

## Ressourcen

Eva Gredel\*, Laura Herzberg\* & Angelika Storrer\*

# Linguistische Wikipedistik Linguistic Wikipedistics

<https://doi.org/10.1515/zgl-2018-0029>

## 1 Einleitung

Die Wikipedia ist nicht nur die größte Online-Enzyklopädie weltweit, sondern auch eines der erfolgreichsten Projekte im Web 2.0: In nur 16 Jahren sind rund 48 Millionen Einträge in 295 Sprachversionen entstanden (Wikimedia 2018). Mit Rang 5 des Alexa-Rankings ist die Wikipedia eine der meistgenutzten Plattformen im Internet (Alexa 2018). Durch ihre Relevanz und Reichweite wird die Wikipedia auch intensiv beforscht. Die Seite „Wikipedistik“ (WP-Wikipedistik; Wikipedia 2018) im Metabereich der deutschsprachigen Wikipedia gibt einen Überblick über nationale und internationale Forschungsaktivitäten und -ergebnisse. Die interessierten Disziplinen, die Erkenntnisinteressen und methodischen Zugänge der Wikipedistik sind vielfältig. Hammwöhner (2007) beschäftigt sich aus informationswissenschaftlicher Perspektive mit Methoden und Ergebnissen der Qualitätsbewertung von Wikipedia-Artikeln. Pscheida (2010) untersucht die Wikipedia unter wissenssoziologischer Perspektive und begründet am Beispiel der Wikipedia interessante Thesen zur „Wissenskultur des digitalen Zeitalters“ (Pscheida 2010: 458 ff.). Stegbauer (2009) untersucht das soziale Rollengefüge und die Motivation der Akteure in der deutschen Wikipedia und gibt einen empirisch sehr gut gestützten Einblick in die sozialen Prozesse im Projekt.

In diesem Beitrag geben wir einen Überblick über die aktuelle Forschung zur Wikipedia aus der Perspektive der Sprach- und Diskursanalyse. Zunächst (Abschnitte 2.1–2.4) verdeutlichen wir das Potenzial der Wikipedia als Forschungsgegenstand an vier Themenfeldern: Text und Interaktion, Diskurslinguistik,

---

**\*Kontaktpersonen:** Dr. Eva Gredel: Germanistische Linguistik, Universität Mannheim, Schloss, D-68131 Mannheim, E-Mail: [eva.gredel@phil.uni-mannheim.de](mailto:eva.gredel@phil.uni-mannheim.de)

Laura Herzberg: Germanistische Linguistik, Universität Mannheim, Schloss, D-68131 Mannheim, E-Mail: [herzberg@uni-mannheim.de](mailto:herzberg@uni-mannheim.de)

Prof. Dr. Angelika Storrer: Germanistische Linguistik, Universität Mannheim, Schloss, D-68131 Mannheim E-Mail: [astorrer@mail.uni-mannheim.de](mailto:astorrer@mail.uni-mannheim.de)

Multimodalität, Sprach- und Kulturvergleich. Der anschließende Abschnitt 2.5 „Wikipedaktik“ beschäftigt sich mit der Wikipedia als lohnenswertem Lerngegenstand in Schule und Hochschule. Wikipedia ist nicht nur interessant als Ressource, an der sich die Besonderheiten digitaler Diskurse, multimodaler Hypertexte und kollaborativer Schreib- und Aushandlungsprozesse gut verdeutlichen lassen. Es ist auch ein Projekt des freien Wissens, das seine Daten unter Creative-Commons-Lizenzen zur Verfügung stellt und es damit ermöglicht, sehr umfangreiche Korpora zu verschiedenen Sprachen und unterschiedlichen Schreibräumen aufzubauen. Abschnitt 3 dokumentiert den umfangreichen Bestand der Wikipedia-Korpusfamilie am Institut für Deutsche Sprache (IDS) und beschreibt, wie diese Korpora für die Sprach- und Diskursforschung linguistisch aufbereitet werden. Abschnitt 4 gibt einen zusätzlichen Einblick in digitale Methoden, die direkt auf den Daten und Metadaten der Wikipedia-Sprachversionen ansetzen und zusätzliche interessante Optionen der Auswertung bieten.

Um die Potenziale der Wikipedia für die Diskurs- und Sprachforschung deutlich zu machen, haben wir die dafür relevanten Komponenten der Wikipedia und deren Vernetztheit in Abbildung 1 visualisiert. Die Wikipedia existiert in verschiedenen Sprachversionen, deren Artikel durch sogenannte Interlanguage-Links miteinander verknüpft sind (Abb. 1, a). Die Wikipedia-Artikel der verschiedenen Sprachversionen sind in den meisten Fällen keine Übersetzungen (vgl. van Dijk 2009), sondern unterscheiden sich im Aufbau, im Inhalt und durch die externen Quellen, mit denen die enzyklopädischen Informationen belegt werden (Abb. 1, d). Dies macht die Wikipedia zu einer wertvollen Ressource nicht nur für den Sprach-, sondern auch für den Kulturvergleich (vgl. Abschnitt 2.4). Die enzyklopädischen Artikel sind der funktionale Kern des Projekts (Abb. 1, a). Es gibt aber weitere Schreib- und Interaktionsbereiche – als Namensräume bezeichnet – mit jeweils speziellen Funktionen im Projekt. So dient der Namensraum der Artikeldiskussionen beispielsweise dazu, Fragen zu diskutieren, die sich bei der gemeinsamen Arbeit an den Artikeln ergeben (Abb. 1, b). Die Gegenüberstellung von Artikeln und Artikeldiskussionen eignet sich sehr gut dafür, das Zusammenspiel von Text und Interaktion in sozialen Medien zu illustrieren (vgl. Abschnitt 2.1). Wikipedia-Artikel sind dynamische Hypertexte, die immer wieder geändert und aktualisiert werden können. Alle Versionen werden in der öffentlich einsehbaren Versionengeschichte (Abb. 1, c) gespeichert, dort kann man sie vergleichen oder neue Versionen durch frühere ersetzen. Die in den Versionengeschichten gespeicherten Daten lassen sich mit digitalen Methoden (vgl. Abschnitt 4) auswerten und bieten wertvolle Anhaltspunkte für die Untersuchung von Diskursdynamiken (vgl. Abschnitt 2.2).

