

Transformationen des Populären. Working Paper Series des SFB 1472

25.10.2024

Working Paper 17

Die Vermessung des Populären?

Explorative Eindrücke aus der datenbasierten Cooperative Research Group „Diskursmetriken des Populären – sprachlich-kommunikative Indikatoren für seine Transformation“

Vanessa Breitkopf
Laura Désirée Haas
Simone Schmid
Friedemann Vogel

Zitierweise:

Breitkopf, Vanessa, Laura Désirée Haas, Simone Schmid und Friedemann Vogel (2024): „Die Vermessung des Populären? Explorative Eindrücke aus der datenbasierten Cooperative Research Group „Diskursmetriken des Populären – sprachlich-kommunikative Indikatoren für seine Transformation“. *Working Paper SFB 1472*, no.17. DOI: <https://doi.org/10.25819/ubsi/10576>.

Abstract

Das explorative Projekt untersucht die quantifizierbaren sprachlichen Spuren des Populären in großen Textkorpora, wie Presseartikeln, Forenbeiträgen und Unternehmenswebsites. Ausgangspunkt der Studie ist die Annahme, dass rekurrente Sprachgebrauchsmuster – etwa Schlagwörter oder feststehende Wortkombinationen – sowohl soziale Praktiken konstituieren als auch Diskursmentalitäten widerspiegeln. Der Beitrag analysiert, inwieweit hochfrequente sprachliche Muster auf eine Wirksamkeit des Populären in Diskursen hinweisen und wie sich Veränderungen im Sprachgebrauch als Indikator für einen Wandel von Mentalitäten interpretieren lassen.

Schlagwörter: Diskursanalyse, Korpuslinguistik, Diskursmetrisierung, Popularität, Medien

The explorative project analyses the quantifiable linguistic traces of the popular in large text corpora, such as press articles, forum posts and company websites. The starting point of the study is the assumption that recurrent patterns of language use – such as catchwords or fixed word combinations – both constitute social practices and reflect discourse mentalities. The article analyses the extent to which high-frequency linguistic patterns indicate the effectiveness of the popular in discourses and how changes in language use can be interpreted as an indicator of a change in mentalities.

Keywords: Discourse analysis, corpus linguistics, discourse metrics, popularity, media

Kontakt:

Prof. Dr. Friedemann Vogel
Universität Siegen
E-Mail: friedemann.vogel@uni-siegen.de

1. Einführung: Von der Sprache zu Diskurs-mentalitäten

Welche quantifizierbaren Spuren hinterlässt das Populäre in sprachlichen Massendaten? Diese Frage beschäftigte uns im Rahmen der CRG „Diskursmetriken des Populären“ sowohl in methodischer als auch in analytischer Hinsicht im Winterhalbjahr 2023/24. Ausgangspunkt war und ist die konstruktivistische und diskurslinguistische Annahme, dass rekurrente Sprachgebrauchsmuster (Schlagwörter, feststehende und als solche erwartbare Mehrworteinheiten usw.) sowohl Konstituenten als auch indexikalische Spuren für soziale Praktiken darstellen: das von der Partei AfD programmatisch verwendete Schlagwort „Altpartei“ zur Bezeichnung aller anderen Großparteien etwa kommuniziert die Selbst- und Fremdpositionierung der AfD (Deus 2020) im politischen Raum; die frequente Zu- und Abnahme von Ausdrücken (Wortfeldern) zur Bezeichnung von Freund-Gegner-Feind gibt Auskunft über die demokratische Verfasstheit (Freiheitsgrade an Dialogizität) eines öffentlichen Diskurses (Vogel 2022); sogar die lexikalisch-syntaktischen Muster als Ganzes – als Gruppensprache oder Jargon – haben identitätsstiftendes Potential und können gemeinsame Interessen, Bedürfnisse, Ängste usw. signalisieren (Spitzmüller 2022). Sprachmuster wie diese finden sich in großen Textkorpora und lassen sich also als Ausdruck von Diskursmentalitäten (Hermanns 1995) interpretieren. Größere Änderungen im Sprachgebrauch einer Gruppe sind zugleich Indizien für Mentalitätswandel zunächst auf Seiten der beobachteten Sprecher-Schreiber (z. B. von Textproduzenten in Presse und Rundfunk), je nach medialer Reichweite der Kommunikate dann auch der Textrezipienten (etwa wenn rekurrenter Sprachgebrauch in Massenmedien die Wirklichkeitsannahmen ganzer Bevölkerungsteile prägt, vgl. Luhmann 2004/1995, Vogel 2010).

Vor diesem Hintergrund wollten wir explorativ prüfen, ob und inwiefern das Populäre in unseren verschiedenen Projektkorpora – Presstexte, Forenbeiträge, Unternehmenswebsites – in Form hochfrequenter Sprachgebräuche sichtbar wird. Finden sich in unseren Untersuchungsdaten quantifizierbare, d. h. durch algorithmusbasierte Analyseverfahren in Massentextdaten statistisch auswertbare Hinweise für eine zu- oder abnehmende Popularisierung zweiter Ordnung¹?

Die nachfolgenden Beobachtungen verstehen sich als erste suchende, auf Oberflächenstrukturen beschränkte Denkanstöße; eine detaillierte, kontextsensitive Untersuchung steht noch aus.

2. Methodisches Verfahren und Datengrundlage

Den konzeptionell-methodischen Rahmen unserer Überlegungen zur ‚Vermessung‘ öffentlicher Diskurse bildet der Ansatz der „Diskursmetrisierung“ (Vogel 2024). Dieser Ansatz knüpft an korpuslinguistische Diskursanalysen an, interessiert sich jedoch weniger für ‚lokale‘ semantische Kämpfe (wie etwa Felder / Luth / Vogel 2016), sondern für ‚globale‘ Häufigkeitsverteilungen in großen Textkorpora als Indizien für themen- bzw. ereignisübergreifende Denkmuster. Ausgangspunkt der Analyse ist

