

Untersuchung des Gender Bias im deutschsprachigen Wikipedia

Sarah Krause¹

Abstract: Stereotypen und Vorurteile bezüglich Frauen und Männer herrschen trotz gesetzlicher Gleichstellung in den Köpfen vieler Menschen. Dieser sogenannte Gender Bias hat Einfluss auf viele Bereiche des gesellschaftlichen Lebens, auch auf vermeintlich neutrale Medien wie die Online-Enzyklopädie Wikipedia. Der Gender Bias in Wikipedia wurde in verschiedenen Abhandlungen anhand von Biographien untersucht und insbesondere hinsichtlich der Sichtbarkeit und Sprache durch verschiedene Parteien festgestellt. Diese Arbeit baut auf vorangegangenen Untersuchungen auf und betrachtet die Ausprägung des Gender Bias im deutschsprachigen Wikipedia hinsichtlich drei verschiedener Aspekte: Textanzahl, Textlänge und Sprache. Insgesamt bestätigten die Untersuchungen einen Gender Bias im deutschsprachigen Wikipedia. Nur 29,16% der betrachteten Biographien handeln von Frauen, auch die Textlänge der Artikel über Frauen ist durchschnittlich um 12,3% kürzer. Hinsichtlich der Sprache konnten weniger Unterscheidungen in der Beschreibungen der Geschlechter festgestellt werden als in vorherigen Abhandlungen. Dies lässt auf einen positiven Trend hin zur Minimierung des Gender Bias hoffen.

Keywords: Gender Bias, Wikipedia

1 Einleitung

Heute sind Frauen und Männer in Deutschland gesetzlich gleichgestellt. Diese Gleichstellung ist jedoch noch nicht in allen Bereichen umgesetzt. Beispielsweise verdienen Männer immer noch mehr als Kolleginnen in vergleichbaren Positionen [Fü21]. Die früheren Rollenbilder von Mann und Frau sind noch nicht aus allen Bereichen der Gesellschaft und allen Köpfen ausgeradiert. Geschlechterstereotypen und Vorurteile führen zu einer unterschiedlichen Betrachtung der Geschlechter, dem sogenannten Gender Bias. Gender Bias hat, oft unbemerkt, Einfluss auf verschiedene Alltagssituationen. Auch in einigen vermeintlich neutralen Quellen kann Gender Bias nachgewiesen werden, ein Beispiel ist die Online-Enzyklopädie Wikipedia.

Der Gender Bias in Wikipedia wurde bereits durch verschiedene Abhandlungen betrachtet, diese Arbeit soll darauf aufbauen. Ziel ist es zu untersuchen, inwiefern der Gender Bias auch im deutschsprachigen Wikipedia vorhanden ist.

¹ Hochschule für Technik und Wirtschaft (HTW) Berlin, Treskowallee 8, 10318 Berlin,
Sarah.Krause@student.htw-berlin.de

Zur Beantwortung der Forschungsfrage wurden im Rahmen dieser Arbeit Biographien in Form von Wikipediaartikeln hinsichtlich verschiedener Aspekte untersucht.

In den folgenden Kapiteln wird zunächst eine thematische Einführung bezüglich des Gender Bias auf Wikipedia, dessen Ausprägungen und Gründe gegeben, bevor auf die praktischen Untersuchungen eingegangen wird. In Kapitel 3 folgt die Erläuterung der Methode. Es wird beschrieben, welche Daten wie ausgewertet werden sollen und wie die Daten gesammelt wurden. Die Ergebnisse der Untersuchungen werden in Kapitel 4 beschrieben. Abschließend wird ein Fazit gezogen und ein Ausblick auf Anknüpfungspunkte der Arbeit gegeben.

2 Hintergrund

Als Online-Enzyklopädie sammelt Wikipedia Artikel zu verschiedenen Themenbereichen und veröffentlicht sie kostenlos, mit einer hohen Reichweite, im Internet [CB12, Gr15]. Es enthält Beiträge über verschiedene Epochen, Disziplinen und Personen und möchte, nach eigener Aussage, das Wissen der Welt darstellen. Die Besonderheit von Wikipedia, im Vergleich zu herkömmlichen Enzyklopädien, ist, dass die Erstellung der Artikel Community basiert ist [Wa16].

Benutzer*innen können sich freiwillig als Autor*innen beteiligen und müssen keine Experten in dem Gebiet ihres Beitrags sein [Gr15]. Ihre Beiträge müssen lediglich durch Quellen gestützt sein.

