

Zeitliche Verlaufskurven in den DTA- und DWDS-Korpora: Wörter und Wortverbindungen über 400 Jahre (1600–2000)

Alexander Geyken, Matthias Boenig, Susanne Haaf, Bryan Jurish,
Christian Thomas, Frank Wiegand, Kay-Michael Würzner
(Berlin-Brandenburgische Akademie der Wissenschaften)

Einführung

In diesem Beitrag werden zwei Referenzkorpora der deutschen Sprache verwendet, um daraus frequenznormierte Verlaufskurven für Wörter und Wortverbindungen zu berechnen: das DWDS-Kernkorpus des 20. Jhs. sowie das Referenzkorpus des Deutschen Textarchivs (1600–1900). Da beide Korpora bezüglich ihrer Metadaten vereinheitlicht und auch mit denselben linguistischen Informationen annotiert wurden, können korpusübergreifende Abfragen gestellt werden. Beispiele hierfür sind schreibweisentolerante Lemmasuchen oder textsortenspezifische Suchen. Die auf dieser Grundlage generierten Verlaufskurven stehen auf der Website des Deutschen Textarchivs für die Abfrage zur Verfügung.

1 Korpusgrundlage und Annotation

Die Grundlage für die zeitlichen Verlaufskurven (Histogramme) bilden die 100 Millionen Textwörter des DWDS-Kernkorpus des 20. Jhs. sowie weitere 140 Millionen Textwörter des Deutschen Textarchivs, welches Werke des 17. bis 19. Jh. als Erstausgaben umfasst.

Beide Korpora sind hinsichtlich der repräsentierten Textsorten aus Belletristik, Gebrauchsliteratur, Wissenschaft und (im DWDS:) Journalistischer Prosa sowie hinsichtlich der enthaltenen Disziplinen ausgewogen (Geyken 2007, 2013; Geyken et al. 2011). Sie wurden beide gemäß den TEI/P5-Richtlinien annotiert (Geyken et al. 2012; Haaf et al., forthcoming) und sind bezüglich der Metadaten untereinander interoperabel, insbesondere bezüglich der in den Histogrammen verwendeten Angaben zum Datum und zu den Textsorten.

Beide Korpora wurden für die Auswertung linguistisch annotiert, insbesondere wurden alle Texte tokenisiert, lemmatisiert, nach lexikalischen Kategorien analysiert (PoS-Tagging) und mit GermaNet-Kategorien versehen. Von besonderer Bedeutung für die Generierung der Histogramme ist die CAB-Analyse der historischen Texte: Mit CAB werden historische Varianten einer Wortform auf die wahrscheinlichste Normalform reduziert und ihrem neuhochdeutschen Lemma zugeordnet (Jurish 2013). Damit ist nicht nur die Suche in den Korpora, sondern auch die Darstellung der Wortverläufe schreibweisenübergreifend möglich. Dies geschieht, indem das Suchwort extensional zu all denjenigen Wortformen expandiert wird, die eine (möglicherweise flektierte) Variante des Suchwortes darstellen. In den

folgenden Abschnitten wird gezeigt, dass diese Vorgenerierung der möglichen Formen für den Benutzer eine effektive Hilfe darstellt (s. Abschnitt 3: Beispiele).

Beide Korpora, das DWDS-Kernkorpus und das DTA-Korpus, sind mit der Suchmaschine DDC (Dialing DWDS Concordancer) indiziert (Jurish et al. 2014). DDC verfügt über reichhaltige Metadatenfilter sowie die Möglichkeit, mehrere Annotationen auf einer Wortposition zu indizieren und abfragbar zu machen. Darüber hinaus können reguläre Ausdrücke, Boolesche Verknüpfungen und Abstandsoperatoren genutzt werden. Insbesondere ist es möglich, Metadatenfilter und mehrfache linguistische Annotationen miteinander zu verbinden. Die Indizes von DDC wurden so optimiert, dass Abfragen über die zeitliche Verteilung (Histogramme) ausreichend schnell für eine dynamische Berechnung zur Laufzeit sind. Dadurch ist es auch möglich, zeitliche Verläufe für die gesamte Mächtigkeit der DDC-Abfragesprache dynamisch zu berechnen.

2 Visualisierung

Grundlage der Visualisierung sind die nach Messpunkten (Jahreszahlen bzw. Datumsintervallen) normierten relativen Häufigkeiten pro Million Textwörter. Da die beiden Korpora weder gleich verteilt noch gleich groß sind und zudem die Wortanzahl je Zeitintervall variiert, ist die Normierung notwendig, um die Histogramme beider Korpora einheitlich zu präsentieren. Wie im vorigen Abschnitt erwähnt, können normierte relative Frequenzen nicht nur für einzelne Wortformen oder Lemmata, sondern auch für Kombinationen aus linguistischen Annotationen und Phrasensuchen ausgegeben werden. Darüber hinaus können Histogramme textsortenspezifisch (z. B. nur für Belletristik oder nur für Wissenschaft) oder textsortenübergreifend gebildet werden.