1 „Popularisierung zweiter Ordnung [im Sinne des SFB 1472 Transformationen des Populären] bezeichnet [vielmehr] solche Popularisierungsverfahren, die Populäres herstellen, indem sie seine Beachtung durch viele feststellen und ausstellen. Dies geschieht durch die Veröffentlichung von quantifizierten Popularitätsrelationen, d. h. durch öffentliche Charts, Rankings und Listen, die wiederum populär gerahmt werden: Hitparaden im Radio und Fernsehen, Einschaltquoten verschiedener Sender im Vergleich, mit Spannung erwartete Bestenlisten, Preisverleihungen für quantitative Beachtungserfolge, Einbindung von Erfolgswerten in die Produktwerbung, Rankings von Museen, Politikerinnen, Universitäten, die allfällige Veröffentlichung von Ergebnissen der Markt- und Meinungsforschung.“ (Döring et al. 2021: 13)

die Entwicklung einer Minimalhypothese, das heißt die Vorauswahl von Ausdrücken, die zumindest auch mit dem zu untersuchenden Zielkonzept assoziiert sind. In unserem Fall stellt sich die Frage, welche Ausdrücke Hinweise auf das Konzept des Populären – vor allem in seiner Ausprägung als Popularisierung zweiter Ordnung – geben könnten. Hierzu wurde zunächst introspektiv – aus analytischen Vorarbeiten in SFB und Teilprojekten – eine Liste an Wörtern und Mehrworteinheiten gesammelt, mit denen Diskursakteure typischerweise eine große oder kleine, eine zu- oder abnehmende Popularität von Sachverhalten markieren. Hierzu zählen insbesondere Wörter, die Beachtung durch Viele explizit thematisieren (z. B. etwas sei *populär*, *im Trend*, *beliebt*, *vielbeachtet*, *gehe viral* oder *finde große Aufmerksamkeit*, *Publikumsresonanz* u. ä.), aber auch Substantive, Phrasen oder Namen von Referenzobjekten, die schematisch als ‚populär‘ (gleich, ob erwünscht oder problematisiert) oder als ‚Popularitätsindikator‘ im kulturellen Gedächtnis etabliert sind (z. B. etwas sei ein *Klassiker*, *Hype*, *Welterfolg*, *Skandal*, *Negativimage*, *Shitstorm* oder eine *Ära*; *Charts*, *Andrang*, *Rating*, *Hitparade*, *Fangemeinde*, *Kassenhit* usw.). Zugleich – allerdings im Rückblick weniger systematisch – wurden Ausdrücke berücksichtigt, die weniger Popularität, als vielmehr qualitative Wertattribute konstituieren (z. B. etwas sei *ehrenwert*, *(in)diskutabel*, *erstklassig*, *erhellend*, *überragend* usw.). Erweitert wurde diese introspektiv gebildete Liste mit potenziell gebrauchswandten Ausdrücken mithilfe der Kookkurrenzdatenbank CCDB (Keibel / Belica 2007), einer Datenbank, die Wortgebrauchsprofile von über 200.000 Wörtern des Deutschen induktiv und statistisch miteinander vergleichen kann und damit Gebrauchsähnlichkeit (Quasisynonyme, Metonyme) empirisch ermittelt. Die daraus entstehende Liste an Ausdrücken umfasste über 600 Wörter und Mehrworteinheiten, die den Skopus unserer Untersuchung zugleich erweiterte, andererseits auch größere Unschärfen hineinbrachte (vgl. etwa Ausdrücke wie *Rufmord*, *Liebe*, *Liga*).

Anschließend wurden für möglichst alle dieser Ausdrücke Häufigkeitsverteilungen in den beteiligten Teilprojekt-Korpora erhoben und – sofern häufiger als 10-mal belegt – grafisch ausgewertet. Von einer Aggregation der Häufigkeitsverteilung (also der Bildung von Summen über mehrere Ausdrucksfrequenzen hinweg) haben wir Abstand genommen, weil uns die semantische Verteilung der gesammelten Ausdrucksliste zu diffus schien; wir haben stattdessen versucht, die einzelnen Verteilungen kontrastiv und qualitativ einzuordnen.

Die herangezogenen Korpora waren:

- **Diachrones Pressekorpus** (Teilprojekt C04): LeMonde Diplomatie / LMD: 12,0 Mio. Token – 1995 bis 2020; Der SPIEGEL / SPON: 238,0 – 1947 bis 2019; Jungle World / JW: 40,6 – 1997 bis 2020; Junge Freiheit / JF: 59,8 – 1997 bis 2023; Ladenburger Zeitung / LAZ: 50,5 – 1950 bis 2021; Süddeutsche Zeitung / SZ: 746,4 – 1945 bis 2020; Neues Deutschland / ND: 177,1 – 1999 bis 2011.
- Beiträge aus dem von 2012 bis 2024 existierenden **Perry Rhodan-Forum** (Teilprojekt A01): Spoilerbereich zu Zyklus „Neuroversum“ (Heft 2600–2699), Zyklus „Das Atopische Tribunal“ (Heft 2700–2799), Zyklus „Die Jenzeitige Lande“ (Heft 2800–2874), Zyklus „Sternengruft“ (Heft 2875–2899), Zyklus „Genesis“ (Heft 2900–2999), Zyklus „Mythos“ (Heft 3000–3099).

– Textinhalte von Unternehmenswebsites von 871 sogenannten Hidden Champions (Teilprojekt B06). Hidden Champions sind oft inhaber- bzw. familiengeführte mittelständische Unternehmen, die unbemerkt von der heimischen Öffentlichkeit als Weltmarktführer in engen Nischenmärkten agieren und die eine durch die Wissenschaft mitgeprägte populäre Marke darstellen, die auch das öffentliche Interesse weckt (Pahnke / Welter 2019).

Die insgesamt 3712 semiautomatisch generierten Verlaufsdiagramme zum (ausgewählten) Sprachgebrauch in den Presstextkorpora wurden zum einen einzeln betrachtet, zum anderen mithilfe des weit verbreiteten Clustering-Verfahrens *KMeans* (VanderPlas 2017) bzw. seiner Implementierung in der Python-Bibliothek *tslearn* nach Ähnlichkeit (*euclidean distance*) gruppiert. Die Zeitreihen wurden zuvor z-standardisiert. Im Ergebnis lagen drei verschiedene Verlaufsgrafiken für die induktive Erschließung vor: ein Verlaufsdiagramm, das die absolute und relative Häufigkeit eines einzelnen Ausdrucks vor dem Hintergrund der Teilkorpusgrößen darstellt (Abb. 1); korpusbezogene Matrizen mit Verlaufsdiagrammen zu allen Ausdrücken, jedoch in einer Form, die Frequenzunterschiede über die Zeit stark überbetont (zwecks Tendenzerkennung und Vergleich; Abb. 2); eine Matrix aus Clustern von ähnlichen Verlaufsdiagrammen als Hilfsmittel zur leichteren Orientierung (Abb. 3).