Trotz dieser Art der Informationssammlung zeigen Studien, dass Wikipediaartikel eine vergleichbare Qualität mit herkömmlichen Enzyklopädien aufweisen. Gegen einen Grundsatz von Enzyklopädien verstößt Wikipedia jedoch: Das Verfassen aus einem neutralen Standpunkt (englisch: neutral point of view, NPOV) heraus. Auf der Plattform kommt es verstärkt zur Überrepräsentation von bestimmten Meinungen bzw. Sichtweisen (Englisch: point of view, POV), dem sogenannten Bias. Bias tritt dabei in verschiedenen Ausprägungen auf, in kultureller, politischer und geschlechterspezifischer Sicht [Hu17]. Diese Arbeit fokussiert sich auf die unterschiedliche Betrachtungsweise der Geschlechter, dem Gender Bias, in Wikipedia. Gender Bias beschreibt die asymmetrische Assoziation von Ereignissen mit Frauen und Männern, die zu Geschlechterstereotypen führen können [SP11].

Der Gender Bias in Wikipedia wurde bereits in verschiedenen wissenschaftlichen Abhandlungen betrachtet [CB12, Gr15, Hu17, RR11, SP11, Wa15, Wa16]. Dabei wurden verschiedene Ausprägungen des Bias in Wikipedia untersucht und festgestellt. Als Grundlage der Untersuchungen dienen Wikipediaartikel in Form von Biographien. Da in diesen einzelne Personen thematisiert werden, eignen sie sich, um Unterschiede in der Behandlung und Betrachtung der Geschlechter festzustellen [Gr15]. Unterschiede zwischen den Biographien der Geschlechter wurden insbesondere in der Sprache beziehungsweise der Inhalte der Texte und der Sichtbarkeit der Artikel festgestellt [CB12, Gr15, Hu17, RR11, SP11, Wa15, Wa16].

Die Sprache, die Autor*innen innerhalb eines Textes verwendet, kann, trotz des in Wikipedia üblichen neutralen Schreibstils, ein Indikator für die Meinung der Autor*innen über das Thema des Artikels sein. Dabei kann zwischen explizitem und implizitem Bias unterschieden werden. Explizit beschreibt, dass Autor*innen ihre Meinung direkt in den Text einfließen lassen und ist dadurch oft erkennbar. Bei einem impliziten Bias hingegen werden bestimmte Fakten oder Aspekte eines Themas weggelassen, die eigentlich zum Sachverhalt gehören. Damit wird, durch die Leser*innen des Textes unerkannt, der Fokus nur auf die beschriebenen Inhalte gelegt [Hu17]. Gerade in einer Enzyklopädie wie Wikipedia ist ein impliziter Bias besonders gefährlich, da Leser*innen auf die Korrektheit und Vollständigkeit der Inhalte vertrauen und auf deren Basis Meinungen bilden und sich Wissen aneignen [SP11].

In mehreren Untersuchungen zur Ausprägung des Gender Bias in der Sprache bzw. in den Themen der Wikipediaartikel-Biographien wurden Unterschiede zwischen Männern und Frauen festgestellt [Gr15, SP11, Wa15, Wa16]. In Artikeln über Männern werden häufig Themen wie Sport und ihr Beruf in den Fokus gestellt und verstärkt kognitive Begriffe genutzt [Gr15, Wa15]. In Biographien über Frauen wird hingegen verstärkt auf ihren Familienstand (insbesondere verheiratet oder geschieden) und ihre Kinder eingegangen [Wa15]. Bei Artikeln über Frauen werden häufig persönliche Ereignisse mit beruflichen vermischt, was bei Biographien über Männer nicht der Fall ist. Diese Vermischung stärkt Geschlechterstereotypen in Bezug auf Karriere, Ehe und Elternschaft und gefährdet die Gleichstellung der Geschlechter in der Gesellschaft [SP11]. In Biographien über Frauen gibt es außerdem häufiger geschlechtsbezogene Wörter und Inhalte als in denen über Männer [Gr15, Wa15, Wa16]. Frauenartikel enthalten eher Wörter wie Frau, weiblich oder Dame (Original in Englisch: woman, female, lady) als Wörter wie Mann, männlich oder Herr (Original in Englisch: man, masculine, gentleman) in Artikeln über Männern vorkommen [Wa15]. Wagner et al. (2015) bringen dies mit der Theorie in Zusammenhang, dass das männliche Geschlecht in vielen Situationen als Null-Gender (Standardgeschlecht) angesehen wird. Männliche Wörter scheinen in diesen Situation als nicht notwendig, da der Kontext das männliche Geschlecht vorzugeben scheint [Wa15].

Diese unterschiedliche Beschreibung der Geschlechter ist wahrscheinlich nicht bei Betrachtung einer einzelnen Biographie erkennbar, doch es zeigt in der Gesamtheit einen Spiegel der Meinung der Autor*innen.