Wikipedia-Artikel sind multimodale Hypertexte, die vor allem Bilder, aber auch andere mediale Objekte wie Audio- und Videodateien enthalten. Diese

sind meist nicht direkt in der Datenbasis der jeweiligen Wikipedia-Sprachversion gespeichert, sondern in der Datenbasis des Schwesterprojekts „Wikimedia Commons“ (Hammwöhner 2013, Abb. 1, e). Dies hat den Vorteil, dass die Medienobjekte nur einmal kategorisiert und rechtlich geprüft werden müssen und dann in alle Sprachversionen der Wikipedia (und andere Wiki-Projekte der Wikimedia-Foundation) eingebunden werden können. Hierdurch eröffnen sich interessante Möglichkeiten, Multimodalität in Texten und Diskursen auch sprachvergleichend zu untersuchen (vgl. Abschnitt 2.3 und 2.4).

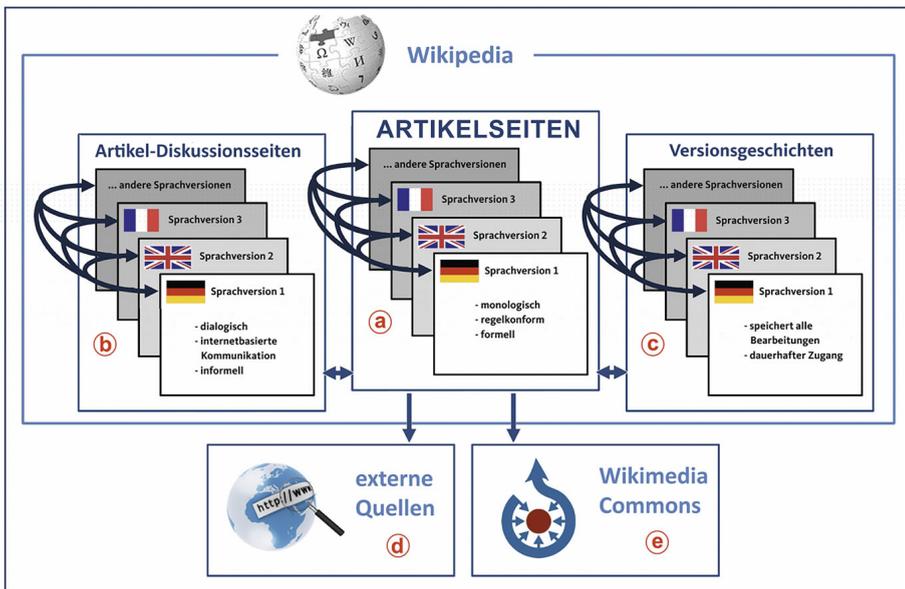


Abb. 1: Struktur der Wikipedia (Eigene Darstellung)

## 2 Linguistische Wikipedistik: Forschungsfelder und Potenziale

Im Folgenden möchten wir an vier Bereichen der Sprach- und Diskursforschung erläutern, welches Potential die Wikipedia als Forschungsgegenstand hat. Wir verweisen dazu im Detail auf publizierte bzw. im Druck befindliche Studien, die u. a. die in den Abschnitten 3 und 4 beschriebenen Korpora und Methoden nutzen.

## 2.1 Text und Interaktion

Im Internet wird Schriftsprache nicht nur für die Gestaltung von Texten genutzt, sondern auch für die Interaktion zwischen den Nutzern des „Social Web“. Dabei entwickelt sich eine neue interaktionsorientierte Haltung zum Schreiben, die sich in wichtigen Merkmalen vom planvollen Verfassen redigierter Schrifttexte, dem textorientierten Schreiben, unterscheidet. Beim interaktionsorientierten Schreiben steht das Gelingen der laufenden sozialen Interaktion im Fokus; beim textorientierten Schreiben das Verfassen eines Textes, der ohne weitere Interaktion zwischen Autor und Leser aus sich heraus verständlich ist (vgl. Storrer 2018b). An der Wikipedia kann man die Unterschiede zwischen text- und interaktionsorientiertem Schreiben gut verdeutlichen: Artikeltexte präsentieren Wissen zum Artikelgegenstand so, dass sich Leser schnell und ohne direkte Rückfragen Informationen erschließen können. Wichtige Merkmale des textorientierten Schreibens – das reflektierte Planen, Strukturieren und Redigieren – sind für die Arbeit an den Artikeln charakteristisch, auch wenn die Wikipedia als dynamischer Hypertext nicht auf einen Abschluss des Produktionsprozesses, sondern auf kontinuierliche Aktualisierung und Optimierung der Artikel hin ausgerichtet ist. Artikeldiskussionen demgegenüber dienen den Nutzern dazu, sich über die Verbesserung und Aktualisierung der Artikel auszutauschen und strittige Aspekte zu diskutieren. Hierbei wird überwiegend in einer interaktionsorientierten Grundhaltung geschrieben, bei der das Gelingen des aktuellen Handlungsziels – die Verbesserung des Artikels, die Diskussion um strittige Formulierungen oder die Zuverlässigkeit von Quellen – im Mittelpunkt steht. Die Unterschiede zwischen text- und interaktionsorientierter Schreibhaltung schlagen sich in den sprachlichen Formen und Strukturen nieder. Dies macht die in Abschnitt 3 beschriebenen linguistisch aufbereiteten Korpora mit Wikipedia-Artikeln und Artikeldiskussionen zu interessanten Ressourcen für die korpusbasiert arbeitende Varietätenlinguistik. Kupietz et al. (2014) nutzen Korpora mit Artikeln und Diskussionen aus der Wikipedia-Korpusfamilie des IDS für eine Vergleichsstudie zu Phänomenen, die in der populären Sprachkritik als Sprachverfallserscheinungen diskutiert werden (schwache Präteritumsformen bei starken Verben, attributive Verwendung von Adverbien auf *-weise*, analytische Konjunktivformen). Storrer (2017) nimmt das Korpus mit deutschen Wikipedia-Artikeldiskussionen des IDS als Datengrundlage, um die Verwendung der Interjektion *hm* (in allen Schreibvarianten) beim interaktionsorientierten Diskutieren mit der Verwendung in Chats und in literarischen Werken zu vergleichen. In zwei weiteren Fallstudien vergleicht Storrer (i. Dr.) Frequenz und Verwendungsmuster der Konnektoren *weil*, *d. h.* und *sprich* im Artikelkorpus mit denen im Korpus der Artikeldiskussionen und kontrastiert die Ergebnisse mit Beschreibungen zur Verwendung in redigierten Texten und mündlichen Gesprächen.

