ Es handelt sich um folgende Ausdrücke: *Atlantikwall*, *Blutschutzgesetz*, *Bonzokratie*, *Braunhemd(en)*, *deutschblütig*, *Feierabendgestaltung*, *Führergrundsatz*, *Gauleiter*, *Hakenkreuzfahne/-flagge*, *Hitlerjugend*, *Jugendwalter*, *Kinderlandverschickung*, *NSDAP*, *Reichsautobahn(en)*, *Reichsberufswettkampf*, *Reichskulturkammer*, *Schutzstaffel*, *Sippenamt*, *Staatsjugendtag*, *Sturmlokal*.

<i>Hitlerjugend:</i>	Erstbeleg 1873 in den <i>Mitteilungen der Deutschen Gesellschaft für Natur- und Völkerkunde Ostasiens</i> , Vol. 28–29; insgesamt 7 Belege zwischen 1873 und 1924; ¹⁶
<i>Reichsautobahn(en):</i>	Erstbeleg 1856 in <i>Westermanns Monatshefte</i> , Vol. 175–176; insgesamt 5 Belege zwischen 1856 und 1910; ¹⁷
<i>Kinderlandverschickung:</i>	Erstbeleg 1890 in <i>Die Neue Rundschau</i> , Vol. 98, Nr. 1–2;
<i>Schutzstaffel:</i>	Erstbeleg 1829 in <i>The Encyclopedia Americana</i> , Vol. 30;
<i>NSDAP:</i>	erste Belege ab 1800; insgesamt 46 Belege zwischen 1800 und 1919. ¹⁸

Dass diese Angaben fehlerhaft sind, ist offenkundig und geht zumeist auch sofort aus den Korpusdaten selbst hervor. Beispielsweise wird die *Aschaffener Zeitung: amtliches Organ der NSDAP und der Staats- und Gemeindebehörden* auf 1833 datiert, und Ähnliches trifft auf viele der anderen zu früh datierten Belege des Ausdrucks *NSDAP* zu. Der auf 1907 datierte Beleg von *Reichsautobahn* trägt als Quellenverweis *Law Foreign Dissertations*, die einschlägigen Seiten in »Google Books« verweisen aber unzweideutig auf Publikationen aus den 1920er und 1930er Jahren. Diese Teilanalyse bestätigt übrigens, dass die einseharen Belegstellen dem gesamten Korpus von »Google Books«, und nicht nur dem selektierten Teilkorpus entstammen (s. Abschnitt 3.1). Welches die Gründe dafür sind, dass bestimmte Quellen falsch datiert sind, ist oft unerfindlich.¹⁹ Auch ist unklar, in welchem Ausmaß die auf diese Weise generierten Phantombelege die quantitativen Grundlagen der Korpusrecherchen mittels »Ngram Viewer« beeinträchtigen, auch wenn der Einfluss solcher Falschdatierungen auf den ersten Blick nicht sehr groß erscheint, zumal sie im Prinzip reparabel sind, insofern falsch datierte Belege eliminiert werden können.

Schauen wir uns zum Schluss die Wortbildung *charakterlich* an, von der Klemperer (³1957: 236) wie bereits erwähnt behauptet, sie sei eine »Neuprägung der Nazis« (vgl. auch Schmitz-Berning 2000: 132–133 und Michael/Doerr 2002: 112 »adjective coined by the Nazis meaning the German character«, sowie Sternberger/Storz/Süstkind ³1968: 13, 37–44). Eine Auswertung anhand von »Ngram Viewer« erbringt das folgende Ergebnis.