Artikel über Männer und Frauen in Wikipedia unterscheiden sich neben der Sprache, die in ihren Biographien verwendet wird, auch in ihrer Sichtbarkeit [CB12, Gr15, RR11, Wa15, Wa16]. Um Sichtbarkeit zu erreichen ist einerseits entscheidend, dass Artikel zum jeweiligen Geschlecht existieren und vorhandene Artikel eine gute Netzwerkstruktur innerhalb Wikipedias haben.

Graells-Garrido et al. stellten in ihren Untersuchungen fest, dass nur 15,5% der Biographien in ihren Daten von Frauen handeln [Gr15]. Auch die Betrachtung von Regale und Rhue bestätigen, dass in Wikipedia und anderen Online Datenbanken weniger Einträge über Frauen als über Männer existieren. [RR11]. Doch woran liegt das? Es ist wahrschein-

licher, dass eine Biographie über eine bedeutende Frau fehlt, als über einen bedeutenden Mann [CB12]. Wagner et al. (2016) interpretieren dies als Ergebnis des Effekts der gläsernen Decke. Der Begriff der gläsernen Decke wird vorrangig im beruflichen Kontext genutzt. Sie beschreibt eine unsichtbare Decke, die es Frauen unmöglich, beziehungsweise schwerer macht, in höheren Positionen eingestellt zu werden. Wagner et al. (2016) vermuten, dass Wikipedia diese gläserne Decke in Bezug auf die Erstellung von Biographien besitzt. So werden Biographien nur für Frauen erstellt, die höhere Fähigkeiten und Fertigkeiten haben als Männer über die eine Biographie verfasst wird [Wa16].

Ein Merkmal von Wikipedia ist die Verlinkung zwischen den Artikeln. Werden in Artikeln Personen oder Themen genannt, die ebenfalls einen Wikipediaartikel besitzen, werden sie verlinkt, sodass der Leser*innen sie direkt aufrufen kann. Umso öfter ein Artikel in anderen verlinkt ist, desto wahrscheinlicher ist es, dass ein Wikipedianutzer auf ihn aufmerksam wird. Eine häufige Verlinkung ist deshalb ein Merkmal für die Relevanz eines Artikels. Untersuchungen zeigen, dass Biographien von Männern eine größere Netzwerkstruktur haben und deswegen häufiger zu finden sind [Gr15]. Frauenartikel werden hingegen weniger häufig verlinkt [Wa16]. Wagner et al. (2015) fanden außerdem heraus, dass Frauenartikel häufiger zu Biographien über Männer verlinken als umgedreht. Dies hat einen negativen Effekt auf die Sichtbarkeit von Frauen in Wikipedia [Wa15].

Es gibt verschiedene Erklärungen, warum es in Wikipedia einen Gender Bias hinsichtlich der beschriebenen Aspekte gibt. Ein Ansatz ist die unterschiedliche Sicht auf die Geschlechter in der Vergangenheit [Wa15]. Frauen sind in der Geschichte oft unterrepräsentiert, da sie einer bestimmten Rolle (oft der Hausfrau und Mutter) zugeschrieben wurden und oft ausschließlich hinsichtlich dieser betrachtet wurden. Biographien zu Personen aus dieser Zeit sind deshalb oft auf diese Thematik beschränkt [Wa15].

Der aber wohl entscheidendere Ansatz ist die Übertragung des Bias der Autor*innen auf ihre Wikipediabeiträge [CB12, Gr15, Wa15, Wa16]. In der Autorenschaft von Wikipedia herrscht ein Gender Gap (Geschlechterunterschied), nur 16,1% der Autor*innen sind weiblich [HS13]. Es gibt aber nicht nur weniger Frauen, die zu Wikipediaartikeln beitragen, Autorinnen schreiben durchschnittlich weniger und hören häufig früher auf, Beiträge auf Wikipedia zu verfassen und zu bearbeiten. Da Autorinnen sich auf andere Themen fokussieren als männliche Autoren, ist der Gender Gap dahingehend ein Problem, dass diese Themen und Sichtweisen vernachlässigt werden [Gr15]. Warum der Anteil der Autorinnen in Wikipedia so gering ist, begründen Collier und Bear anhand von zwei Faktoren: Konflikten und Wissen. Autoren in Wikipedia erstellen nicht nur neue Artikel, die Hauptarbeit besteht in der Erweiterung oder Bearbeitung bereits bestehender Beiträge. Das Löschen und Aktualisieren von Texten anderer Autor*innen führt zu Konflikten auf der Plattform. Frauen neigen wiederum eher dazu als Männer, Konflikte zu vermeiden [CB12]. Der Faktor Wissen bezieht sich nicht auf tatsächliche Unterschiede in Fähigkeiten und Kenntnissen von Männern und Frauen. Männer haben durchschnittlich ein höheres Selbstvertrauen bezüglich ihres Fachwissens, was in einer stärkeren Motivation mündet, ihr Wissen mit der Welt teilen zu wollen [CB12].