Das Zusammenspiel von Text und Interaktion ist nicht nur für empirische Studien zur diamesischen (also mediumsabhängigen) Variation relevant. An der Wikipedia lassen sich sehr gut typische Merkmale des Social Web verdeutlichen (vgl. z. B. Marx/Weidacher 2014: 66 f., Storrer 2018b). Weiterhin bildet die Wikipedia eine hervorragende empirische Grundlage dafür, Charakteristika und Praktiken der schriftbasierten Interaktion herauszuarbeiten (Beißwenger 2016) und Prozesse des gemeinschaftlichen Schreibens aus der Perspektive der Schreibprozessforschung (Kallass 2015) bzw. mit computerlinguistischen Verfahren zu untersuchen (Ferschke et al. 2013).

## 2.2 Die diskurslinguistische Perspektive auf Wikipedia

Als relativ neue Teildisziplin der Linguistik geht die Diskurslinguistik der Frage nach, wie soziale Wirklichkeiten in transtextuell organisierten Einheiten konstruiert werden. Bereits vorgelegte diskursanalytische Arbeiten zur Wikipedia aus den Medien- und Kommunikationswissenschaften (Pentzold 2007, Beyersdorff 2011) und aus der Linguistik (Liebert/Kohl 2004, Arendt/Dreesen 2015, Gredel 2016, Gredel 2017) zur Wikipedia zeigen ihr großes Potential als Untersuchungsgegenstand diskursanalytischer Studien.

Zentral für digitale Diskursanalysen zur Wikipedia ist es, die Artikelseiten nicht isoliert als Diskursfragmente zu betrachten. Zielführend ist die Zusammenschau der verknüpften Artikel- und Diskussionsseite sowie der Versionsgeschichte zu einem Thema: Stellen die Artikelseiten das Produkt diskursiver Aushandlung dar, wird über die Diskussionsseiten der Prozess diskursiver Aushandlung in der sozialen Interaktion der Wikipedia-Autoren deutlich (Gredel 2016). Dies legt für die Analyse digitaler Diskurse in der Wikipedia die Integration diskurs- und interaktionslinguistischer Ansätze nahe (Gredel 2017), um die Aushandlungsprozesse auf den Diskussionsseiten adäquat analysieren zu können.

Die Versionsgeschichten erlauben es zudem, diskursive Dynamiken im digitalen Diskurs der Wikipedia zu rekonstruieren: So lässt sich über die Versionsgeschichte zum englischen Eintrag *Annexation of Crimea by the Russian Federation* zeigen, dass Autoren im Rahmen eines Edit-Wars den Begriff *Annexation* immer wieder gegen alternative Termini wie *Incorporation* oder *Accession* austauschen, um mit diesen konkurrierenden Termini verschiedene Sichten zu den Ereignissen auf der Krim 2014 sprachlich zu konstruieren (Gredel 2017: 106). Digitale Diskurse in der Wikipedia zeichnen sich zudem durch ihre multimodale und multilinguale Dimension aus, was anhand multimodaler Metaphern in verschiedenen Sprachversionen der Wikipedia zum Lemma *Burnout* sichtbar gemacht werden kann (Gredel 2018a).

## 2.3 Wikipedia aus Sicht der Multimodalitätsforschung

Im Rahmen der kollaborativen Wissensproduktion in der Wikipedia ist es möglich, Bild-, Audio- und Videomaterial zu integrieren. Für Sprach- und Diskursanalysen ergibt sich daraus die Notwendigkeit, die multimodalen Textrealitäten der Online-Enzyklopädie zu berücksichtigen, um alle Aspekte der kollaborativen Textproduktion adäquat zu erfassen.

Systematisierend stellt Hammwöhner für die Wikipedia und für Wikimedia Commons fest, dass „jedes Bild [...] in eine Vielfalt von Verwendungszusammenhängen gestellt“ (2013: 290) werden kann, woraus sich komplexe Bild-Text-Relationen und Bild-Bild-Relationen ergeben. Insgesamt macht er vier Typen solcher Verwendungszusammenhänge aus. Zunächst ist dies der direkte Ko(n)text der Bilder, der neben den Bildunterschriften auch durch die Textpassagen der jeweiligen Artikelseiten konstituiert wird. Auch die zu Bildern hinterlegten Metainformationen wie z. B. Bildtitel und Verschlagwortung sind als Kontext einzustufen. Des Weiteren benennt Hammwöhner (2013: 290) auch Bild-Bild-Relationen in den sogenannten Bildgalerien der Wikipedia als Kontext. Auch die Prämierung von Bildern stellt er als Wikipedia-spezifischen Verwendungszusammenhang dar, der die multimodale Ausgestaltung der Wikipedia beeinflusst (Hammwöhner 2013: 290).