Da die Häufigkeiten von Wortformen (Types) in Textkorpora zipfsch verteilt sind, ist aufgrund der Größe der beiden Referenzkorpora bereits bei Wortformen einer mittleren Häufigkeit damit zu rechnen, dass die Histogramme bei einzelnen Messpunkten Nullstellen aufweisen können. Umgekehrt kann auch die Unausgewogenheit der Textkorpora an einem Messpunkt zu „Ausschlägen“ mit zu hoher Frequenz führen. Aus diesem Grund wurden in die Visualisierungskomponente verschiedene Parameter zur Glättung implementiert.

Aus Platzgründen soll hier nur auf die beiden wichtigsten eingegangen werden: die Parameter „window“ und „pruning“. Der Parameter „window“ gibt die Fensterbreite (als natürliche Zahl) für die Glättung nach dem gleitenden Mittelwert an. Der Parameter „pruning“ implementiert ein zweistufiges Verfahren. Im ersten Schritt wird eine Fehlerverteilung für die normierten Datenpunkte berechnet. Die beobachteten „Fehler“ werden unter Annahme einer Normalverteilung in p-Werte überführt, und alle Datenpunkte mit p-Werten außerhalb des angegebenen Konfidenzbereichs ($p=0.05$) werden als Ausreißer behandelt. Datenpunkte, die Ausreißern entsprechen, werden durch eine lineare Interpolation der nächstliegenden Datenpunkte, die innerhalb des Konfidenzbereichs liegen, ersetzt.

Die Visualisierung selbst erfolgt mittels der Javascript-Bibliothek Highcharts¹ und ist auf der Website des DTA abfragbar.²

3 Beispiele, Ergebnisse und Diskussion

Im folgenden, abschließenden Abschnitt sollen einige Vorzüge der Visualisierung auf der Grundlage der oben beschriebenen Referenzkopora und ihrer linguistischen Annotationen anhand konkreter Beispiele illustriert werden. Die Abfragemöglichkeiten und Ergebnisse werden mit den entsprechenden Ergebnissen im *Google Ngram Viewer*³ in Beziehung gesetzt.

Beispiel 1: *Tatsache*

Als erstes Beispiel dient das Lemma „Tatsache“, ein Begriff, der laut dem Etymologischen Wörterbuch (Pfeifer) erst mit einer Publikation aus dem Jahr 1756 Eingang in die deutsche Sprache fand.⁴ Die Wortverlaufskurve in *Abb._1* bestätigt den Beginn dieser „Wortkarriere“ und zeichnet ihn von der Mitte des 18. Jhs. bis in das 20. Jh. nach.⁵

Die Verlaufskurve im *Google Ngram Viewer*, *Abb._2*, zeigt eine vergleichbare Tendenz. Der Vergleich beider Histogramme lässt dennoch Probleme bei der Arbeit mit dem *Ngram Viewer* sichtbar werden: Historische Schreibweisen wie „Thatsache“ – sowie ggf. etliche weitere möglich Varianten⁶ und idealerweise auch alle möglichen Flexionsformen bzw. Expansionen – müssen bei der Eingabe der Anfrage explizit ergänzt werden, um die historische Entwicklung des Begriffs zu visualisieren. Auf den Punkt Flexionsformen/Expansionen wird im dritten Beispiel („billig“) noch näher eingegangen; hier sind zunächst nicht erklärbare Ausschläge der Kurve vor 1756 von Interesse, als Spalding das Wort in die deutsche Sprache brachte. Besonders deutlich zeigt sich darin ein Problem des *Google-Books*-Korpus. Eine Recherche in den Dokumenten daselbst zeigt, dass es sich bei sämtlichen früheren Treffern um falsch datierte Dokumente handelt, darunter so prominente Beispiele aus dem 19. Jh. wie [Goethes Schriften zur Morphologie \(II. Teil, datiert auf 1659\)](#) und die [Fliegenden Blätter \(in Google Books datiert auf 1692\)](#). Die Wortverlaufskurven werden hier also durch Metadatenfehler verfälscht.

¹ <http://www.deutschestextarchiv.de/search/plot>.

² <http://www.highcharts.com/products/highcharts>.

³ <https://books.google.com/ngrams>.