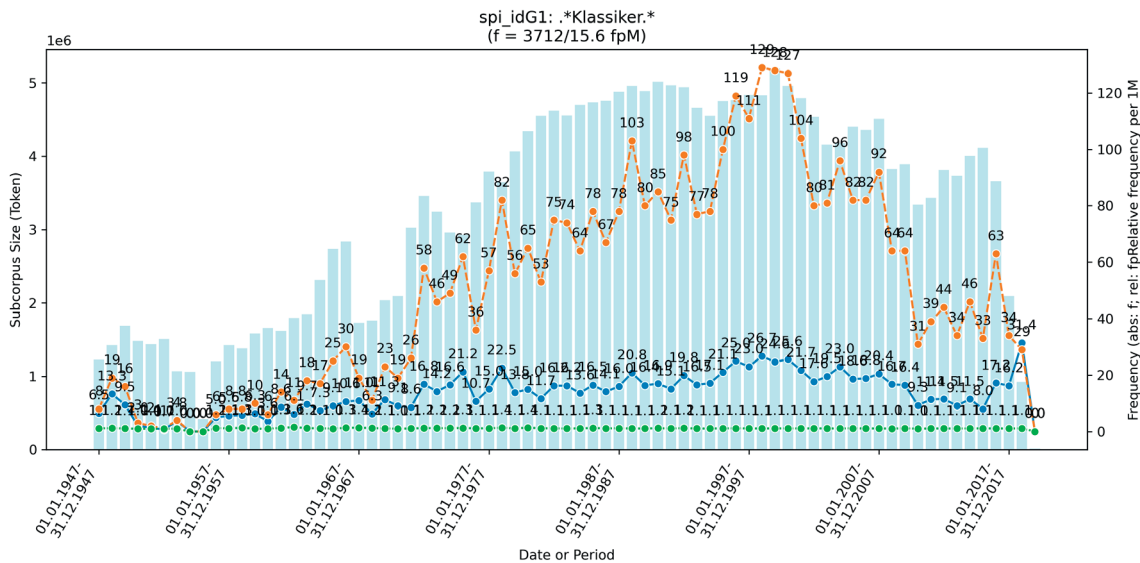


Abb. 1: Verlaufsdiagramm zum regulären Ausdruck .*Klassiker.* im Korpus des SPIEGEL (inkl. Online) nach Jahren. Blaue Säulen: Subkorpusgröße in Token; gelbe Linie: absolute Häufigkeit; blaue Linie: relative Häufigkeit; grüne Linie: durchschnittliche Häufigkeit pro Text.

Abb. 2: Matrix mit Verlaufsdiagrammen zur Häufigkeit verschiedener Ausdrücke im SPIEGEL-Korpus.

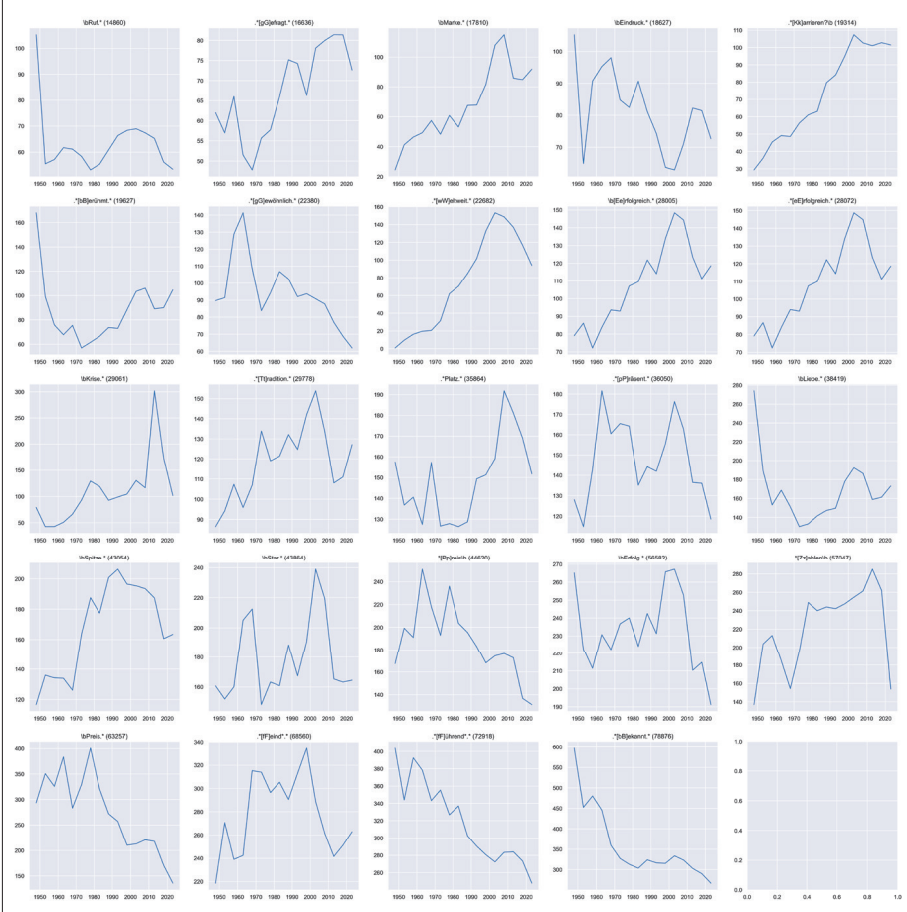
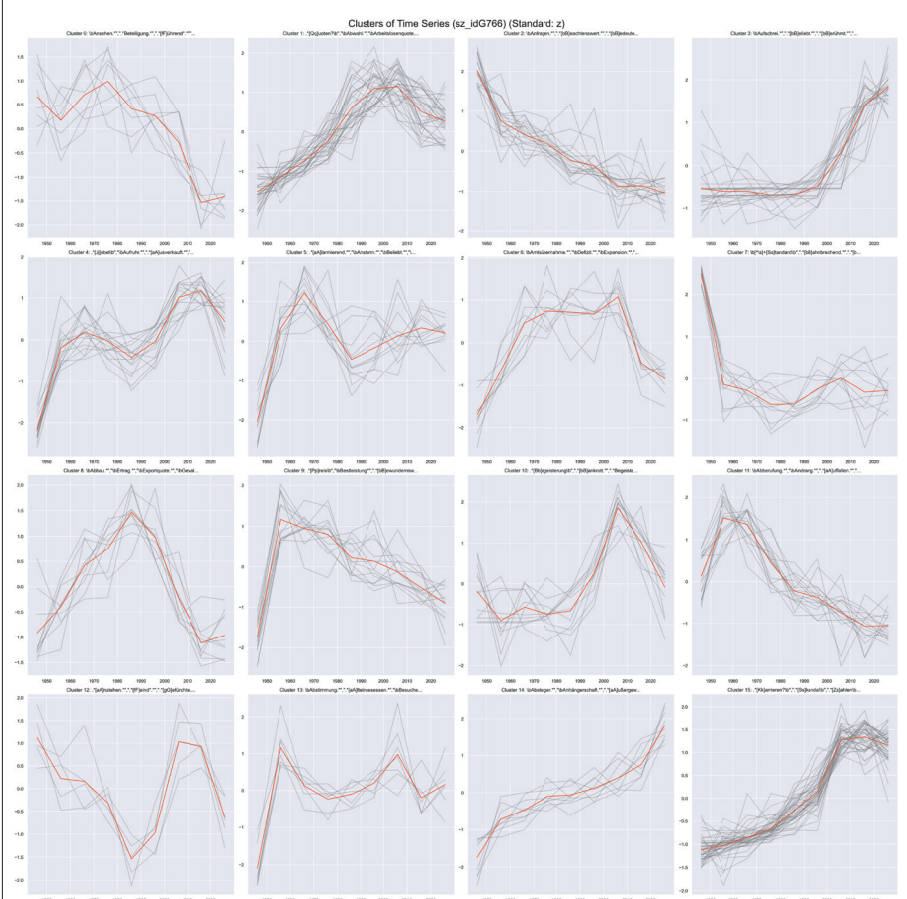


