

Die Identifikation psychologischer Forschungsthemen mit Topic Modeling

Inauguraldissertation zur Erlangung des Doktorgrades (Dr. rer. nat.)
im Fach Psychologie, Fachbereich I an der Universität Trier

Vorgelegt von
André Bittermann

im Februar 2020

Gutachter:
Prof. Dr. Michael Bosnjak
Prof. Dr. Armin Günther

Dissertationsort:

Trier

Danksagung

Die vorliegende Arbeit wäre ohne die Unterstützung vieler Personen nicht entstanden. Daher möchte ich die Gelegenheit nutzen, mich bei ihnen an dieser Stelle zu bedanken.

Mein Dank gilt zunächst meinem Betreuer Prof. Dr. Michael Bosnjak für die Förderung des Dissertationsprojektes von Beginn an. Seine hilfreichen Anmerkungen und gezielten Vorschläge zur richtigen Zeit ermöglichten ein zielführendes Vorankommen. Ich danke auch für die Möglichkeit, im Rahmen der strategischen Ausrichtung des ZPID meinem Interesse an Topic Modeling nachgehen zu können.

Meinem Zweitbetreuer Prof. Dr. Armin Günther danke ich für das spontane Interesse an meiner Arbeit sowie für den einsichtsreichen, weitsichtigen und kritischen Input.

Mein ganz besonderer Dank gilt Dr. Veronika Kuhberg-Lasson. Veronika hat von Anfang an den kreativen Prozess im Rahmen meiner Studienprojekte gefördert, half jederzeit dabei, sehr rohe Ideen zu konkretisieren und gab in allen Phasen äußerst konstruktives und ermutigendes Feedback. Ohne ihre Impulse und ihr offenes Ohr hätte die vorliegende Dissertation nicht so entstehen können. Vielen Dank auch für die immer spontane Bereitschaft, Passagen dieser Dissertation und der publizierten Studien zu lesen und zu kommentieren.

Andreas, Eva und Nina danke ich für ihr direktes Interesse an den Studien und die wunderbare Zusammenarbeit. Andreas danke ich besonders für den intensiven, lehrreichen und unterhaltsamen Austausch zu R sowie für die vielen Diskussionen zu methodischen Details bei Topic Modeling und darüber hinaus. Es ist mir immer eine große Freude!

Meinen Kolleginnen und Kollegen am ZPID danke ich für das freundliche Umfeld, das über interdisziplinären und zwischenmenschlichen Austausch dafür gesorgt hat, dass ich mit viel Freude an dieser Dissertation gearbeitet habe. Für die jederzeit schnelle und ausführliche Unterstützung in allen englischsprachigen Angelegenheiten danke ich Katja. Für ihre Angebote, Teile dieser Dissertation und der Studien kritisch zu lesen, danke ich in chronologischer Reihenfolge Ute, Oliver, Anita, Veronika B. und Martin. Jannik und Andreas K. danke ich für die immer schnelle Hilfe in allen Hard- und Software-Angelegenheiten. Ich danke auch meinem ehemaligen Bereichsleiter Jürgen Wiesenhütter, der Anfang 2017 mit der Frage, ob ich eine Idee für eine Publikation hätte, den Stein ins Rollen gebracht hatte.

Meinen "Hot Topics" Isabel, Anna und Anton danke ich für alles, vor allem für die große Unterstützung in den letzten Phasen dieser Dissertation.