Diagramm 19 zeigt, dass die Frequenz von *charakterlich* in der NS-Zeit zwischen 1928 (1.0e-5) und 1941 (1.3e-4) um das Dreizehnfache zunimmt. Dem Detaildiagramm 20 zufolge lässt sich das Adjektiv aber bereits viel früher im Korpus deutscher Texte nachweisen: »Ngram Viewer« generiert zwischen 1768 (Erstbeleg) und 1915 nicht weniger als 246 Belege. Diese frühen Belege erweisen sich aber mehrheitlich als Fehlanzeigen, und zwar nicht, weil die Quellen fehlerhaft datiert wären, sondern weil das OCR-Programm auch andere, dem Wort *charakterlich* ähnelnde Formen unter das Lemma *charakterlich* rubriziert. So ist beim ältesten »Beleg« (*Allgemeine deutsche Bibliothek*, Berlin/Stettin:

¹⁶ Die HJ wurde 1926 gegründet.

¹⁷ Von Reichsautobahnen ist erst ab den 1910er Jahren die Rede.

¹⁸ Die NSDAP wurde 1920 gegründet.

¹⁹ Hiervon auszunehmen sind nur gelegentlich auftretende Scanfehler, die darauf zurückzuführen sind, dass Erscheinungsjahre älterer Quellen, die manchmal schwer lesbar sind, mitunter falsch eingescannt werden; z. B. ist der 7. Jahrgang von Carl Caesar Leonhards *Taschenbuch für die gesammte Mineralogie* nicht 1815, sondern 1813 in Frankfurt erschienen.

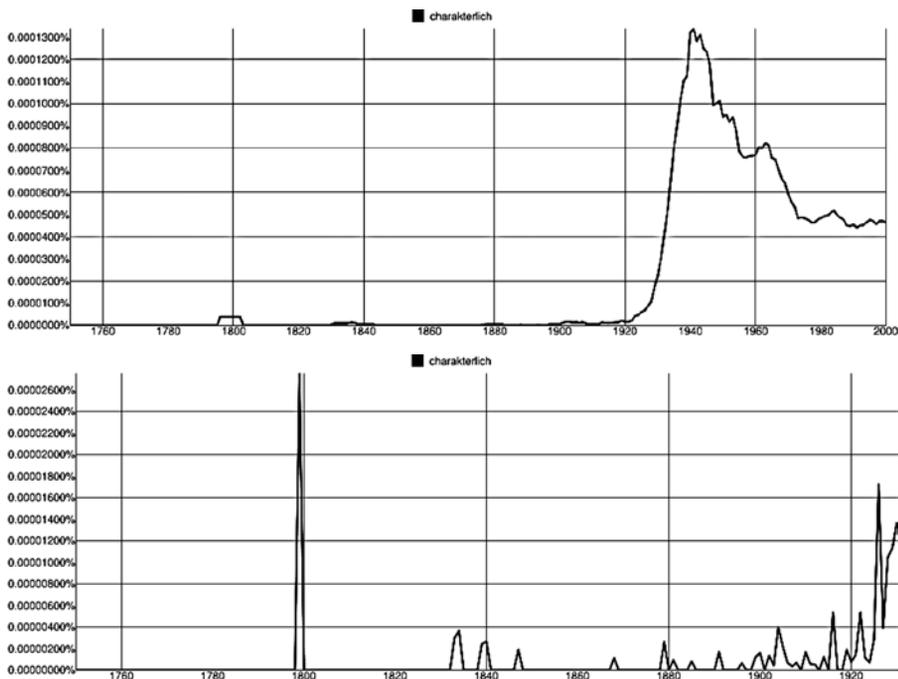


Diagramme 19 und 20: Die Häufigkeit des Adjektivs *charakterlich* zwischen 1750 und 2000 (19; Smoothing 3) und im Detail zwischen 1750 und 1930 (20; Smoothing 0)