Um die manipulative Kraft von Bildern geht es in den explorativen Fallstudien von Wessler et al. 2017, in der u. a. Bildmaterial in Wikipedia-Artikeln zum griechischen Politiker Alexis Tsipras und seiner Partei Syriza in mehreren Sprachversionen der Wikipedia untersucht wurde. Der Vergleich zeigt, dass sich die Artikel in der Auswahl der Bilder, in der Formulierung der Bildunterschriften und in den Bezugnahmen auf die Bilder im Artikeltext deutlich unterscheiden und dass diese Unterschiede interessante und bislang vernachlässigte Ansatzpunkte für die Untersuchung von Bias in den verschiedenen Sprachversionen der Wikipedia bieten.

## 2.4 Wikipedia als Ressource zum Sprach- und Kulturvergleich

Die Wikipedia stellt aufgrund ihrer 295 Sprachversionen und der verschiedenen Namensräume eine einzigartige Ressource zum Sprach- und Kulturvergleich dar, die den heterogenen Charakter natürlicher Sprachen eindrucksvoll belegt. Kontrastiv angelegte Sprach- und Diskursanalysen können den Vergleich in der areal-horizontalen Dimension zum Gegenstand haben und dabei international oder intranational konzipiert werden. Studien zum internationalen Vergleich zwischen Sprachversionen zu Nationalsprachen (etwa zum Deutschen, Fran-

zösischen und Englischen) wurden bereits in mehreren Teildisziplinen der Linguistik durchgeführt. Aus der Perspektive der Grammatikforschung analysierte das Projekt *EuroGr@mm* Wikipedia-Korpora zu insgesamt sechs europäischen Sprachen bezüglich grammatischer Variation (Dalmas/Hansen-Fabricius/Schwinn 2016). Herzberg/Storrer (erscheint) zeigen am Beispiel der interaktiven Einheit *okay*, wie Wikipedia-Korpora weiterhin für interaktionsanalytische Zugänge zur Verwendung einzelner sprachlicher Einheiten in verschiedenen Sprachversionen (deutsch-französisch) genutzt werden können. Die diskurslinguistische Analyse zu multimodalen Metaphern in verschiedenen Sprachversionen der Wikipedia wurde bereits in Abschnitt 2.2 erwähnt (Gredel 2018a). Auch der intranationale Vergleich zwischen dialektalen Versionen der Wikipedia ist in der areal-horizontalen Dimension möglich. Eine erste textlinguistische Analyse vergleicht Wikipedia-Einträge der deutschen und der alemannischen Sprachversion, wobei deutlich wird, dass sich die textuellen Makrostrukturen der untersuchten Wikipedia-Einträge in den beiden Versionen maßgeblich unterscheiden (Gredel 2018b).

## 2.5 Wipedaktik

Wie eingangs erwähnt, gehört die Wikipedia zu den weltweit am häufigsten genutzten Plattformen im Internet. Allerdings rufen die meisten Nutzer nur Informationen ab, sind aber selbst nicht aktiv an der Verbesserung und Erweiterung der Artikeltexte beteiligt und wissen auch wenig über die komplexen Prozesse und Strukturen, die hinter dem Wikipedia-Projekt stecken. Vor diesem Hintergrund ist es wichtig, das Wissen über die sozialen und inhaltlichen Prozesse und die Maßnahmen der Qualitätssicherung didaktisch aufzubereiten und in die Bildungsinstitutionen hineinzutragen. Hierzu dienen Initiativen wie das von der Wikimedia Foundation koordinierte Projekt „Wikipedia macht Schule“ [Wikipedia-Schulprojekt], das Workshops an Schulen und Institutionen der Lehrerbildung anbietet sowie frei verfügbare Materialien für den Unterricht entwickelt. Deutschsprachige Einführungen in die Leitlinien und Prinzipien der Arbeit an der Wikipedia sind das unter WikiBooks verfügbare [Wikipedia-Lehrbuch] oder die in Buchform erschienene Einführung von van Dijk (2010). Stöcklin (2010) bietet viele Anregungen, Beispiele und Hintergrundinformationen, um die Fallstricke, aber auch die Chancen der Wikipedia-Nutzung bewusst zu machen und am Beispiel der Wikipedia allgemeine Kompetenzen im Umgang mit Quellen und der Informationsrecherche online und offline zu entwickeln. Storrer (2012) zeigt, wie man an der Wikipedia die in den Bildungsstandards (KMK-BSG 2011) verankerten Kompetenzen im Umgang mit Hypertexten vermitteln kann und wie sich durch den Vergleich von Artikeln und Diskussionen Unterschiede zwi-

schen text- und interaktionsorientierter Schreibhaltung herausarbeiten lassen. Gredel (i. Dr.) knüpft an das Strategiepapier „Bildung in der digitalen Welt“ (KMK 2016) der Kultusministerkonferenz an und zeigt, welches didaktische Potential die Wikipedia als Lehr- und Lerngegenstand in der Lehrerbildung und in der Schule (Deutschunterricht und Fremdsprachendidaktik, vgl. dazu auch Gredel 2018c) bietet. Der Artikel kartographiert das Feld der Wikipedaktik in Schule und Hochschule, beschreibt die zu entwickelnden digitalen Kompetenzen, erörtert relevante Themen- und Lernfelder und gibt einen aktuellen Überblick über vorhandene digitale Werkzeuge und Ressourcen, u. a. die Plattform *WikiBu* ([www.wikibu.ch](http://www.wikibu.ch)), die für schulische Kontexte entwickelt wurde und dabei helfen soll, die unterschiedlichen Qualitäten von Wikipedia-Artikeln zu bewerten. Wie sich das für die Wikipedia entwickelte Wiki-System MediaWiki in verschiedenen didaktischen Szenarien (Wiki-Schreibkonferenzen, Hypertextlabor etc.) im Kontext Schule und Hochschule einsetzen lässt, beschreiben Beißwenger/Storrer (2010) in einem Überblicksartikel an verschiedenen Fallbeispielen.