Abb. 3: Cluster ähnlicher Verlaufsdiagramme zum Korpus der Süddeutschen Zeitung als Ergebnis eines kmeans-Clusterings; die rote Linie repräsentiert den Durchschnitt (arith. Mittel) aller Verläufe im jeweiligen Cluster.



Die Häufigkeitsanalyse der Beiträge des *Perry Rhodan*-Forums wurde in zwei Schritten durchgeführt: Erstens wurden die Frequenzen der ausgewählten Ausdrücke für den gesamten Korpus ausgewertet und zweitens wurde für jeden Spoilerbereich eine eigene Auszählung vorgenommen. Ein Zyklus beschreibt jeweils einen eigenen, abgeschlossenen Zeitabschnitt: „Neuroversum“ erschien von 2011 bis 2013, „Das Atopische Tribunal“ von 2013 bis 2015, „Die Jenzeitigen Lande“ von 2015 bis 2016 usw. Mit Beginn eines neuen Zyklus wurden die Unterforen des vorherigen kaum noch bis gar nicht mehr bespielt und in ein Archiv verschoben. Die isolierte Analyse ermöglichte somit in einer vergleichenden Gegenüberstellung einen Entwicklungsverlauf der Häufigkeiten.

Um zu untersuchen, welche Ausdrücke auf den Unternehmenswebsites von sogenannten Hidden Champions als Indikatoren für Popularität verwendet werden, wurden die Häufigkeitsverteilungen ermittelt und induktiv in Kategorien eingeteilt, die quantitative, qualitative, hierarchisierende, und komparative (d. h., domänenspezifisch valorisierende) Bewertungsregime abbilden.

3. Ergebnisse der induktiven Datensichtung

Bei der Auswertung zeigt sich, dass – ohne eine weitere qualitative Detailanalyse in den Texten – nicht alle Verteilungen plausibel eingeordnet werden können. Außerdem ist erst bei Sichtung aufgefallen, dass sich bei der (teilweise automatisch erfolgten) Formulierung von Suchanfragen zuweilen Fehler eingeschlichen haben: manche Suchanfragen (vor allem solche, die mittels Platzhalter auch flektierte Formen oder Komposita einbeziehen sollen, erzeugten zu viele heterogene und/oder falsch-positive Treffer (Beispiel: **gehoben.** → *hervorgehoben*), bei anderen haben wir Varianten in der Schreibung (z. B. infolge Rechtschreibreform) übersehen. Auf Basis von Konkordanzen (zeilenweise Sichtung von Treffern) konnten solche fehlerhaften Verteilungen ausgeschlossen bzw. Häufigkeitsverteilungen korrigiert werden.

Folgende Beobachtungen finden sich in jenen **Pressedaten**, für einen Zeitraum ab Mitte des 20. Jahrhunderts vorliegen. *Süddeutsche Zeitung* (SZ) und *Der Spiegel* / *Spiegel Online* (SPIEGEL) verhalten sich dabei tendenziell ähnlich; die *Ladenburger Zeitung* (LAZ) als Lokalzeitung weicht ein Stück weit davon ab (u. a. auch deshalb, weil die Zeitung über die Jahrzehnte hinweg mehrere tiefgreifende Veränderungen erfuhr).

(1) Die folgenden Ausdrücke (mit einer Mindestfrequenz von 1000) weisen eine stetig steigende Häufigkeitsverteilung in SZ und SPIEGEL auf, das heißt, sie werden über die Zeit seit Mitte des 20. Jahrhunderts tendenziell immer häufiger verwendet (gemessen in relativen Zahlen):

gewöhnlich [v. a. negiert: *un/außer-*], *angefeindet*, *Anhängerschaft*, *Ära*, *Aufmerksamkeit-erregen*, *Aufbruch*, *Aufschrei*, *beäugt*, *bedenkenswert*, *bejubelt*, *Bekanntheit*, *berühmt-berüchtigt*, *Berühmtheit*, *Bestseller*, *Bestsellerliste*, *brisant*, *Bürgerbeteiligung*, *Charts*, *Eklat*, *erstaunlich*, *etabliert*, *exzellente*, *Fangemeinde*, *Fehltritt*, *Forbes*, *gelistet*, *global*, *hochgelobt*, *Hype*, *Image**, *im-Trend*, *Klassiker* [als populär ausgezeichnete Phänomene], *kontrovers*, *Krise*, *Kultstatus*, *legendär*, *Mainstream*, *Marke*, *populär*, *prägend*, *Präsenz* [v. a. Problematisierung lokaler Präsenz], *provokant*,

Publikumsliebbling, Ranking [ab 2000er], Rating, Renommee, Reputation, Resonanz [mit Rückgang 2000er], Score, Shitsstorm, Skandal [deutlich zunehmend ab 1980er], skandalös, Social Media, Soziale Medien, Star, Top-X/Top-Ten/Top 100, umstritten, verbreitet, Verkaufshit, Verkaufsschlager, Verkaufszahlen, verrissen, verspottet, virulent, vogue, wahrgenommen, Warteschlangen, weltweit, Zulauf

Bei diesen Ausdrücken handelt es sich um Wörter bzw. Mehrworteinheiten, die

- eine große oder kleine / zu- oder abnehmende Beobachtung, Konflikte oder Stereotype thematisieren;
- Phänomene der Internet-Kultur bezeichnen;
- oder auf Außeralltägliches, Unerwartetes, Ungewöhnliches verweisen.