⁴ Vgl. z. B. das Etymologische Wörterbuch des Deutschen (nach W. Pfeifer), digitale Version via DWDS: „[...] nachgebildet (1756) von dem Theologen Spalding für engl. matter of fact [...]“. Siehe *Bestätigung der natürlichen und geoffenbarten Religion*, Leipzig, 1756, sowie Johann Joachim Spaldings Übersetzung von Joseph Butlers *The analogy of religion, natural and revealed, to the constitution and course of nature*, 1736. Die Textfassung dieses Werks ist für das DTA in Arbeit; derzeit stammt der früheste Beleg im DTA-Korpus aus Münter 1772.

⁵ Aus Platzgründen kann hier auf den Abfall der Kurve ab Mitte des 20. Jh. nicht eingegangen werden.

⁶ Im DTA-Korpus sind für den in dieser Hinsicht einfach erscheinenden Begriff „Tatsache“ immerhin 15 Expansionen belegt, vgl. <http://kaskade.dwds.de/dstar/dta/lizard.perl?q=Tatsache>.

Beispiel 2: *merkwürdig*

Das zweite Beispiel illustriert die Möglichkeiten, die eine Textsortendifferenzierung für die Interpretation der Histogramme bietet. Der Begriff „merkwürdig“ wurde Pfeifers Etymologischem Wörterbuch zufolge ab dem 19. Jh. vorrangig in der Bedeutung „seltsam, verwunderlich“ verwendet; bis dahin dominierte die seit dem 17. Jh. gebräuchliche Verwendung im Sinne von „bemerkenswert, bedeutsam“. *Abb._3* zeigt das nach Textsorten differenzierte Histogramm. Belegt ist die relativ hochfrequente Verwendung von „merkwürdig“ in den Textsorten Wissenschaft und Gebrauchsliteratur bis über die Mitte des 19. Jhs. hinaus. Folgt man der These (wie die im DTA verfügbaren Belege bestätigen), dass die Verwendung in der Wissenschaft und der Gebrauchsliteratur vorrangig im Sinne von „bemerkenswert, bedeutsam“ geschah, legt dies einen etwas länger andauernden Gebrauch des Wortes in dieser Bedeutung nahe, als bei Pfeifer angegeben. Auch hier zeigt sich der Vorteil gegenüber der Verlaufskurve des *Google Ngram Viewers* (*Abb._4*), wo die Textsortendifferenzierung fehlt.⁷

Beispiel 3: *billig*

Das abschließende Beispiel („billig“) illustriert zwei weitere Vorzüge der auf den Referenzkorpora beruhenden Histogramme gegenüber dem *Google Ngram Viewer*: die bessere Abdeckung der DTA-Korpora bezüglich des 17. Jhs. und die bessere Handhabung der großen graphematischen Varianz v. a. in diesen historischen Texten. *Abb._5* zeigt das Histogramm für „billig“ aus DTA und DWDS; die Kurve veranschaulicht die breite Verwendung des Begriffs im 17. Jh., die dazugehörigen Belege zeigen das Bedeutungsspektrum zwischen ‘angemessen, gerechtfertigt’ und ‘mäßig, wohlfeil, günstig’. Die Kurve aus dem *Google Ngram Viewer*, *Abb._6*, zeigt dagegen keinen kontinuierlichen Verlauf im 17. Jh., was angesichts der u. a. im DTA-Korpus belegten Verbreitung des Begriffs die verhältnismäßig geringe Substanz des *Google-Books*-Korpus für diesen Zeitraum belegt. Daher erscheint auf dessen Grundlage quellenbasierte Forschung zumindest in diesem Zeitraum kaum möglich. Ein weiteres, bereits angesprochenes Problem, kommt erschwerend hinzu: Insbesondere bei Abfragen für Zeiträume vor 1700 liefert das *Google-Books*-Korpus aufgrund der sehr heterogenen Graphie schlichtweg keine befriedigende Anzahl von Belegen; hier liefert die CAB-Analyse der DTA-Korpora klare Vorteile. Allein für das in dieser Hinsicht relativ unproblematisch erscheinende Lemma „billig“ sind im DTA-Korpus die in *Abb._7* gezeigten immerhin 67 Flexions- und Expansionsformen belegt. Bei einer Abfrage via DTA/DWDS werden alle diese Formen berücksichtigt, während der Nutzer des *Google Ngram Viewers* sie manuell eingeben (und zu diesem Zweck selbstverständlich überhaupt erst einmal präsent haben) müsste.