### 3 Korpuslinguistische Ressourcen: Wikipedia-Korpora am Institut für Deutsche Sprache

Die Wikipedia steht unter freien Lizenzen zur Verfügung und bietet damit der Forschung die Möglichkeit, die Daten in linguistisch aufbereitete Korpora zu überführen. Das Institut für Deutsche Sprache (IDS) bietet seit 2005 über seine Korpusinfrastruktur Zugriff auf linguistisch aufbereitete Korpora u. a. zu Wikipedia-Artikeln, Wikipedia-Diskussionen und Wikipedia-Benutzerdiskussionen zu verschiedenen Jahrgängen und in mehreren Sprachen. Der Bestand wird laufend erweitert und ausgebaut: Seit 2011 erscheint alle zwei Jahre eine neue Konvertierung der deutschsprachigen Wikipedia (vgl. Längen/Kupietz 2017). Die Wikipedia-Korpora sind Bestandteil des deutschen Referenzkorpus (*DeReKo*), das mit 42 Milliarden Wörtern (Release vom 23.02.2018) das größte Archiv von Korpora geschriebener Gegenwartssprache darstellt (vgl. Längen 2017). Aktuell werden die Wikipedia-Korpora über die Korpusrechercheschnittstelle *COSMAS II* in drei Archiven verwaltet, die sich nach der jeweiligen Wikipedia-Sprachversion richten. Die deutschen und englischen Wikipedia-Korpora sind jeweils in einem Archiv organisiert; weitere acht europäische Sprachversionen sind in einem weiteren Archiv abrufbar. Neben dem Zugriff über *COSMAS II* stehen die Wikipedia-Korpora verschiedener Sprachversionen und aus unterschiedlichen Jahren in Samples bzw. vollständig unter der CC-BY-SA-Lizenz auch zum Download zur Verfügung. Darüber hinaus gibt es auch die Möglichkeit, die deutschsprachigen

Wikipedia-Korpora von 2013 und 2015 mit Wortartenannotierungen des TreeTaggers herunterzuladen.

Der Aufbau von linguistischen Wikipedia-Korpora am IDS aus Wikipedia-Dumps erfolgt zunächst durch eine Aufbereitung gemäß der am IDS verwendeten TEI-Customisierung *I5* (vgl. Längen/Sperberg-McQueen 2012, Beißwenger 2018). Bei der Konvertierung wird der Wikitext unter Verwendung des Wikitext-Parsers *Sweble* und XSLT-Stylesheets in *I5* transformiert (vgl. Margaretha/Längen 2014). Der am IDS verwendete Wikipedia-Konverter steht zum Download zur Verfügung. Die neueste Konvertierung von 2017 enthält Korpora aus Wikipedia-Artikeln, Wikipedia-Artikeldiskussionen, Wikipedia-Nutzerdiskussionen und erstmalig auch Wikipedia-Redundanzdiskussionen. Diese Korpora beinhalten insgesamt 1,5 Milliarden Tokens und enthalten diskursstrukturelle Annotationen, u. a. die Segmentierung der Wiki-Texte in Threads und Postings, die Annotation von Signaturen und Zeitstempeln, mit denen Postings in Wikipedia-Diskussionen versehen sind, sowie die Repräsentation der sogenannten Interlanguage-Links, mit denen Artikel unterschiedlicher Sprachversionen verlinkt sind (vgl. Längen/Kupietz/Margaretha (erscheint)).

## 4 Digitale Methoden und Werkzeuge

Digitale Methoden ergänzen die Möglichkeiten, die korpuslinguistische Verfahren zur Analyse der Wikipedia bieten, da sie direkt auf den Daten und Metadaten der Wikipedia-Sprachversionen ansetzen und zusätzliche interessante Optionen der Auswertung bieten. Sie ermöglichen etwa den Zugriff auf die in Abschnitt 2.3 erläuterte multimodale Dimension der Wikipedia. Digitale Methoden (*digital methods*) gehören zum Methodenspektrum der Sozial- und Medienwissenschaften und sind dadurch charakterisiert, dass sie sich weniger an den sprachlichen Formen als an der Logik des Mediums orientieren („follow the medium“, Rogers 2013: 24). Mittlerweile existieren zahlreiche Tools, die frei verfügbar sind und nach den Prinzipien der Digital Methods auf Wikipedia-Daten operieren. Im Folgenden sollen diese mit ihrem Nutzen für Sprach- und Diskursanalysen beschrieben werden.

Kontroversen in Wikipedia können über das Tool *Contropedia* (<http://contropedia.net>) gefunden werden. Das Tool färbt diejenigen Links in einem Wikipedia-Eintrag mit unterschiedlichen Intensitäten ein, zu denen es auf den Diskussionsseiten und in der Versionsgeschichte besonders viele (sprachliche) Aktivitäten gibt.

Das Benutzerskript *WhoColor* (<https://f-squared.org/whovisual/>) unterstützt die akteurs- bzw. handlungsorientierte Perspektive linguistischer Untersuchun-

gen. Es ordnet über farbliche Markierungen Wikipedia-Autoren die Textpassagen eines Artikels zu, die diese verfasst haben. Der prozentuale Anteil an der Erstellung des jeweiligen Artikels wird nach Autoren geordnet am rechten Rand der Seite gelistet.

Ein Tool der *Digital Methods Initiative* Amsterdam, das die multimodale Dimension der Wikipedia fokussiert und bei der Betrachtung von Bildinventaren in den verschiedenen Sprachversionen zu einem Thema hilfreich ist, ist *Wikipedia Cross-lingual Image Analysis* (<https://tools.digitalmethods.net/beta/wikipediaCrosslingualImageAnalysis/>). Das Tool extrahiert ausgehend von einem Wikipedia-Eintrag die Bilder und Grafiken aller zum Lemma verlinkten Sprachversionen.