Interessant sind dabei zum Beispiel Ausdrücke, die auf den ersten Blick ein Qualitätsattribut im Sinne der Popularisierung erster Ordnung² konstituieren, wie *Klassiker*. Bei einem Blick in die Belege (in allen Medien) zeigt sich aber schnell, dass vor allem populärkulturelle Phänomene (und gerade nicht die als bildungsbürgerliche ‚Hochkultur‘ konnotierten Artefakte oder Themen) als *Klassiker* bezeichnet und als ‚vielbekannt‘ bzw. ‚vielgemocht‘ ausgewiesen werden:

- *große Spielesammlung mit allen Klassikern*
- *Ein Krimi-Klassiker, mit viel Witz und Charme von der TiL auf die Bühne gebracht*
- *und das Triathlonfestival (23. 7.) sind längst zu Klassikern geworden*
- *Ein Klassiker der Unterhaltung mit beliebten Künstlern: Magda Schneider*
- *Neben dem Klassiker Scholle mit Specksoße und neuen Kartoffeln*

(2) Die folgende Liste an Ausdrücken zeigt eine (in relativen Zahlen gemessene) stetig abnehmende Häufigkeitsverteilung über die Zeit hinweg. Dabei fällt auf, dass viele dieser Wörter Qualitätsattribute darstellen und als Teil von Werturteilen fungieren, wie wir sie in einer High-Low-Axiomatik im Sinne des SFBs erwarten würden. Die Liste ist jedoch heterogen: *Publikumsinteresse* thematisiert explizit Beachtung.

[in!]diskutabel, anerkennenswert, Ansehen, außerordentlich, beachtenswert, beachtlich, Beachtung, bedenklich, bedeutend [~10% alle Belege negierend], bedeutungsvoll, begrüßenswert, beklagenswert, bewundernswert, brauchbar, Ehrenpreis, erhaltenswert, erstklassig, gebräuchlich, hervorzuheben, lobenswert, Publikumsinteresse, Sensation, unbeachtet, verehrt, Verehrung, Verkaufserfolg, weltbekannt, weltberühmt, wohlbekannt

(3) Manche relativ hochfrequent genutzten Ausdrücke zeigen eine unimodale Verteilung über die Zeit, das heißt: sie nehmen in der Häufigkeit zunächst zu und fallen nach einem Peak dann wieder.

Popularität [Rückgang seit 2000er], Hitparade, hochgeschätzt, gepriesen, förderungswürdig, dominierend, Bürgernähe [meist in problematisierender Schlagwort-Verwendung: mehr/Defizit an Bürgernähe usw.], Beteiligung, Bekanntheitsgrad, Besucherandrang/-ansturm/-aufkommen, attraktiv

2 „Diese Aushandlungen zwischen Experten und Laien fasst das Forschungsprogramm „Transformationen des Populären“ als Popularisierung erster Ordnung, deren Formen, Verfahren, aber auch deren Risiken auf Seiten der Eliten zumeist im Modell von Diffusion und Defizit beschrieben werden (Hilgartner 1990; Schwarz 1999): Um für die Inklusion des Volkes im Dienste höherer Ziele zu sorgen, müssten die legitimen Kultur- und Wissensbestände vereinfacht und verbreitet werden – z. B. durch Ratgeberliteratur, Volksausgaben, Benimmbücher, Klassikersynopsen, Opern- und Reiseführer oder günstige Reproduktionen.“ (Döring et al. 2021: 12)

(4) Eine kleinere Anzahl an Ausdrücken – **beliebt** [mit steigender Grundtendenz], *Beliebtheit, bemerkenswert, begehrt, begeistern, begeistert* – zeigen zwei ausgeprägte Peaks.

(5) Daneben gibt es eine Vielzahl an Ausdrücken, die vorübergehende Popularität erreichen und / oder an bestimmte Typen von historischen Diskursereignissen (v. a. ‚Problem‘-Ereignisse) geknüpft sind, etwa *Aufbruch, aufgeflammt, Bankrott*.

(6) Die Ladenburger Zeitung hat sich über die Jahrzehnte hinweg von einem vielseitigen Blatt mit eigener internationaler und bundesweiter Berichterstattung zu einem kleinen Kleinstadtblättchen entwickelt. Die Ausdrücke (unserer Liste) mit deutlicher steigender oder fallender Gebrauchsfrequenz vermitteln kein klares Bild:

- steigend: *außergewöhnlich, ausverkauft, beeindruckend* [steigt stärker als bei anderen Medien], *herausragend, Resonanz, Klassiker* [ab 2000], *kontrovers* [aber nur verhalten steigend]
- fallend: *verbreitet, erstaunlich, umstritten, populär, Skandal*

Nur einzelne Verteilungen lassen sich einordnen: Mit der Verkleinerung und dann dem Wegfall eines internationalen / bundesweiten Berichtsteils geraten auch weit(er)reichende Diskursereignisse (*Skandal*) eher aus dem Blickfeld.

(7) Soziale Medien und der damit einhergehende Bedeutungszuwachs für Beobachtungswahrnehmung spielen erwartungsgemäß erst in den 2000er Jahren eine Rolle in der Presse, allerdings mit durchaus überraschenden Unterschieden zwischen den einzelnen Medien:

- *Blogs* allg.: vor allem linksliberale Medien wie *Jungle World* (f = 42,0 per Mio) und *LMD* (19,2), aber auch *JF* (18,6); dagegen kaum in *SPIEGEL* (4,4) *ND* (4,3), *SZ* (1,4) oder *LAZ* (0,3)
- *Twitter* (in allen Zeitungen, aber mit Abstand am häufigsten in *JW* mit 42,1 fpM und *JW* mit 41,9 gegenüber *SPIEGEL* (8,2), *LMD* (7,4), *SZ* (5,1)
- *Telegram* findet Nennung seit den 2000er Jahren vor allem bei *JF* (5,4), *JW* (3,0; zunehmend) und *LMD* (2,1; abnehmend)
- *LinkedIn, Pinterest, Snapchat, Twitch*: vor allem *JF* (in den Übrigen vernachlässigbar oder gar nicht)
- *Tumblr* und *Tinder*: insg. sehr selten, aber am ehesten in der *JW*
- *Reddit*: *JW* (0,7), *JW* (0,5), *SPIEGEL* / *SZ* (je 0,1)
- *Weibo*: nur sehr selten in der *LMD*

Es fällt auf, dass die rechtskonservative Wochenzeitung *JF* nicht nur deutlich häufiger auf soziale Medien rekurriert als die anderen, sondern auch ein größeres Spektrum aufnimmt. Blogs, aber auch *Tumblr* sind tendenziell eher ein Metier linksliberaler Medien.