Inhaltsverzeichnis

Danksagung	3
Inhaltsverzeichnis	4
Zusammenfassung	7
Abstract	8
Abbildungsverzeichnis	9
Tabellenverzeichnis	10
Glossar	11
Liste der in dieser kumulativen Dissertation enthaltenen Publikationen	13
1. Zur Relevanz der Identifikation psychologischer Forschungsthemen	14
2. Methoden zur Erfassung und Analyse von Forschungsthemen	18
2.1 Traditionelle Ansätze in der Psychologie	18
2.1.1 Erfahrungsgeleitete Bestimmung	18
2.1.2 Datengeleitete Bestimmung	19
2.2 Topic Modeling	23
2.2.1 Grundprinzip	24
2.2.2 Zentrale Merkmale und ihre Implikationen	29
2.2.3 Latent Dirichlet Allocation	35
2.2.4 Structural Topic Model	43
2.2.5 Themen-Validität und Themen-Reliabilität	50
2.3 Diskussion der automatisierbaren Ansätze und Forschungslücken	52
3. Gegenstand der vorliegenden Dissertation	58
3.1 Forschungsziele und Datengrundlage	58
3.2 Anmerkungen zum methodischen Vorgehen und zu den Schwerpunkten der drei Studien	60
4. Studie 1: “How to Identify Hot Topics in Psychology Using Topic Modeling”	63
4.1 Hintergrund	63
4.2 Fragestellungen	63
4.3 Methode	63
4.3.1 Daten	63
4.3.2 Topic Modeling und Themen-Validität (Schwerpunkte 1 und 2)	64
4.3.3 Trends (Schwerpunkt 3)	65
4.4 Ergebnisse	65
4.5 Diskussion	70

5. Studie 2: “Ist die psychologische Forschung durchlässig für aktuelle gesellschaftliche Themen?”	74
5.1 Hintergrund	74
5.2 Fragestellungen	75
5.3 Methode	75
5.3.1 Daten	75
5.3.2 Aufbereitung der Textdaten	76
5.3.3 Topic Modeling (Schwerpunkte 1 und 2)	76
5.3.4 Themen-Validität und Publikationsspitzen (Schwerpunkte 3 und 4)	78
5.4 Ergebnisse	79
5.5 Diskussion	83
6. Studie 3: “Unterscheiden sich die Forschungsinteressen von Frauen und Männern in der Psychologie?”	87
6.1 Hintergrund	88
6.2 Fragestellungen	88
6.3 Methode	89
6.3.1 Daten	89
6.3.2 Topic Modeling (Schwerpunkte 1 und 2)	89
6.3.3 Themen-Reliabilität (Schwerpunkt 3)	91
6.3.4 Prävalenzdifferenzen	91
6.3.5 Trend-Dyaden (Schwerpunkt 4)	92
6.4 Ergebnisse	92
6.5 Diskussion	96
7. Entwicklung einer nutzerfreundlichen App zur Exploration und Analyse psychologischer Forschungsthemen (“PsychTopics”)	99
7.1 Hintergrund	99
7.2 Ziele	100
7.3 Methode	101
7.3.1 Daten	101
7.3.2 Oberfläche der App	102
7.3.3 Topic Modeling	102
7.4 Ergebnisse	107
7.4.1 Themen	103
7.4.2 Ansichten der App	103
7.5 Diskussion	107
7.5.1 Erste Rückmeldungen und (methodische) Schlussfolgerungen	107
7.5.2 Perspektiven zur Weiterentwicklung	109
8. Allgemeine Diskussion	111
8.1 Wissenschaftlicher Ertrag vor dem Hintergrund der Forschungsziele	119
8.2 Anknüpfungspunkte für forschungsbasierte Infrastruktureinrichtungen in der Psychologie am Beispiel des ZPID	119