Das Tool *Manypedia* (<http://www.manypedia.com>) macht ebenfalls von den Interlanguage-Links Gebrauch und bietet die Möglichkeit, die verknüpften Seiten zu einem Thema in zwei Sprachversionen in parallelen Fenstern anzeigen und in eine gewünschte Zielsprache übersetzen zu lassen. Grundlage für die Übersetzung ist *Google Translate*. Das Tool unterstützt somit die Nutzung der Wikipedia für den Sprach- und Kulturvergleich.

## 5 Fazit

Unser Überblick sollte deutlich machen, dass es sich bei der Wikipedistik um ein innovatives und aktives Forschungsfeld handelt, das auch mit Blick auf die Thematisierung kollaborativer Wissensproduktion im Social Web für Bildungsinstitutionen relevant ist, u. a. um im Rahmen der „Wikipedaktik“ Bewusstheit für die dahinter stehenden Prozesse und Kompetenzen zur Bewertung der dabei entstehenden Produkte zu entwickeln. Anders als viele kommerziell betriebene Social-Media-Plattformen stellt sich die Wikipedia explizit in den Kontext der Bewegung des freien Wissens und offener Ressourcen. Sie wird von ehrenamtlich arbeitenden Nutzern erstellt und unter Lizenzen verfügbar gemacht, die es erlauben, daraus linguistisch aufbereitete Korpora, wie die in Abschnitt 3 beschriebene Wikipedia-Korpusfamilie am Institut für Deutsche Sprache IDS, aufzubauen. Weiterhin werden in der interdisziplinär betriebenen Wikipedistik digitale Werkzeuge und Methoden entwickelt, die auch für interaktions- und diskurslinguistische Untersuchungen interessant sind; Beispiele hierfür haben wir in Abschnitt 4 beschrieben. Insgesamt verfügt die linguistische Wikipedistik damit über sehr umfangreiche digitale Ressourcen, die – wie wir beispielhaft in Abschnitt 2 gezeigt haben – für Untersuchungen von Sprache, Interaktion und Diskursen auf vielfältige Weise genutzt werden können.

## Literatur

- Arendt, Birte/Dreesen, Philipp (2015): Kontrastive Diskurslinguistik – Werkstattbericht zur Analyse von deutschen und polnischen Wikipedia-Artikeln. Standortbestimmung – Eine reflektierende Vorbemerkung. In: Kämper, Heidrun/Warnke, Ingo H. (Hrsg.): Diskurs interdisziplinär. Zugänge, Gegenstände, Perspektiven. Berlin/Boston: de Gruyter, 427–445.
- Beißwenger, Michael (2016): Praktiken in der internetbasierten Kommunikation. In: Deppermann, Arnulf/Feilke, Helmuth/Linke, Angelika (Hrsg.): Sprachliche und kommunikative Praktiken (= Jahrbuch 2015 des Instituts für Deutsche Sprache). Berlin/Boston: de Gruyter, 279–310.
- Beißwenger, Michael (2018): Internetbasierte Kommunikation und Korpuslinguistik: Repräsentation basaler Interaktionsformate in TEL. In: Lobin, Henning/Schneider, Roman/Witt, Andreas (Hrsg.): Digitale Infrastrukturen für die germanistische Forschung (= Germanistische Sprachwissenschaft um 2020, 6). Berlin/New York: de Gruyter, 307–349.
- Beißwenger, Michael/Storrer, Angelika (2010): Kollaborative Hypertextproduktion mit Wiki-Technologie. Beispiele und Erfahrungen im Bereich Schule und Hochschule. In: Jakobs, Eva-Maria/Lehnen, Katrin/Schindler, Kirsten (Hrsg.): Schreiben und Medien. Schule, Hochschule, Beruf (= Textproduktion und Medium, 10). Frankfurt: Peter Lang, 13–36.
- Beyersdorff, Marius (2011): Wer definiert Wissen? Wissensaushandlungsprozesse bei kontrovers diskutierten Themen in „Wikipedia – Die freie Enzyklopädie“ – Eine Diskursanalyse am Beispiel der Homöopathie. Münster: Lit-Verlag.
- Dalmas, Martine/Fabricius-Hansen, Catherine/Schwinn, Horst (2016): Einleitung. In: Dalmas, Martine/Fabricius-Hansen, Catherine/Schwinn, Horst (Hrsg.): Variation im europäischen Kontrast. Untersuchungen zum Satzanfang im Deutschen, Französischen, Norwegischen, Polnischen und Ungarischen. Berlin: de Gruyter.
- Diewald, Nils/Hanl, Michael/Margaretha, Eliza/Bingel, Joachim/Kupietz, Marc/Bański, Piotr/Witt, Andreas (2016): KorAP Architecture – Diving in the Deep Sea of Corpus Data. In: Calzolari, Nicoletta/Choukri, Khalid/Declerck, Thierry/Goggi, Sara/Grobelnik, Marko/Maegaard, Bente/Mariani, Joseph/Mazo, Hélène/Moreno, Asunción/Odijk, Jan/Piperidis, Stelios (Hrsg.): Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia. Paris: European Language Resources Association (ELRA), 3586–3591.
- Ferschke, Oliver/Daxenberger, Johannes/Gurevych, Iryna (2013): A survey of NLP methods and resources for analyzing the collaborative writing process in Wikipedia. In: Gurevych, Iryna/Kim, Jungi (Hrsg.): The People’s Web meets NLP. Collaboratively constructed language resources. Berlin/Heidelberg: Springer, 285–310.
- Gredel, Eva (2016): Digitale Diskursanalysen: Kollaborative Konstruktion von Wissensbeständen am Beispiel der Wikipedia. In: Jaki, Sylvia/Sabban, Anette (Hrsg.): Wissensformate in den Medien (= Kultur – Kommunikation – Kontakte, 25). Berlin: Frank & Timme, 317–339.
- Gredel, Eva (2017): Digital discourse analysis and Wikipedia: Bridging the gap between Foucauldian discourse analysis and digital conversation analysis. In: Journal of Pragmatics (115), 99–114.
- Gredel, Eva (2018a): Metaphorische Muster zur Europäischen Integration in verschiedenen Sprachversionen der Wikipedia – Die Online-Enzyklopädie als Ressource zum Sprach- und