3 Methode

Dieses Kapitel beschreibt die Ziele der Arbeit sowie die zu erwartenden Ergebnisse. Es wird außerdem der Erhalt der Daten und die Vorgehensweise bei der Auswertung erläutert.

3.1 Untersuchung

Im Großteil der Abhandlungen, die sich mit dem Gender Bias in Wikipedia auseinandersetzen, wird das englischsprachige Wikipedia als Datenquelle genutzt, da es die größte Anzahl von Biographien enthält. Mit ca. 102.000 Biographien ist das deutschsprachige Wikipedia auf dem dritten Platz. Bezüglich des Anteils der Biographien über Frauen bildet es das Schlusslicht unter den 20 größten Wikipedia Editionen mit 13,2 % [Wa16].

Diese Arbeit untersucht das bisher weniger betrachtete deutschsprachige Wikipedia. Basis der Untersuchungen bilden Biographien. Dabei werden nur Biographien von Personen betrachtet, die in den Jahren 1996 – 2000 geboren sind. Umso später die Geburtsjahre, desto weniger Einfluss sollte die unterschiedliche historische Betrachtung der Geschlechter auf die Inhalte der Biographien haben. Würden noch aktuellere Geburtsjahre gewählt werden, so würde die Stichprobe zu klein werden.

Anhand der ausgewählten Biographien soll überprüft werden, inwiefern in den modernen Biographien des deutschsprachigen Wikipedias ein Gender Bias zu erkennen ist. Dazu werden die Biographien hinsichtlich drei Aspekten untersucht: Textanzahl, Textlänge und Sprache. Die Auswahl der untersuchten Gesichtspunkte erfolgte basierend auf den Methoden bisheriger Untersuchungen.

Die Textanzahl wird genutzt, um einen Teil der Sichtbarkeit zu überprüfen. Es wird gezählt, wie viele Wikipediaartikel Biographien über Frauen und Männer enthalten. Aufgrund des beschriebenen Effekts der gläsernen Decke wird erwartet, dass es weniger Biographien über Frauen als über Männer gibt.

Neben der Textanzahl wird die Textlänge der vorhandenen Beiträge gesammelt und verglichen. Die Textlänge gibt Auskunft darüber, wie viel Beachtung der beschriebenen Person gegeben wird.

Mit Hilfe der Sprache soll untersucht werden, wie Personen beschrieben werden und somit die Meinung der Autor*innen und ein möglicher Bias ermittelt werden. Die Untersuchung der Sprache erfolgt über Worthäufigkeiten, welche ein Indikator für Unterschiede in Texten sind [Gr15]. Für die Auswertung von Worthäufigkeiten wird das Programm Linguistic Inquiry Word Count (LIWC) verwendet. LIWC ist ein Textanalyseprogramm, welches automatisiert Wörter in mehr als 80 psychologische Kategorien einordnet [TP09]. Die deutsche Version DE-LIWC2015 enthält 18.711 Wörter. Wörter können in mehr als einer Kategorie vorkommen [Me17]. In dieser Arbeit werden insbesondere die Kategorien bezüglich

sozialer, kognitiver und biologischer Prozesse (Original Englisch: social, cognitive and biological processes) betrachtet. Abbildung 1 zeigt Beispielwörter der drei Kategorien sowie ihren Unterkategorien.

Social processes	social	gesellig, kumpel, reden
Family	family	papa, tochter, tante
Friends	friend	bro, kumpel
Female references	female	Frau, mädchen, weiblich
Male references	male	bruder, mann, onkel
Cognitive processes	cogproc	denken, weil, Wissen
Insight	insight	denken, realisieren
Causation	cause	deswegen, grund
Discrepancy	discrep	sollte, wollte
Tentative	tentat	eventuell, vielleicht
Certainty	certain	immer, sicher
Differentiation	differ	aber, sonst
Biological processes	bio	essen, blut, schmerz
Body	body	arm, kopf, muskel
Health	health	erkältet, klinik, medikament
Sexual	sexual	geil, heiß, nackt
Ingestion	ingest	hunger, mahlzeit, pizza

Abb. 1: Betrachtete LIWC Hauptkategorien und deren Unterkategorien [Me17].

Analog zu den Ergebnissen von Graells-Garrido. et al. wird erwartet, dass in Biographien über Frauen häufiger Begriffe der Kategorien soziale und biologische Prozesse auftreten, bei Biographien über Männer hingegen Wörter die Kategorie kognitive Prozesse [Gr15].