In den Beiträgen des offiziellen Forums zur Science-Fiction-Heftromanserie *Perry Rhodan* zeigt sich, dass dort die Popularitätsindikatoren kaum bis gar nicht verwendet werden, die in den anderen Korpora häufig auftauchen. Dies scheint plausibel, da es sich bei den untersuchten Daten um Diskussionen über den jeweils aktuellen Zyklus einer Serie in einem mehr oder

weniger geschlossenem Raum handelt, an dem sich nur nach Anmeldung beteiligt werden konnte. Das Korpus ist zudem wesentlich kleiner. Anders als in sozialen Medien oder Pressedaten geht es im Forum nicht darum, den eigenen Text oder die Serie weit oder plattformübergreifend zu verbreiten und vor eine breite Masse zu stellen; es geht darum, in einer gefestigten Gruppe an Leser:innen und vor der Redaktion, den Autor:innen, etc. Beachtung durch qualitative Beiträge zu erhalten – und sich so als angesehene Expert:innen oder Kritiker:innen hervorzuheben.

Es ist demnach nicht verwunderlich, dass Charts oder Ratings, aber auch populäre soziale Medien wie Facebook, Twitter /X oder Instagram in dieser Diskurskultur kaum eine Rolle spielen, da sich der Aktionsraum auf das Forum beschränkt. Eine einigermaßen hohe Frequentierung des Ausdrucks „X“ lässt sich beispielsweise nicht auf die gleichnamige Plattform, sondern auf serienspezifische Ausdrücke wie Namen von Raumschiffen („X-1“, „Xaxata“) oder Bezeichnungen sowie auf diskursspezifische Ausdrücke wie allgemeine Verknüpfungen („Sonnensystem X“, „X Romane“) oder auf Multiplikationszeichen für technische Berechnungen zurückführen. Diesem Prinzip folgen die relativ hochfrequent genutzten Ausdrücke im Forum (Mindestfrequenz von 800):

(1) Serienspezifische Ausdrücke beinhalten beispielsweise *unsterblich*, *Star* (Raumschiff *Stardust*), *Hype* (*Hyperraum*) oder *Feind*.

Interessant ist dabei die Häufigkeitsverteilung je nach Zyklus. Dies weist daraufhin, dass je häufiger ein Ausdruck im jeweiligen Zyklus-Bereich vorkommt, desto wahrscheinlicher ist es, dass dieser Teil der aktuellen Zyklus-Handlung ist – und, dass dieser von Forist:innen wahrgenommen wird.

(2) Diskursspezifische Ausdrücke wie *Forum* (auch *Forist*), *Liebe* (oftmals als Anrede wie *Liebe:r XY* oder als Abschiedsgruß wie *Liebe Grüße*), *Eindruck* (*mein Eindruck ist*), *führend* (*irreführend*, *zielführend*, *wortführend*, *weiterführend*), *un/bekannt* (*viel/wenig bekannt*, *Fakten, die bekannt sind*).

Es ist dabei auffällig, dass emotional aufgeladene Wörter wie *Liebe* auch dazu verwendet werden, um etwas als sehr gelungen darzustellen (*Ich liebe es*). Die Einleitung in einen serienspezifischen Diskurs erfolgt größtenteils z. B. durch *interessant*. Dies kann als Indikator verwendet werden, um herauszufinden, was im Text bei vielen im Forum Beachtung findet. Begriffe wie *bekannt* werden dazu genutzt, um die Expertise der Forist:innen darzustellen (*es ist allgemein bekannt*) und sie von Lai:innen (*es ist unbekannt*) zu differenzieren. Die *high/low*-Axiologie reproduziert sich somit im Populären.

In der serienspezifischen Kommunikation eines Forums werden eigene Popularitätsindikatoren und eine eigene Popularitätslogik entwickelt, die immer wieder auf den Serientext sowie den Diskurs zurückverweisen. Es gibt jedoch Ausnahme wie *Platz*. In diesem Fall wird zwar größtenteils auf die Serie (*neuer Platz einnehmen*, *Platz im Kosmos*) oder auf eigene Eindrücke (*nicht mehr so viel Platz im Detail*) verwiesen, allerdings werden gelegentlich auch Ranking-Hinweise dadurch beschrieben (*Aktuell Platz 33 der Amazon Bestsellerliste E-Books*).

Im Teilprojekt B06 wurden keine diachronen Häufigkeitsverteilungen analysiert, wie in den Teilprojekten C04 und A01. Die Datenbasis bilden Unter-

nehmenswebseiten – ein zentraler Kommunikationskanal, um mit Kund:innen, (potenziellen) Mitarbeiter:innen und der Öffentlichkeit in den Dialog zu treten. Im Gegensatz zu den Presstexten obliegt es den Unternehmen selbst, wie sie sich präsentieren und welche Signale sie über ihr Unternehmen senden. Die Selbstdarstellungen der Unternehmen auf Websites (digitale Visitenkarte) ist ein wesentlicher Aspekt ihrer organisationalen Identität (DiMaggio / Powell 1983); die Art der Selbstdarstellung wiederum beeinflusst, wie Unternehmen von außen und innen wahrgenommen werden. Von besonderer Bedeutung ist hierbei die sprachliche Ausgestaltung. Beispielsweise ermöglichen Metaphern und Sentiments die Vermittlung spezifischer Merkmale (Luri / Schau / Ghosh 2023), wodurch die Identifikation mit den jeweiligen Unternehmensbildern erleichtert oder erschwert werden kann. Die öffentliche Darstellung von Auszeichnungen oder die Referenzierung einer Spitzenposition kann das Image und die Reputation eines Akteurs beeinflussen (Sharkey / Kovacs / Hsu 2022). Die am häufigsten vorkommenden Worthäufigkeiten anhand der Unternehmenswebsites lassen sich in Indikatoren unterteilen, die auf

- quantitative Beachtung (z.B. *global, weltweit, bekannt, Bekanntheit, Präsenz, präsent*),
- qualitative Beachtung (z.B. *Begeisterung, begeistert*),
- Hierarchie (z.B. *Top-, Platz, Ranking*),
- Erfolg (z.B. *Weltmarktführer, Marktführer, Champion, Expansion, Erfolg, erfolgreich, Vorreiter, Bestseller*) und
- Status (z.B. *Tradition, Ruf, etabliert*) verweisen.