- Kulturvergleich. In: Kämper, Heidrun/Gredel, Eva/Mell, Ruth M./Polajnar, Janja (Hrsg.): Diskurs – kontrastiv. Diskurslinguistik als Methode zur Erfassung sprachübergreifender und transnationaler Diskursrealitäten (= Sprache – Politik – Gesellschaft 23). Bremen: Hempen, 97–121.
- Gredel, Eva (2018b): Di Alemannisch Wikipedia – Di frei Enzyklopedi, wo alli chöi mitschaffe: Eine text- und variationslinguistische Analyse der alemannischen Sprachversion der Wikipedia. In: Adamzik, Kirsten/Maselko, Mateusz (Hrsg.): Variationslinguistik trifft Textlinguistik. Tübingen: Narr, 161–182.
- Gredel, Eva (2018c): Vom Bannspruch zur Wikipedaktik? – Wikipedia als Ressource zum Sprach- und Kulturvergleich: Potenziale kontrastiver Diskursanalysen der Online-Enzyklopädie in der Fremdsprachendidaktik. In: Zeitschrift für Romanische Sprachen und ihre Didaktik 12(1), 99–117.
- Gredel, Eva (i. Dr.): Wikipedaktik: Kollaborative Sachlexikografie als Lehr- und Lerngegenstand im Deutschunterricht. In: Lexicographica. 2(2018). Erscheint im Frühjahr 2019.
- Hammwöhner, Rainer (2007): Qualitätsaspekte der Wikipedia. In: kommunikation@gesellschaft 8 (2007). URL: [http://www.soz.uni-frankfurt.de/K.G/B3\\_2007\\_Hammwoehner.pdf](http://www.soz.uni-frankfurt.de/K.G/B3_2007_Hammwoehner.pdf) (Stand: 28. 08. 2018).
- Hammwöhner, Rainer (2013): Bilddiskurse in den Wikimedia Commons. In: Frank-Job, Barbara/Mehler, Alexander/Sutter, Tilmann (Hrsg.): Die Dynamik sozialer und sprachlicher Netzwerke. Konzepte, Methoden und empirische Untersuchungen am Beispiel des WWW. Wiesbaden: Springer, 285–311.
- Herzberg, Laura/Storrer, Angelika (im Erscheinen): Investigating OKAY across Genres, Modes and Languages: A Corpus-based Study on German and French. Cahiers du Laboratoire de Recherche sur le Langage.
- Kallass, Kerstin (2013): Schreibprozesse in der Wikipedia. Eine linguistische Analyse. Dissertation Universität Koblenz-Landau. URL: <https://kola.opus.hbz-nrw.de/frontdoor/index/index/docId/726> (Stand: 12. 03. 2018).
- Kallass, Kerstin (2015): Schreiben in der Wikipedia. Prozesse und Produkte gemeinschaftlicher Textgenese. Wiesbaden: Springer VS.
- Kupietz, Marc/Belica, Cyril/Lüngen, Harald/Perkuhn, Rainer (2014): Zwischen Empirie und Ästhetik – Ansätze zur korpuslinguistischen Untersuchung und Bewertung von Sprachwandel. In: Plewnia, Albrecht/Witt, Andreas: Sprachverfall? Dynamik – Wandel – Variation. Berlin/New York: de Gruyter, 149–169.
- Liebert, Wolf-Andreas/Kohl, Christian (2004): Selbstorganisation der Wissenschaftsvermittlung: Quellentransparenz, Kontroversität und Qualitätssicherung in der Internet-Enzyklopädie Wikipedia. In: Fachsprache (3–4), 134–148.
- Lüngen, Harald (2017): DeReKo – Das Deutsche Referenzkorpus. Schriftkorpora der deutschen Gegenwartssprache am Institut für Deutsche Sprache in Mannheim. *Zeitschrift für germanistische Linguistik* 45 (1), 161–170.
- Lüngen, Harald/Kupietz, Marc (2017): CMC Corpora in DeReKo. In: Bański, Piotr/Kupietz, Marc/Lüngen, Harald/Rayson, Paul/Biber, Hanno/Breiteneder, Evelyn/Clematide, Simon/Mariani, John/Stevenson, Mark/Sick, Theresa (Hrsg.): Proceedings of the Workshop on Challenges in the Management of Large Corpora and Big Data and Natural Language Processing (CMCLC-5+BigNLP) 2017 including the papers from the Web-as-Corpus (WAC-XI) guest section. Birmingham, 24 July 2017. Mannheim: Institut für Deutsche Sprache, 20–24. URL: [https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/6259/file/Luengen\\_Kupietz\\_CMC+Corpora\\_2017.pdf](https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/6259/file/Luengen_Kupietz_CMC+Corpora_2017.pdf) (Stand: 15. 07. 2018).