4 Ausblick

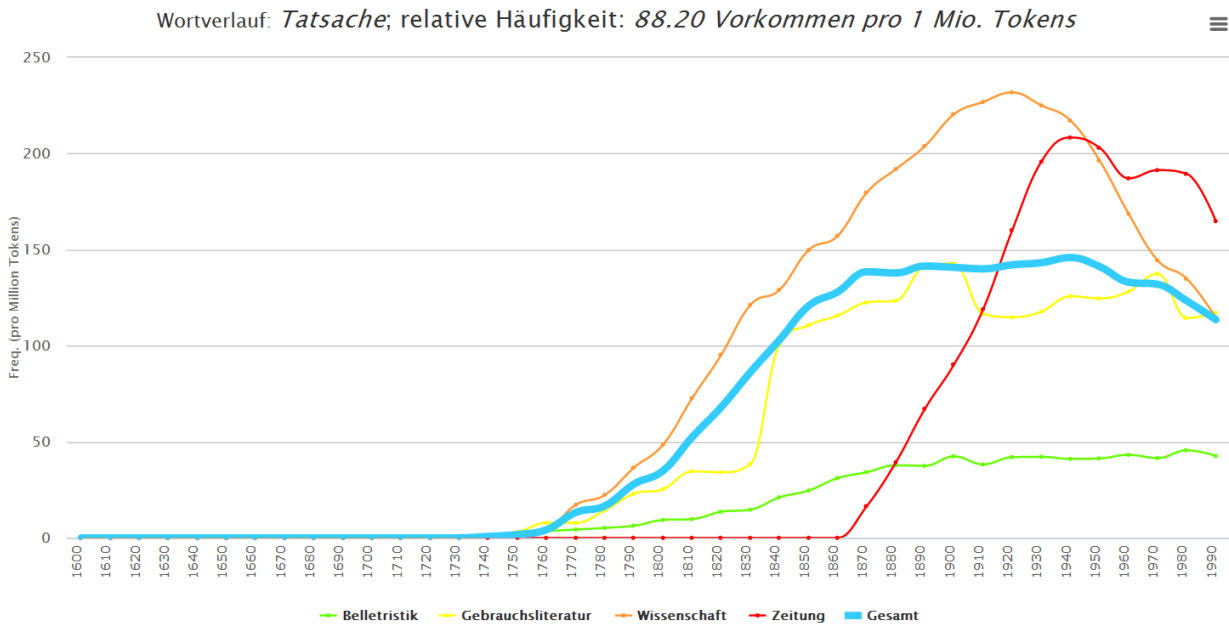
Wie im vorigen Abschnitt gezeigt werden konnte, liefert die Zeitverlaufskurve auf der Grundlage der beiden Referenzkorpora interessante Ergebnisse, die mit dem wesentlich

⁷ NB: Überraschend ist zudem, dass die Graphien „merkwürdig“ und „merckwürdig“ in dem der Kurve zugrundeliegenden Korpus *German* (das nicht identisch mit dem abfragbaren aktuellen GB-Korpus ist) nicht belegt sind.

größeren *Google-Books*-Korpus entweder nicht oder nur mit großem Rechercheaufwand ermittelbar gewesen wären. Grund dafür sind die verlässlichen Metadaten, die Zuordnung nach Textsorten sowie die aufgrund der genauen Texterfassung möglichen Erschließungsmethoden (CAB-Software).

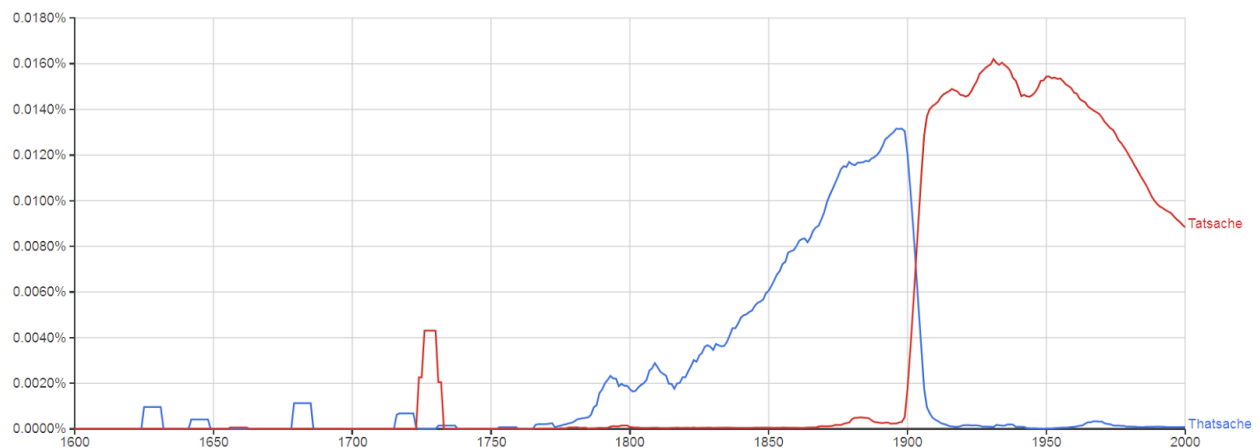
Es ist damit zu rechnen, dass sich die Lage der auf Referenzkorpora basierten zeitlichen Verlaufskurven künftig weiter verbessern wird. Zum einen liegt dies daran, dass immer mehr historische Volltexte in hoher Qualität entstehen. Diese Bemühungen werden dadurch verstärkt, dass die DFG erst unlängst eine spezifische OCR-Förderlinie aufgelegt hat. Zum anderen hat das DTA für die sich dynamisch verändernden Korpusgrundlagen bereits die geeigneten technischen Lösungen: Das Korpus des DTA wird automatisch im Wochenrhythmus indiziert.