Zudem sind Indikatoren, die auf soziale Medien und Plattformen (z.B. *Xing, WhatsApp, YouTube, Facebook, Twitter, LinkedIn*) verweisen bzw. verlinken festzustellen. Diese Tatsache ist nicht überraschend, da es sich hierbei um Standardfunktionen von Websites handelt.

Hidden Champions betonen in ihrer Selbstdarstellung Aspekte, die im Sinne der *high-Axiologie* (Döring et al. 2021) nicht nur ihre Qualitätsführerschaft und die Qualität ihrer Produkte und / oder Dienstleistungen, sondern auch Werthaltungen unterstreichen.

In der Gesamtschau lässt sich festhalten, dass sich die Popularisierung von Hidden Champions, die anhand von Unternehmenswebsites untersucht wurde, sich nicht gänzlich von der Popularisierung erster Ordnung löst. Unternehmen greifen in ihrer sozialen Identitätskonstruktion auf Verweise zurück, die sowohl markante, quantifizierbare Merkmale als auch qualitative Wertattribute aufgreifen. Eine Umkehr der Beweislast ist somit nur tentativ feststellbar, da sowohl qualitative als auch quantitative, komparative und hierarchisierende Bewertungsregime ineinandergreifen.

Entrepreneurship-Studien setzen zunehmend Forschungstechniken ein, die über interpretative Analysen hinausgehende, quantitative Auswertungen ermöglichen. Dies umfasst unter anderem computergestützte Sprachanalysen (beispielsweise zur Identifizierung von Persönlichkeitsmerkmalen von Superstar-Unternehmern vgl. Obschonka / Fisch / Boyd 2017, zur Zusammenhangsermittlung von Sprachstil von Crowdfunding-Erfolg vgl. Parhankangas / Renko 2017) sowie korpuslinguistische Diskursanalysen (Chandra 2017; Roundy / Asllani 2019). Auch Machine-Learning (ML)-

Ansätze gewinnen in der Entrepreneurshipforschung zunehmend an Bedeutung (Obschonka / Audretsch 2020). Die Fortschritte im Bereich des maschinellen Lernens eröffnen neue methodische Möglichkeiten und Techniken für die Konstruktion von Modellen zur Vorhersage verschiedener menschlicher Verhaltensweisen und bieten differenzierte Einblicke in die Vorhersagbarkeit unternehmerischer Aktivitäten. Der Einsatz computergestützter (Sprach-)Analysen auf Basis großer Textkorpora (frei verfügbare Daten in sozialen Medien) eröffnet die Möglichkeit, eine ähnliche prognostische Validität zu erreichen wie auf Basis etablierter Selbstberichte mittels Fragebogen (Boyd / Pennebaker 2017; Obschonka / Lee / Rodríguez-Pose / Eichstaedt / Ebert 2020). Neben den Vorteilen wird jedoch zunehmend konstatiert, dass ML-Modelle oft eine Black-Box darstellen (Rudin 2019). Auch wird darauf hingewiesen, dass eine unkritische Reflexion von Datenquellen, Indikatoren und daraus folgende Konsequenzen die Theorieentwicklung beeinträchtigen können (Lindebaum / Moser / Islam 2024). In diesem Zusammenhang sei auf die Bedeutung interdisziplinärer Arbeiten (Obschonka / Audretsch 2020) hingewiesen, wie sie im SFB-Verbund und der CRG geleistet werden. Die CRG-Arbeit verdeutlicht, dass die Interpretationsfähigkeiten der Forschenden von entscheidender Bedeutung für den integrierten Forschungsprozess unter Nutzung von ML-Ansätze sind: Das menschliche Verständnis und die logischen Denkprozesse stehen im Mittelpunkt der ML-basierten Theoriebildung und -prüfung (Lévesque / Obschonka / Nambisan 2020).

4. Fazit

Die hier skizzierten Beobachtungen verstehen sich als Eindrücke aus der diskursmetrischen CRG, die zu weitergehenden Anschlussuntersuchungen einladen sollen.

Der globale Blick auf die diachronen Häufigkeitsverteilungen von ausgewählten Ausdrücken in Presstexten gibt Indizien für einen Mentalitätswandel über die vergangenen 50 Jahre: viele diskursreflexive Wörter und Mehrworteinheiten, die die zunehmende oder nachlassende (öffentliche) Wahrnehmung eines Sachverhalts thematisieren, nehmen tendenziell oder deutlich zu.

Die Beobachtung des zeitlichen Verlaufs von Häufigkeitsverteilungen ausgewählter Ausdrücke in Beiträgen eines Forums, an denen unterschiedliche Akteursgruppen wie Redakteur:innen, Autor:innen, Zeichner:innen, Leser:innen und Fans beteiligt sind, gibt Aufschlüsse über die diskursive Entwicklung serieneigener Popularitätsindikatoren. Sie markiert ebenso, wann und wie eine Serie evolviert. Der zuvor skizzierte Mentalitätswandel in Presstexten deutet im Perry Rhodan-Kontext auf das Anregen oder den Abbruch von serien-, aber auch fandomspezifischer Aushandlungs- und Stabilisierungsprozesse hin.

Die Analyse der Ergebnisse, welche Signale sogenannte Hidden Champions via Unternehmenswebsites im öffentlichen Diskurs nutzen, legt nahe, dass Unternehmen in ihrer sozialen Identitätskonstruktion auf Verweise zurückgreifen, welche sowohl markante, quantifizierbare Merkmale als auch qualitative Wertattribute miteinbeziehen.

5. Literatur

11/13

Boyd, Ryan L. und James W. Pennebaker (2017): „Language-based personality: a new approach to personality in a digital world“, in: *Current Opinion in Behavioral Sciences* 18, S. 63–68. DOI: <https://doi.org/10.1016/j.cobeha.2017.07.017>.

Chandra, Yanto (2017): „Social Entrepreneurship as Institutional-Change Work: A Corpus Linguistics Analysis“, in: *Journal of Social Entrepreneurship* 8 (1), S. 14–46. DOI: <https://doi.org/10.1080/19420676.2016.1233133>.

Deus, Fabian (2020): „Altpartei“, in: *Diskursmonitor. Glossar zur strategischen Kommunikation in öffentlichen Diskursen*. Hg. von der Forschungsgruppe Diskursmonitor und Diskursintervention. Veröffentlicht am 02.07.2020. URL: <https://diskursmonitor.de/glossar/altpartei>.