- Lüngen, Harald/Sperberg-McQueen, Christopher M. (2012): A TEI P5 Document Grammar for the IDS Text Model. In: Bański, Piotr/Modignani Picozzi, Eleonora Litta/Witt, Andreas (Hrsg.): TEI and Linguistics. *Journal of the Text Encoding Initiative* 3.2012.
- Lüngen, Harald/Kupietz, Marc/Margaretha, Eliza (im Erscheinen): Neue Features der linguistischen Wikipedia-Korpora am IDS Mannheim. In: *Diskurse – digital* 2018/1 (= Sonderband 1: Linguistische Wikipedistik). URL: <https://journals.bib.uni-mannheim.de/diskurse-digital> (Stand: 28. 08. 2018).
- Margaretha, Eliza/Lüngen, Harald (2014): Building linguistic corpora from Wikipedia articles and discussions. In: *Journal of Language Technologie and Computational Linguistics JLC* 29/2, 59–83.
- Marx, Konstanze/Weidacher, Georg (2014): *Internetlinguistik. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr.
- Pentzold, Christian (2007): *Wikipedia. Diskussionsraum und Informationsspeicher im neuen Netz* (= Internet Research 29). München: Fischer.
- Pscheida, Daniela (2010): *Das Wikipedia-Universum: Wie das Internet unsere Wissenskultur verändert*. Bielefeld: transcript.
- Rogers, Richard (2013): *Digital Methods*. Cambridge/London: The MIT Press.
- Stegbauer, Christian (2009): *Wikipedia. Das Rätsel der Kooperation*. Wiesbaden: Springer.
- Stöcklin, Nando (2010): *Wikipedia clever nutzen – in Schule und Beruf*. Zürich: Orell Füssli.
- Storrer, Angelika (2012): *Neue Text- und Schreibformen im Internet: Das Beispiel Wikipedia*. In: Feilke, Helmuth/Köster, Juliane/Steinmetz, Michael (Hrsg.): *Textkompetenzen in der Sekundarstufe II*. Freiburg: Filibach bei Klett, 277–304.
- Storrer, Angelika (2017): *Grammatische Variation in Gespräch, Text und internetbasierter Kommunikation*. In: Konopka, Marek/Wöllstein, Angelika (Hrsg.): *Grammatische Variation. Empirische Zugänge und theoretische Modellierung*. Berlin/New York: de Gruyter, 105–125.
- Storrer, Angelika (2018a): *Web 2.0: das Beispiel Wikipedia*. In: Janich, Nina/Birkner, Karin (Hrsg.): *Handbuch Text und Gespräch* (= Handbücher Sprachwissen, 5). Berlin/Boston: de Gruyter, 397–417.
- Storrer, Angelika (2018b): *Interaktionsorientiertes Schreiben im Internet*. In: Deppermann, Arnulf/Reineke, Silke (Hrsg.): *Sprache im kommunikativen, interaktiven und kulturellen Kontext*. Berlin/Boston: de Gruyter, 219–244.
- Storrer, Angelika (i. Dr.): *Text und Interaktion im Internet*. In: Eichinger, Ludwig/Plewnia, Albrecht (Hrsg.): *Neues vom heutigen Deutsch: Empirisch -- methodisch -- theoretisch*. Berlin/Boston: de Gruyter.
- van Dijk, Ziko (2009): *Handbuch der vielsprachigen Wikipedia*. URL: [https://de.wikipedia.org/w/index.php?title=Spezial:Buch&bookcmd=download&collection\\_id=fd35e cdc371c50c2407f05a340caca82767f9c0c&writer=rl&return\\_to=Wikipedia%3ABücher%2FHandbuch+der+Vielsprachigen+Wikipedia](https://de.wikipedia.org/w/index.php?title=Spezial:Buch&bookcmd=download&collection_id=fd35e cdc371c50c2407f05a340caca82767f9c0c&writer=rl&return_to=Wikipedia%3ABücher%2FHandbuch+der+Vielsprachigen+Wikipedia) (Stand: 28. 08. 2018).
- van Dijk, Ziko (2010): *Wikipedia. Wie Sie zur freien Enzyklopädie beitragen*. München.
- Wessler, Hartmut/Theil, Christoph Kilian/Stuckenschmidt, Heiner/Storrer, Angelika/Debus, Marc (2017): *Wikiganda: Detecting Bias in Multimodal Wikipedia Entries*. In: Seizov, Ognyan/Wildfeuer, Janina (Hrsg.): *New Studies in Multimodality*. London/New York: Bloomsbury, 201–224.

## Quellenverzeichnis

- Alexa (2018): The top 500 sites on the web. URL: <https://www.alexa.com/topsites> (Stand: 30. 07. 2018).
- KMK-BSG 2011 = Kultusministerkonferenz (2011): Bildungsstandards für Deutsch Gymnasium – Klassen 6, 8, 10, Kursstufe. URL: [http://www.bildung-staerkt-menschen.de/unters-tuetzung/schularten/Gym/faecher/D/listing\\_bildungsstandards](http://www.bildung-staerkt-menschen.de/unters-tuetzung/schularten/Gym/faecher/D/listing_bildungsstandards) (Stand: 12. 12. 2011).
- KMK 2016 = Kultusministerkonferenz (2016): Bildung in der digitalen Welt. Berlin. URL: [https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2016/Bildung\\_digitale\\_Welt\\_Webversion.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2016/Bildung_digitale_Welt_Webversion.pdf) (Stand: 06. 04. 2018).
- Wikimedia (2018): Wikipedia Statistics. URL: [https://stats.wikimedia.org/EN/TablesWikipedia\\_ZZ.htm](https://stats.wikimedia.org/EN/TablesWikipedia_ZZ.htm) (Stand: 30. 07. 2018).
- Wikimedia Commons (2018): Main Page. URL: [https://commons.wikimedia.org/w/index.php?title=Main\\_Page&oldid=205347442](https://commons.wikimedia.org/w/index.php?title=Main_Page&oldid=205347442) (Stand: 01. 09. 2016).
- Wikipedia (2018): Wikipedia:Wikipedistik. URL: <https://de.wikipedia.org/wiki/Wikipedia:Wikipedistik> (Stand: 30. 07. 2018).
- Wikipedia-Lehrbuch (2016): URL: <http://de.wikibooks.org/wiki/Wikipedia-Lehrbuch> (Stand: 28. 08. 2018).
- Wikipedia-Schulprojekt (2012): Wikipedia macht Schule. URL: Informationsseite zum Wikipedia-Schulprojekt: <https://wikimedia.de/wiki/Schulprojekt> (Stand: 28. 08. 2018).
- WP-Wikipedistik = Wikipedia:Wikipedistik (2015). URL: <http://de.wikipedia.org/wiki/Wikipedia:Wikipedistik> (Stand: 28. 08. 2018).