Friedrich Nicolai, 1768: 58) ohne weiteres erkennbar, dass es sich um die Form *charakteristisch* handelt, und der zweitälteste »Beleg« (Johann Jakob Hess, *Geschichte der drey letzten Lebensjahre Jesu*. 3. Band, Zürich: Orell, Gessner, Füssli und Comp., 1773: 440) ist zweifelsfrei als *Charakterisches* lesbar (bis ins 19. Jahrhundert ist *charakterisch* in den Quellen ziemlich gebräuchlich). In noch anderen Fällen werden *Charakter auch* (1784), *Charakterzeich(-nung)* (u. a. 1790, 1795, 1797, 1831 und öfter) oder sogar zwei sich teilweise überschneidende Zeilen (»Bau und *Charakter/So reich* die Chinesische Schrift ...«, 1784) fälschlicherweise als *charakterlich* erfasst. Hieraus geht hervor, dass das optische OCR-Programm nicht fehlerfrei ist und noch verbessert werden muss. Allem Anschein nach ist die Tatsache, dass viele ältere deutsche Quellen in Fraktur gedruckt sind, der Hauptgrund für solche Mängel.²⁰ Immerhin taucht auch die richtige Form *charakterlich* in den 1890er Jahren hin und wieder in Texten auf, so u. a. im 5. Jahrgang (1895) der *Deutschen Zeitschrift für Nervenheilkunde*, im 21. Jahrgang (1899) der *Zeitschrift für französische Sprache und Literatur* in einem Artikel über Rousseau sowie im 24. Jahrgang (1899) der *Deutschen Revue über das gesamte nationale Leben der Gegenwart*. Auch Anfang des 20. Jahrhunderts kommt das Wort u. a. in Schriften zur Psychologie vor. Dass es sich um eine »Neuprägung« der Nationalsozialisten handeln würde, lässt sich mittels »Ngram Viewer«

²⁰ Vergleichbare Fehler kommen auch bei anderen Ausdrücken vor. U. a. wird *Schutzkraft* in naturwissenschaftlichen Quellen des 19. Jahrhunderts als *Schutzhaft* gelesen; die NS-Bezeichnung *ostisch* (insbesondere in der Fügung *ostische Rasse*) wird einerseits mit der nahezu homographischen Sprachbezeichnung *oskisch*, andererseits aber auch mit dem zweiten Teil wissenschaftlicher Begriffe wie *geognostisch*, *oryktognostisch* verwechselt usw.

somit als eine irrige Annahme zurückweisen, aber ebenso fehlerhaft wäre es, ohne detaillierte Kontrolle einfach das Ergebnis von »Ngram Viewer« zu übernehmen und zu glauben, das Wort *charakterlich* gäbe es schon im 18. oder frühen 19. Jahrhundert.

5 Schlussbemerkung

Dieser Beitrag hat sich nur mit einem Bruchteil der Ausdrücke beschäftigt, die als typisch für die NS-Zeit gelten, obwohl sie in vielen Fällen keine »Neuprägungen« aus dieser Zeit darstellen. Auf viele weitere Ausdrücke, die man Klemperers Buch sowie anderen Quellen entnehmen kann, wurde im vorliegenden Beitrag nicht eingegangen, u. a. bevorzugte Eigennamen (Klemperer ³1957: 93–97), bestimmte Satzzeichen (Klemperer ³1957: 90–92) usw. Der Beitrag dürfte indes ein Bild dessen vermitteln, wozu das neue Google-Tool »Ngram Viewer« für historische lexikologische Studien fähig ist und welche möglichen Probleme dabei berücksichtigt werden müssen.

Zwar haben verschiedene Autoren, darunter Victor Klemperer, bereits vor mehr als einem halben Jahrhundert erläutert, dass die Sprache der NS-Zeit keinen Bruch mit der Geschichte der deutschen Sprache darstellte. Nicht nur baute sie, zumal im Bereich der Wortbildung, auf dem bereits bestehenden Lexikon und System der Sprache auf, sondern viele Ausdrücke blieben auch nach dem Zweiten Weltkrieg im Wortschatz der Sprache erhalten (»unbewältigt«), wenn auch meist mit anderem semantischen Wert und geänderter Häufigkeit. Einerseits haben wir korrigierend darauf hingewiesen, dass bestimmte, vor allem mehr »technische« Teile des NS-Vokabulars (insbesondere Nomenklatur), durchaus ursprünglich auf die NS-Zeit zurückgehen. Andererseits ist klar geworden, dass die Geschichte von NS-Ausdrücken nach dem Zweiten Weltkrieg keineswegs einheitlich ist und aus unterschiedlichen Gründen stark variieren kann. Die Tatsache aber, dass man anhand von »Ngram Viewer« nunmehr im Stande ist, die Thesen Klemperers und vieler anderer Forscher über die Sprache der Nationalsozialisten auf breiter Korpusbasis zu überprüfen und ggf. zu nuancieren, zu bestätigen oder zu widerlegen, zeigt an einem einzigen Beispiel, dass das neue Google-Tool generell auch für Sprachwissenschaftler ein willkommenes und nützliches Instrument sein dürfte.