Abbildungen⁸



<http://www.deutschestextarchiv.de/search/ddc/lemmata/?lemma=Tatsache&mode=extended:norm=date%2Bclass&smooth=spline&single=0&grand=1&slice=10&prune=1&window=3&wbase=0&logavg=0&logscale=0&xrange=1740%3A2000&totals=0>

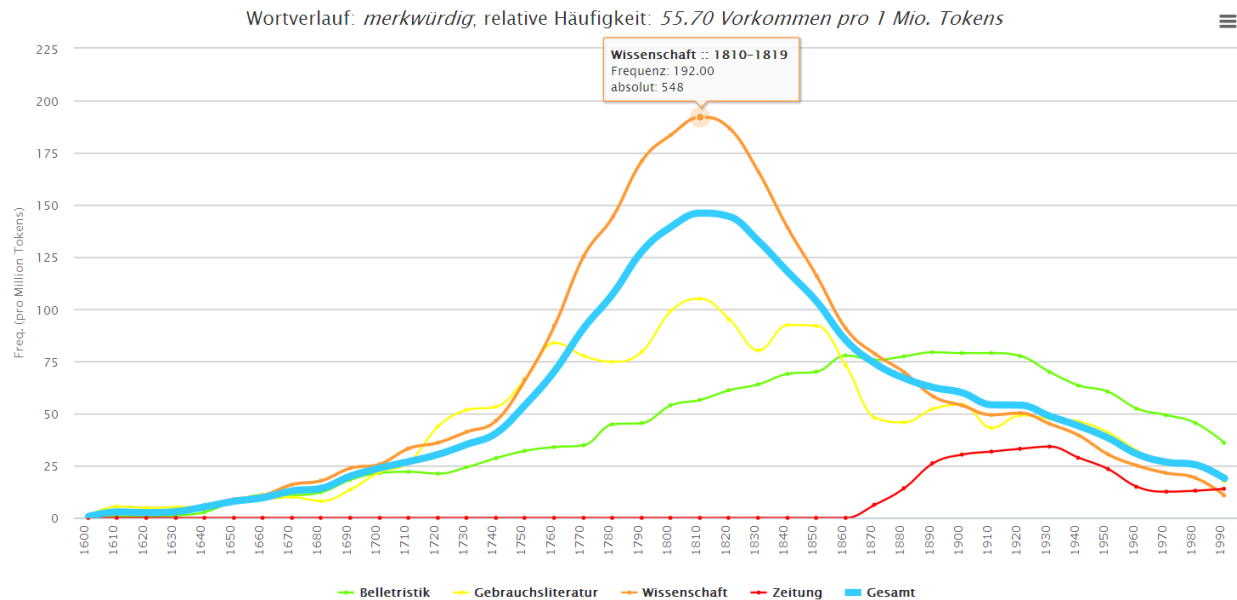
Abb._1: DTA-DWDS-Histogramm „Tatsache“



https://books.google.com/ngrams/graph?content=Thatsache%2CTatsache&year_start=1600&year_end=2000&corpus=20&smoothing=3&share=&direct_url=t1%3B%2CTatsache%3B%2Cc0%3B.t1%3B%2CTatsache%3B%2Cc0