3.2 Datenquelle

Als Datenquelle für die Untersuchungen dienen Wikipediaartikel in Form von Biographien von Personen, die in den Jahren 1996 bis 2000 geboren sind. Die Daten aus den Biographien wurden mit Hilfe eines Web Scrapers direkt aus Wikipedia extrahiert. Der Web Scraper wurde in Python programmiert.

Wikipediaartikel sind Kategorien zugeordnet, die am Ende des Artikels aufgeführt werden. Bei Artikeln über Personen enthalten diese Kategorien unter anderem das Geburtsjahr (z.B. Geboren 1998) und das Geschlecht (z.B. Frau). Wird eine der Kategorien ausgewählt, werden alle Wikipediaartikel, die dieser Kategorie zugeordnet sind, angezeigt. Die Liste eines Geburtsjahrs dient als Ausgangspunkt für die Ausführung des Web Scrapers.

Der Web Scraper durchläuft automatisch die Liste der Namen der in dem ausgewählten Jahr geborenen Personen und öffnet die individuellen Wikipediaartikel. Aus dieser Seite

wird der Text extrahiert und geprüft, welches Geschlecht die Person hat. Dafür wird überprüft ob die Kategorie Frau oder Mann zugeordnet ist. Das Geschlecht sowie die Länge des extrahierten Textes werden in einer CSV-Datei gespeichert. Die Länge des Textes wird in Anzahl der Zeichen gemessen. Für die Analyse mit LIWC wird nicht nur die Textlänge sondern der gesamte Text benötigt. Alle Texte einer Geschlechts werden in einer TXT-Datei gespeichert. Dieser Vorgang wird für die fünf ausgewählten Geburtsjahre wiederholt.

Die Datensammlung fand im Januar 2022 statt.

3.3 Vorgehen in der Auswertung

Für die Auswertung der Textanzahl und der Textlänge wurden die CSV-Dateien der einzelnen Jahre in Excel importiert und mit Hilfe von Pivot-Tabellen ausgewertet und visualisiert.

Wie schon beschrieben, wurde die Sprache in Form der Worthäufigkeiten mit Hilfe des DELIWCV2015 ausgewertet. Dazu wurden die TXT-Dateien in das Programm geladen. LIWC gibt die Worthäufigkeiten in Form von Prozent wieder [TP09]. Die Ergebnisse wurden wiederum in Excel exportiert. Dort wurde der Mittelwert zwischen den Ergebnissen der einzelnen Jahre pro Geschlecht gebildet. Anhand dieser Mittelwerte wurden die Ausprägungen der Kategorien verglichen.

Die Ergebnisse der Datenauswertung werden im folgenden Kapitel erläutert.

4 Ergebnisse

Insgesamt wurden 16.311 Biographien für die Geburtsjahre 1996 bis 2000 extrahiert. Die Ergebnisse der Analyse hinsichtlich Textanzahl, Textlänge und Sprache werden in diesem Kapitel beschrieben.

4.1 Textanzahl

Die Anzahl der Biographien pro Jahr nimmt, wie in Abbildung 2 zu erkennen, stetig ab. Waren es für im Jahr 1996 geborene Personen noch 4528 Biographien, waren es nur noch 2295 für Personen die 2000 geboren sind. Dies ist nicht verwunderlich, da die in diesen Jahren geborenen Personen noch relativ jung sind. Mit den Jahren steigt die Anzahl an Personen, deren Biographien in Wikipediaartikeln abgebildet sind. In Zukunft werden die Anzahlen der Biographien für diese Geburtsjahre also weiter steigen.

Der Anteil von Biographien über Frauen hat sich, wie in Abbildung 2 zu erkennen, während der betrachteten Jahre kaum verändert. Insgesamt beträgt der Frauenanteil der Biographien 29,16%. Die Hypothese, dass es weniger Biographien über Frauen gibt, ist somit bestätigt.