Internetquellen

- »Deutscher Wortschatz«, <http://wortschatz.uni-leipzig.de/>
- »Google Books«, <http://books.google.com/>
- »Ngram Viewer«, <http://books.google.com/ngrams>
- »Ngram Viewer: What's all this do?«, <http://books.google.com/ngrams/info>

Literaturverzeichnis

- Baayen, R. Harald (2001): *Word Frequency Distributions*. Dordrecht.
- Bork, Siegfried (1970): *Mißbrauch der Sprache. Tendenzen nationalsozialistischer Sprachregelung*. Bern u. a.
- Doerr, Karin (2002): »Nazi-Deutsch: An Ideological Language of Exclusion, Domination, and Annihilation.« In: Michael, Robert/Doerr, Karin (2002), S. 27–46.
- Ehlich, Konrad (Hg.) (1989): *Sprache im Faschismus*. Frankfurt am Main.
- Frind, Sigrid (1966): »Die Sprache als Propagandainstrument des Nationalsozialismus.« In: *Muttersprache* 76 (1966), S. 129–135.

- Klemperer, Victor (1947; ³1957): *LTI. Notizbuch eines Philologen*. Halle, Saale.
- Klemperer, Victor (1995): *Ich will Zeugnis ablegen bis zum letzten. Tagebücher 1933–1945*. 2 Bände. Berlin.
- Michael, Robert/Doerr, Karin (2002): *Nazi-Deutsch/Nazi German: An English Lexicon of the Language of the Third Reich*. Newport, CT.
- Michel, Jean-Baptiste/Shen, Yuan Kui/Aiden, Aviva Presser u. a. (2011): »Quantitative analysis of culture using millions of digitized books.« In: *Science*, 2011–14 January, 331 (2011), S. 176–182.
- Polenz, Peter von (1999): *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart*. Bd. III: *19. und 20. Jahrhundert*. Berlin/New York.
- Schmitz-Berning, Cornelia (2000): *Vokabular des Nationalsozialismus*. Berlin/New York.
- Seidel, Eugen/Seidel-Slotty, Ingeborg (1961): *Sprachwandel im Dritten Reich. Eine kritische Untersuchung faschistischer Einflüsse*. Halle, Saale.
- Sternberger, Dolf/Storz, Gerhard/Süskind, Wilhelm E. (1957; ³1968): *Aus dem Wörterbuch des Unmenschen. Neue erweiterte Ausgabe mit Zeugnissen des Streites über die Sprachkritik*. Hamburg/Düsseldorf.
- Wildt, Michael (2008): *Geschichte des Nationalsozialismus*. Stuttgart.
- Willems, Klaas (2011): »Essen oder fressen? Eine Fallstudie zum Verhältnis von Korpusdaten und Intuition in der lexikalischen Semantik.« In: *Deutsche Sprache* 39 (2011), S. 154–183.
- Zentner, Christian/Bedürftig, Friedemann (1985): *Das große Lexikon des Dritten Reiches*. München.
- Zipf, George K. (1935): *The Psycho-Biology of Language*. Cambridge.

Prof. Dr. Klaas Willems
Algemene Taalwetenschap
Universiteit Gent
Blandijnberg 2
9000 Gent
Belgien