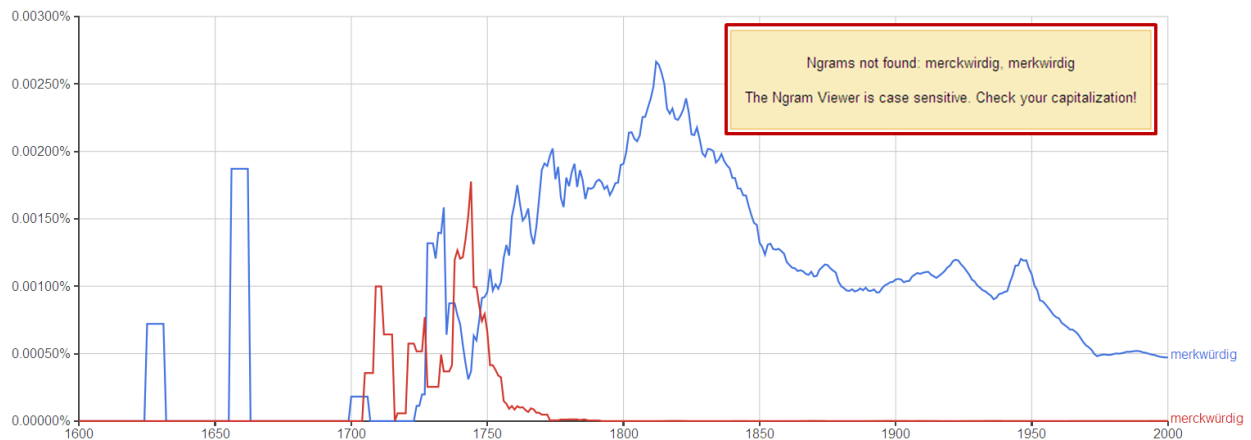
Abb._2: Google-Ngram-Histogramm „Tatsache, Thatsache“

⁸ Zu den gewählten Parametern der Kurvengenerierung siehe die jeweils angegebene URL.



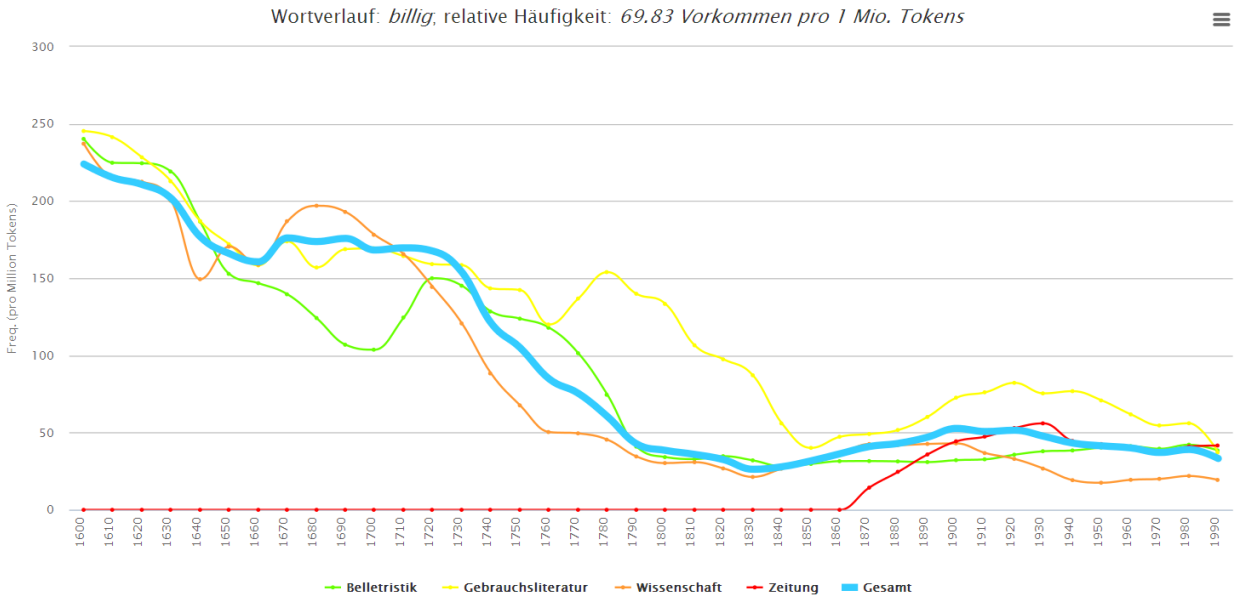
<http://www.deutschestextarchiv.de/search/ddc/lemmata/?lemma=merkw%C3%BCrdig&mode=extended;norm=date%2Bclass&smooth=spline&single=0&grand=1&slice=10&prune=1&window=3&wbase=0&logavg=0&logscale=0&xrange=1600%3A2000&total=0>

Abb._3: DTA-DWDS-Histogramm „merkwürdig“



https://books.google.com/ngrams/graph?content=merkw%C3%BCrdig%2Cmerckw%C3%BCrdig%2Cmerckwürdig%2Cmerkwirdig&year_start=1600&year_end=2000&corpus=20&smoothing=3&share=&direct_url=t1%3B%2Cmerkw%C3%BCrdig%3B%2Cc0%3B.t1%3B%2Cmerckw%C3%BCrdig%3B%2Cc0