Geschlecht	1996	1997	1998	1999	2000	Gesamt
Keine Angabe	0.16%	0.29%	0.18%	0.07%	0.09%	0.17%
Frau	29.87%	28.67%	28.80%	28.05%	30.46%	29.16%
Mann	69.96%	71.05%	71.02%	71.87%	69.46%	70.67%

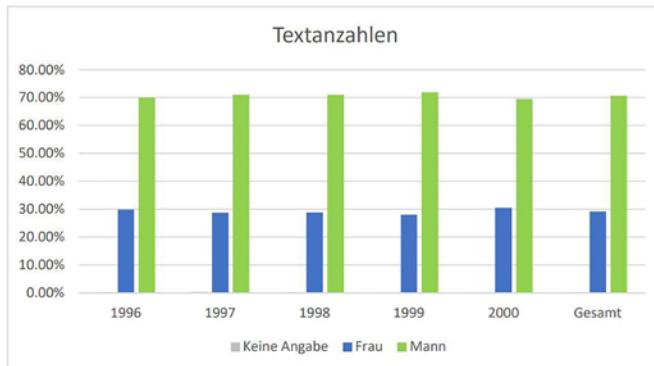


Abb. 2: Übersicht der Textanzahlen (absolut und prozentual) und Visualisierung

Damit liegt der Anteil an Frauen deutlich über dem für das gesamte deutschsprachige Wikipedia gemessenen Anteil von 13,2% [Wa16]. Dies könnte daran liegen, dass ausschließlich jüngere Jahrgänge in die Betrachtung eingeflossen sind. Das frühere Rollenbild, was es vielen Frauen erschwert hat in der Geschichte sichtbar zu werden, ist für diese Geburtsjahre nicht mehr vorhanden.

Trotzdem sind die erhaltenen 29,16% noch weit von 50% entfernt, was dem ungefähren realen Frauenanteil entspricht. Dies kann darauf hindeuten dass die gläserne Decke, die die Erstellung von Biographien über Frauen verhindert, weiterhin besteht.

4.2 Textlänge

Auch die Hypothese hinsichtlich der Textlänge hat sich bestätigt. Die Biographien über Frauen sind in jedem Jahr durchschnittlich kürzer als die über Männer.

Biographien über Frauen sind im Durchschnitt 1.711 Zeichen lang, Biographien über Männer hingegen durchschnittlich 1.921 Zeichen. Das ist ein Unterschied von 12,3%. Die genauen Ergebnisse sind in Abbildung 3 abgebildet.

Geschlecht	1996	1997	1998	1999	2000	Gesamt
Frau	1799	1739	1625	1590	1752	1711
Mann	1968	1996	1889	1843	1847	1921
Differenz	169	257	264	253	94	210
Differenz in %	9.4%	14.8%	16.3%	15.9%	5.4%	12.3%

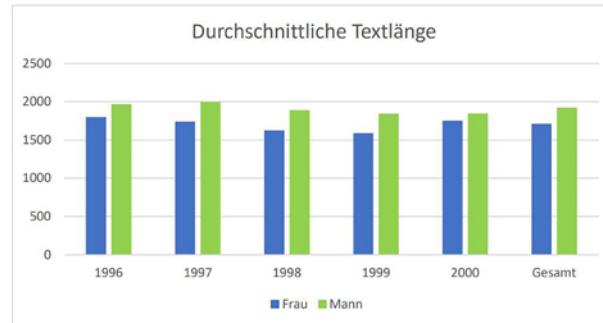


Abb. 3: Übersicht der durchschnittlichen Textlängen und Visualisierung

Einen großen Unterschied gibt es in der maximalen Textlänge. Der längste Artikel über einen Mann ist mit 39.180 Zeichen um 10.524 Zeichen länger als der längster Artikel über eine Frau, was einen Unterschied von 36,7% ergibt. Die kürzesten Artikel umfassen sowohl bei Frauen als auch bei Männern 65 Zeichen. Maximale und minimale Textlänge sind in Abbildung 4 abgebildet.

	Durchschn. Textlänge	Minimum Textlänge	Maximum Textlänge
Frau	1711	65	28656
Mann	1921	65	39180

Abb. 4: Übersicht der Textlängen

4.3 Sprache

Die Ergebnisse der Sprachanalyse mittels Worthäufigkeiten stimmen nicht für alle Kategorien mit den Erwartungen überein. Abbildung 5 zeigt das Ergebnis der Analyse für die drei untersuchten Hauptkategorien soziale, kognitive und biologische Prozesse inklusive ihrer Unterkategorien.