DiMaggio, Paul J. und Walter W. Powell (1983): „The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields“, in: *American Sociological Review* 48 (2), S. 147–160. DOI: <https://doi.org/10.2307/2095101>.

Döring, Jörg, Niels Werber, Veronika Albrecht-Birkner, Carolin Gerlitz, Thomas Hecken, Johannes Paßmann, Jörgen Schäfer, Cornelius Schubert, Daniel Stein und Jochen Venus (2021): „Was bei vielen Beachtung findet. Zu den Transformationen des Populären“, in: *Kulturwissenschaftliche Zeitschrift* 6 (2), S. 1–24. DOI: <https://doi.org/10.25969/mediarep/18843>.

Felder, Ekkehard, Janine Luth und Friedemann Vogel (2016): „„Patientenautonomie“ und „Lebensschutz“. Eine empirische Studie zu agonalen Zentren im Rechtsdiskurs über Sterbehilfe“, in: *Zeitschrift für germanistische Linguistik* 44 (1), S. 1–36. DOI: <https://doi.org/10.1515/zgl-2016-0001>.

Hermanns, Fritz (1995): „Sprachgeschichte als Mentalitätsgeschichte. Überlegungen zu Sinn und Form und Gegenstand historischer Semantik“, in: Andreas Gardt, Klaus J. Mattheier und Oskar Reichmann (Hg.): *Sprachgeschichte des Neuhochdeutschen. Gegenstände, Methoden, Theorien*. Tübingen, S. 69–102.

Keibel, Holger und Cyril Belica (2007): „CCDB: A Corpus-Linguistic Research and Development Workbench“, in: *Proceedings of the 4th Corpus Linguistics conference*. Birmingham.

Lévesque, Moren, Martin Obschonka und Satish Nambisan (2020): „Pursuing Impactful Entrepreneurship Research Using Artificial Intelligence“, in: *Entrepreneurship Theory and Practice* 46 (4), S. 803–832. DOI: <https://doi.org/10.1177/1042258720927369>.

Lindebaum, Dirk, Christian Moser und Gazi Islam (2024): „Big Data, Proxies, Algorithmic Decision-Making and the Future of Management Theory“, in: *Journal of Management Studies* 61 (6), S. 2724–2747. DOI: <https://doi.org/10.1111/joms.13032>.

Luhmann, Niklas (2004 [1995]): *Die Realität der Massenmedien*. 3. Aufl. Wiesbaden.

Luri, Ivan, Hope J. Schau und Bimal Ghosh (2023): „Metaphor-Enabled Marketplace Sentiment Analysis“, in: *Journal of Marketing Research* 61 (3), S. 496–516. DOI: <https://doi.org/10.1177/00222437231191526>.

Obschonka, Martin et al. (2020): „Big data methods, social media, and the psychology of entrepreneurial regions: capturing cross-county personality traits and their impact on entrepreneurship in the USA“, in: *Small Business Economics* 55 (3), S. 567–588. DOI: <https://doi.org/10.1007/s11187-019-00204-2>.

Obschonka, Martin und David B. Audretsch (2020): „Artificial intelligence and big data in entrepreneurship: a new era has begun“, in: *Small Business Economics* 55, S. 529–539. DOI: <https://doi.org/10.1007/s11187-019-00202-4>.

Obschonka, Martin, Christian Fisch und Ryan Boyd (2017): „Using digital footprints in entrepreneurship research: A Twitter-based personality analysis of superstar entrepreneurs and managers“, in: *Journal of Business Venturing Insights* 8, S. 13–23. DOI: <https://doi.org/10.1016/j.jbvi.2017.05.005>.

Pahnke, André und Friederike Welter (2019): „The German Mittelstand: Antithesis to Silicon Valley entrepreneurship?“, in: *Small Business Economics* 52 (2), S. 345–358. DOI: <https://doi.org/10.1007/s11187-018-0095-4>.

Parhankangas, Annaleena und Maija Renko (2017): „Linguistic style and crowdfunding success among social and commercial entrepreneurs“, in: *Journal of Business Venturing* 32 (2), S. 215–236. DOI: <https://doi.org/10.1016/j.jbusvent.2016.11.001>.

Roundy, Philip T. und Adrian Asllani (2019): „Understanding the language of entrepreneurship“, in: *Journal of Economic and Administrative Sciences* 35 (2), S. 113–127. DOI: <https://doi.org/10.1108/JEAS-08-2017-0084>.

Rudin, Cynthia (2019): „Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead“, in: *Nature Machine Intelligence* 1 (5), S. 206–215. DOI: <https://doi.org/10.1038/s42256-019-0048-x>.

Sharkey, Amanda, Balázs Kovács und Greta Hsu (2022): „Expert Critics, Rankings, and Review Aggregators: The Changing Nature of Intermediation and the Rise of Markets with Multiple Intermediaries“, in: *Academy of Management Annals*, Article annals.2021.0025. Advance online publication. DOI: <https://doi.org/10.5465/annals.2021.0025>.

Spitzmüller, Jürgen (2022): „Jargon“, in: *Diskursmonitor. Glossar zur strategischen Kommunikation in öffentlichen Diskursen*. Hg. von der Forschungsgruppe Diskursmonitor und Diskursintervention. Veröffentlicht am 24.03.2022. URL: <https://diskursmonitor.de/glossar/jargon>.

VanderPlas, Jake (2017): *Python Data Science Handbook. Essential Tools for Working with Data*. Beijing u. a.

Vogel, Friedemann (2024): „(Un)Gewissheit im Rechtsdiskurs. Sprachliche Indikatoren für Faktizitätsansprüche in Gerichtsentscheidungen und juristischen Fachaufsätzen aus diskursmetrischer Perspektive“, in: Katja Leyhausen-Seibert, Anna Menzel und Friedemann Vogel (Hg.): *Wissen in Recht und Sprache – Viele Stimmen, vage Grenzen*. Berlin, S. 223–246.

Vogel, Friedemann (2022): „Freund- und Feind-Begriffe“, in: *Diskursmonitor. Glossar zur strategischen Kommunikation in öffentlichen Diskursen*. Hg. von der Forschungsgruppe Diskursmonitor und Diskursintervention. Veröffentlicht am 05.12.2022. URL: <https://diskursmonitor.de/glossar/freund-und-feind-begriffe>.

Vogel, Friedemann (2010): „Linguistische Imageanalyse (LIma). Grundlegende Überlegungen und exemplifizierende Studie zum Öffentlichen Image von Türken und Türkei in deutschsprachigen Medien“, in: *Deutsche Sprache* 4, S.345–377.