Abb._4: Google-Ngram-Histogramm „merkwürdig,merckwürdig,merckwürdig,merkwirdig“



<http://www.deutschestextarchiv.de/search/ddc/lemmata/?lemma=billig&mode=extended;norm=date%2Bclass&smooth=spline&angle=0&grand=1&slice=10&prune=1&window=3&wbase=0&logavg=0&logscale=0&xrange=1600%3A2000&totals=0>

Abb._5: DTA-DWDS-Histogramm „billig“



https://books.google.com/ngrams/graph?content=billich%2Cbillig&year_start=1600&year_end=2000&corpus=20&smoothing=3&share=&direct_url=t1%3B%2Cbillich%3B%2Cc0%3B.t1%3B%2Cbillig%3B%2Cc0

Abb._6: Google-Ngram-Histogramm „billig“

<input checked="" type="checkbox"/> <u>Billich</u>	<input checked="" type="checkbox"/> <u>billich</u>	<input checked="" type="checkbox"/> <u>billicherem</u>	<input checked="" type="checkbox"/> <u>billigerer</u>
<input checked="" type="checkbox"/> <u>BJllich</u>	<input checked="" type="checkbox"/> <u>bil-lig</u>	<input checked="" type="checkbox"/> <u>billicherer</u>	<input checked="" type="checkbox"/> <u>billigeres</u>
<input checked="" type="checkbox"/> <u>BJllig</u>	<input checked="" type="checkbox"/> <u>billigen</u>	<input checked="" type="checkbox"/> <u>billichers</u>	<input checked="" type="checkbox"/> <u>billigern</u>
<input checked="" type="checkbox"/> <u>Billich</u>	<input checked="" type="checkbox"/> <u>bill'g</u>	<input checked="" type="checkbox"/> <u>billiches</u>	<input checked="" type="checkbox"/> <u>billiges</u>
<input checked="" type="checkbox"/> <u>Billiche</u>	<input checked="" type="checkbox"/> <u>bill'ge</u>	<input checked="" type="checkbox"/> <u>billichst</u>	<input checked="" type="checkbox"/> <u>billiglich</u>
<input checked="" type="checkbox"/> <u>Billicher</u>	<input checked="" type="checkbox"/> <u>bill'ger</u>	<input checked="" type="checkbox"/> <u>billichste</u>	<input checked="" type="checkbox"/> <u>billigs</u>
<input checked="" type="checkbox"/> <u>Billig</u>	<input checked="" type="checkbox"/> <u>billg</u>	<input checked="" type="checkbox"/> <u>billichsten</u>	<input checked="" type="checkbox"/> <u>billigst</u>
<input checked="" type="checkbox"/> <u>Billige</u>	<input checked="" type="checkbox"/> <u>billg'</u>	<input checked="" type="checkbox"/> <u>billig</u>	<input checked="" type="checkbox"/> <u>billigste</u>
<input checked="" type="checkbox"/> <u>Billiger</u>	<input checked="" type="checkbox"/> <u>billge</u>	<input checked="" type="checkbox"/> <u>billig'</u>	<input checked="" type="checkbox"/> <u>billigstem</u>
<input checked="" type="checkbox"/> <u>Billigere</u>	<input checked="" type="checkbox"/> <u>billgem</u>	<input checked="" type="checkbox"/> <u>billig's</u>	<input checked="" type="checkbox"/> <u>billigsten</u>
<input checked="" type="checkbox"/> <u>Billigeren</u>	<input checked="" type="checkbox"/> <u>billgen</u>	<input checked="" type="checkbox"/> <u>billige</u>	<input checked="" type="checkbox"/> <u>billigster</u>
<input checked="" type="checkbox"/> <u>Billigeres</u>	<input checked="" type="checkbox"/> <u>billich</u>	<input checked="" type="checkbox"/> <u>billigem</u>	<input checked="" type="checkbox"/> <u>billigstes</u>
<input checked="" type="checkbox"/> <u>Billiges</u>	<input checked="" type="checkbox"/> <u>billiche</u>	<input checked="" type="checkbox"/> <u>billigen</u>	<input checked="" type="checkbox"/> <u>billing</u>
<input checked="" type="checkbox"/> <u>Billigste</u>	<input checked="" type="checkbox"/> <u>billichem</u>	<input checked="" type="checkbox"/> <u>billiger</u>	<input checked="" type="checkbox"/> <u>pillichen</u>
<input checked="" type="checkbox"/> <u>Billigsten</u>	<input checked="" type="checkbox"/> <u>billichen</u>	<input checked="" type="checkbox"/> <u>billigere</u>	<input checked="" type="checkbox"/> <u>pillicher</u>
<input checked="" type="checkbox"/> <u>allerbillichsten</u>	<input checked="" type="checkbox"/> <u>billicher</u>	<input checked="" type="checkbox"/> <u>billigerem</u>	<input checked="" type="checkbox"/> <u>pillig</u>
<input checked="" type="checkbox"/> <u>allerbillichster</u>	<input checked="" type="checkbox"/> <u>billichere</u>	<input checked="" type="checkbox"/> <u>billigeren</u>	