Geschlecht	Jahr	social	family	friend	female	male	cogproc	insight	cause	discrep	tentat	certain	differ	bio	body	health	sexual	ingest
Frau	1996	0.69	0.02	0.04	0.14	0.05	0.19	0.06	0.05	0.02	0.03	0.02	0.04	0.22	0.07	0.05	0.03	0.07
Frau	1997	0.70	0.02	0.04	0.15	0.05	0.17	0.05	0.03	0.01	0.03	0.02	0.04	0.20	0.06	0.06	0.03	0.06
Frau	1998	0.70	0.03	0.05	0.15	0.05	0.16	0.05	0.03	0.01	0.03	0.02	0.03	0.17	0.05	0.04	0.02	0.07
Frau	1999	0.73	0.02	0.06	0.17	0.04	0.18	0.06	0.03	0.01	0.03	0.03	0.03	0.20	0.07	0.04	0.02	0.07
Frau	2000	0.72	0.02	0.05	0.13	0.08	0.15	0.04	0.03	0.02	0.02	0.03	0.03	0.15	0.05	0.03	0.01	0.06
Mann	1996	0.55	0.02	0.03	0.01	0.08	0.15	0.04	0.03	0.01	0.03	0.03	0.02	0.14	0.05	0.04	0.01	0.05
Mann	1997	0.52	0.02	0.02	0.01	0.07	0.16	0.03	0.03	0.01	0.03	0.04	0.03	0.15	0.05	0.04	0.02	0.05
Mann	1998	0.53	0.01	0.03	0.01	0.07	0.14	0.04	0.02	0.01	0.03	0.03	0.02	0.13	0.04	0.04	0.02	0.04
Mann	1999	0.51	0.01	0.03	0.01	0.07	0.12	0.03	0.02	0.02	0.03	0.03	0.02	0.14	0.05	0.03	0.02	0.04
Mann	2000	0.53	0.02	0.03	0.01	0.08	0.12	0.04	0.02	0.01	0.02	0.03	0.02	0.12	0.04	0.03	0.01	0.04
Frau		0.708	0.022	0.048	0.148	0.054	0.17	0.052	0.034	0.014	0.028	0.024	0.034	0.188	0.06	0.044	0.022	0.066
Mann		0.528	0.016	0.028	0.01	0.074	0.138	0.036	0.024	0.012	0.028	0.032	0.022	0.136	0.046	0.036	0.016	0.044
Differenz		0.18	0.006	0.02	0.138	-0.02	0.032	0.016	0.01	0.002	0	-0.008	0.012	0.052	0.014	0.008	0.006	0.022

Abb. 5: Ergebnisse der LIWC Analyse

Wie erwartet werden für die Artikel über Frauen mehr Wörter verwendet, die der Kategorie soziale Prozesse zugeordnet werden können, als in Artikeln über Männern. Der Unterschied beträgt 0,18%. Besonders auffällig ist dabei die Betrachtung der Anteile der weiblichen und männlichen Begriffe. Die Ergebnisse decken sich mit den Ergebnissen von Wagner et al. (2016), dass Frauenartikel eher weibliche Wörter enthalten als Männerartikel männliche Wörter enthalten. Dies kann durch ihre Hypothese des männlichen Geschlechts als Null-Gender begründet werden.

Die Annahme, dass Wörter der Kategorie kognitive Prozesse häufiger in Biographien über Männer verkommen, konnte nicht bestätigt werden. Der Anteil bei Artikeln über Frauen liegt mit 0,17%, 0,032% über dem Anteil bei Artikeln über Männer.

Ebenso konnte nicht bestätigt werden, dass bei Biographien über Frauen deutlich mehr Wörter der Kategorie biologische Prozesse verwendet werden. Zwar ist der Anteil bei Frauen größer als bei Männern, jedoch gibt es keinen signifikanten Unterschied.

5 Grenzen

In dieser Arbeit wird nur zwischen zwei Geschlechtern unterschieden: Frau und Mann. Dies entspricht nicht der Realität. Der Anteil von Artikeln, die keinem Geschlecht zugeordnet werden konnten, die also nicht die Kategorie Frau oder Mann besitzen, beträgt 0,17% (siehe Abbildung 2). Es wurde nicht untersucht ob diese Personen sich keinem der zwei Geschlechter zugehörig fühlen, die Kategorie Frau oder Mann unabsichtlich fehlt, oder ob diese Artikel keine Biographien sind und fälschlicherweise in die Kategorie des Geburtsjahrs eingeordnet wurden. Diese Artikel wurden in der Arbeit nicht betrachtet.

6 Fazit

Der Gender Bias in Wikipedia ist eine vielbetrachtete Thematik, die in verschiedenen Ebenen untersucht und festgestellt wurde. Wikipediaartikel in Form von Biographien über Frauen und Männer unterscheiden sich insbesondere hinsichtlich ihrer Sprache und Sichtbarkeit. Während bei Frauen familienbezogene Themen dominieren, sind bei Artikeln über Männer die Themen Sport und Beruf im Vordergrund. Die Sichtbarkeit unterscheidet sich vor allem in der Anzahl der Artikel pro Geschlecht und der Netzwerk Struktur der Artikel auf der Wikipedia Plattform. Der Gender Bias kann größtenteils auf die Autor*innen der Beiträge zurückgeführt werden. Durch eine starke Mehrheit männlicher Autoren (Gender Gap) wird ihr Bias verstärkt auf die Artikel übertragen. Insbesondere der implizite Bias, das Weglassen von Informationen, die eine Biographie komplett machen würde, ist bei der Betrachtung des Gender Bias wahrscheinlich.

In der Betrachtung deutschsprachiger Biographien der Geburtsjahre 1996 bis 2000 auf Wikipedia konnten Ergebnisse vorheriger Untersuchungen teilweise bestätigt werden. Es gibt auch in diesen Geburtsjahren einen starken Unterschied zwischen der Anzahl weiblicher und männlicher Biographien. Auch hinsichtlich der Textlänge ist bei Artikeln über Frauen ein Defizit zu erkennen. Dies bestätigt die zuvor getroffenen Annahmen, unterstützt die Theorie der gläsernen Decke und zeigt, dass Frauen im Durchschnitt nicht so ausführlich betrachtet werden wie Männer.

Hinsichtlich der Sprache konnten nicht alle Annahme bestätigt werden. Wie vermutet konnte in Texten über Frauen eine Häufung von sozial assoziierten Wörtern festgestellt werden. Insbesondere die Theorie des Mannes als Null-Gender konnte durch die Verhältnisse weiblicher und männlicher Wörter bestätigt werden. Die Annahmen, dass Männer im Anteil von Wörtern kognitiver Prozesse und Frauen im Anteil von Wörtern biologischer Prozesse höher abschneiden, konnte nicht nachgewiesen werden.

Insgesamt bestätigen die Untersuchen, dass ein Gender Bias im deutschsprachigen Wikipedia vorhanden ist. Hinsichtlich der Textanzahlen wurde eine Verbesserung im Vergleich zum gesamten deutschsprachigen Wikipedia festgestellt. Dies lässt eine weitere Angleichung des Verhältnisses, von Artikel über Frauen und Männer, für spätere Generationen vermuten. Die Untersuchung der Sprache zeigt, dass Unterschiede in den Kategorien kleiner werden, was bedeutet, dass sich die Sprache, die genutzt wird um Frauen und Männer zu beschreiben, immer weiter angleicht. Dies ist wichtig, um den Gender Bias zu minimieren und Leser*innen eine neutrale Sicht auf beschriebene Personen zu bieten.

Initiativen von Wikipedia, mehr Autorinnen für ihre Plattform zu gewinnen, würde die Angleichung der Betrachtung der Geschlechter weiter vorantreiben.

6.1 Ausblick

Diese Arbeit bietet Punkte, die zur Anknüpfung weiterer Untersuchungen geeignet sind.

Es könnte ein Vergleich zwischen verschiedenen Generationen durchgeführt werden. Werden verschiedene Generationen separat betrachtet, kann ein Trendanalyse in der Entwicklung der Betrachtung der Geschlechter durchgeführt werden.

Ein anderer Ansatzpunkt könnte die erneute Betrachtung der in der dieser Arbeit untersuchten Geburtsjahre, nach einer gewissen Anzahl von Jahren, sein. So könnte festgestellt werden, ob neue Artikel und Bearbeitungen von Artikeln zu einer ausgeglichenen Sicht auf die Geschlechter geführt haben

Literaturverzeichnis

- [CB12] Collier, B.; Bear, J. (2012): Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions. Proceedings of ACM 2012 conference on Computer Supported Cooperative Work.
- [Fü21] Fünffinger, A. (2021): Frauen verdienen 18 Prozent weniger, <https://www.tagesschau.de/inland/lohn-frauen-maenner-101.html>, 22.01.2022.
- [Gr15] Graells-Garrido, E. et al. (2015): First Women, Second Sex: Gender Bias in Wikipedia. Proceedings of 26th ACM Conference on Hypertext & Social Media.
- [HS13] Hill, B.; Shaw, A. (2013): The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation.
- [Hu17] Hube, C. (2017): Bias in Wikipedia. Proceedings of the 26th International Conference on World Wide Web Companion.
- [Me17] Meier, T. et al. (2018): “LIWC auf Deutsch”: The Development. Psychometrics, and Introduction of DE-LIWC2015.
- [RR11] Reagle, J.; Rhue, L. (2011): Gender Bias in Wikipedia and Britannica. International Journal of Communication 5, 2011.
- [SP11] Sun, J.; Peng, N. (2011): Men Are Elected, Women Are Married: Events Gender Bias on Wikipedia.
- [TP09] Tausczik, Y.; Pennebaker, J. (2009): The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods.
- [Wa15] Wagner, C. et al. (2015): It’s a Man’s Wikipedia? Assessing Gender Inequality in an Online Encyclopedia.
- [Wa16] Wagner, C. et al. (2016): Women through the glass ceiling: gender asymmetries in Wikipedia.