<http://kaskade.dwds.de/dstar/dta/dstar.perl?fmt=expand-html&q=billig&x=Token>

Abb._7: im DTA belegte Expansions- und Flexionsformen des Lemma „billig“

Bibliographie

- Geyken 2007: Alexander Geyken: *The DWDS corpus – A reference corpus for the German language of the 20th century*. In: Fellbaum, Christiane (Hg.): *Idioms and Collocations: Corpus-based Linguistic, Lexicographic Studies*. London: Continuum Press, 2007, S. 23–40.
- Geyken et al. 2011: Alexander Geyken, Susanne Haaf, Bryan Jurish, Matthias Schulz, Jakob Steinmann, Christian Thomas, Frank Wiegand: *Das Deutsche Textarchiv: Vom historischen Korpus zum aktiven Archiv*. In: *Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland*, 20./21. September 2010. Beiträge der Tagung. Hrsg. von Silke Schomburg, Claus Leggewie, Henning Lobin und Cornelius Puschmann. 2., ergänzte Fassung. hbz, 2011, S. 157–161.
http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/Tagung_Digitale_Wissenschaft.pdf#page=159
- Geyken et al. 2012: Alexander Geyken, Susanne Haaf, Frank Wiegand: *The DTA 'base format': A TEI-Subset for the Compilation of Interoperable Corpora*. In: 11th Conference on Natural Language Processing (KONVENS) – Empirical Methods in Natural Language Processing, Proceedings of the Conference. Hrsg. von Jeremy Jancsary. Wien, 2012 (= Schriftenreihe der Österreichischen Gesellschaft für Artificial Intelligence 5).
http://www.oegai.at/konvens2012/proceedings/57_geyken12w/57_geyken12w.pdf
- Geyken 2013: Alexander Geyken: *Wege zu einem historischen Referenzkorpus des Deutschen: das Projekt Deutsches Textarchiv*. In: *Perspektiven einer corpusbasierten historischen Linguistik und Philologie*. Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch“ an der Berlin-Brandenburgischen Akademie der Wissenschaften, 12.–13. Dezember 2011. Hrsg. von Ingelore Hafemann, Berlin 2013, S. 221–234. [urn:nbn:de:kobv:b4-opus-24424](http://nbn-resolving.org/urn:nbn:de:kobv:b4-opus-24424)
- Haaf et al., forthcoming: Susanne Haaf, Alexander Geyken, Frank Wiegand: *The DTA 'Base Format': A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources*. To appear in: *Journal of the Text Encoding Initiative (jTEI)*, Issue 8. [Abstract des korrespondierenden Vortrags im Rahmen der Veranstaltung: TEI Conference and Members Meeting, 2.–5. Oktober 2013, Sapienza, Università di Roma (IT):
<http://digilab2.let.uniroma1.it/teiconf2013/program/papers/abstracts-paper#C137>]
- Jurish 2013: Bryan Jurish: *Canonicalizing the Deutsches Textarchiv*. In: *Perspektiven einer corpusbasierten historischen Linguistik und Philologie*. Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch“ an der Berlin-Brandenburgischen

Akademie der Wissenschaften, 12.–13. Dezember 2011. Hrsg. von Ingelore Hafemann, Berlin 2013, S. 235–244. [urn:nbn:de:kobv:b4-opus-24433](https://nbn-resolving.org/urn:nbn:de:kobv:b4-opus-24433)

Jurish et al. 2014: Bryan Jurish, Christian Thomas, Frank Wiegand: *Querying the Deutsches Textarchiv*. In: Proceedings of the Workshop MindTheGap 2014: Beyond Single-Shot Text Queries: Bridging the Gap(s) between Research Communities (co-located with iConference 2014, Berlin, Germany, 4th March, 2014). Hrsg. von Udo Kruschwitz, Frank Hopfgartner, Cathal Gurrin, Berlin 2014, S. 25–30.
http://ceur-ws.org/Vol-1131/mindthegap14_7.pdf

Münter 1772: Balthasar Münter: Bekehrungsgeschichte des vormaligen Grafen [...] Johann Friederich Struensee. Kopenhagen, 1772. In: Deutsches Textarchiv, www.deutschestextarchiv.de/muenter_bekehren_1772, abgerufen am 07.11.2014.