8.2.1 Informations- und Recherchedienste: Brennpunktthemen	114
8.2.2 PsychPorta: Literaturempfehlungen	115
8.2.3 PsychAuthors: Autorenprofile und -netzwerke	116
8.3 Einschränkungen und Implikationen für weitere Forschung	116
8.4 Abschließende Bewertung und eine Topic-Modeling-Entscheidungsheuristik	119
Literaturverzeichnis	121
Anhang	135
Anhang A: Originalpublikation von Studie 1 (accepted version)	135
Anhang B: Originalpublikation von Studie 2	147
Anhang C: Originalpublikation von Studie 3 (accepted version)	159
Anhang D: Supplementmaterialien zu Studie 3	180
Ergebnisse der Analyse mit Abstracts	181
Tabelle D1. Themen bei Verwendung von Abstracts als Textquelle.	183
Tabelle D2. Auflistung aller Themen mit der Anzahl erfolgreicher Reproduktionen.	186
PsychArchives-ESM 1: Detaillierte Beschreibung des methodischen Vorgehens.	192
PsychArchives-ESM 2: R-Code der Analysen	201
PsychArchives-ESM 3: Abbildung zu Frauenanteilen an den Promotionen.	202
PsychArchives-ESM 4: Auflistung aller Themen.	204
PsychArchives-ESM 5: Zeitliche Verläufe aller Themen.	215
Anhang E: Originalpublikation zu Abschnitt 7	219
Elektronische Supplementmaterialien (ESM)	222
Eidesstattliche Erklärung	223

Zusammenfassung

Vor dem Hintergrund eines fachlichen und gesellschaftlichen Interesses an psychologischen Forschungsthemen wurde der Einsatz von Topic Modeling zu deren automatisierten und validen Identifikation erprobt und mit einem traditionellen Ansatz verglichen, der auf Klassifikationskategorien von Fachdatenbanken beruht. Kern der Arbeit sind drei Studien mit unterschiedlichen Anwendungsszenarien: (1) Die Ermittlung psychologischer Brennpunktthemen, (2) das Verhältnis zwischen gesellschaftlichen Herausforderungen und Themen der psychologischen Fachliteratur sowie (3) die Forschungsinteressen von Psychologinnen und Psychologen. In jeder dieser Studien wurden sukzessive neue methodische Schwerpunkte gesetzt. Die Erkenntnisse daraus flossen in die Entwicklung einer nutzerfreundlichen App zur Exploration und Analyse von psychologischen Forschungsthemen ein. Im Vergleich mit dem klassifikationsbasierten Ansatz erwies sich Topic Modeling als Methode der Wahl zur automatisierten und differenzierten Identifikation psychologischer Forschungsthemen. Zur externen Validierung der Themeninhalte stellte sich ein Klassifikationssystem für psychologische Fachliteratur jedoch als geeignet heraus. Einschränkungen, Implikationen für weitere Forschung sowie Folgerungen für die Weiterentwicklung digitaler Produkte von forschungsbasierten Infrastruktureinrichtungen in der Psychologie werden diskutiert. Abschließend wird eine Entscheidungsheuristik zur Auswahl geeigneter Topic-Modeling-Varianten im Kontext von Forschungsthemen vorgelegt.

Schlüsselwörter:

Topic Modeling, Brennpunktthemen, gesellschaftliche Themen, Forschungsinteressen, App

Abstract

Against the background of a professional and societal interest in psychological research topics, the use of topic modeling for their automated and valid identification was tested and compared to a traditional approach which is based on a database classification code system. Three studies with different application scenarios are the main focus of the current work: (1) The identification of psychological hot topics, (2) the relationship between social issues and topics of psychological literature, and (3) the research interests of male and female psychologists. In each of these studies, new methodological focuses were successively set. The findings from these studies were incorporated into the development of a user-friendly app for the exploration and analysis of psychological research topics. Compared to the classification-based approach, topic modeling proved to be the method of choice for the automated and differentiated identification of psychological research topics. However, a classification system for psychological literature was found to be suitable for the external validation of topic contents. Limitations, implications for future research as well as for consequences for the design of digital products of research-based infrastructures in psychology are discussed. Finally, a decision heuristic for the selection of suitable variants of topic modeling in the context of research topics is presented.

Keywords:

Topic Modeling, Hot Topics, Social Issues, Research Interests, App

Abbildungsverzeichnis

Abbildung 1. Verteilung der Dokument-Thema-Wahrscheinlichkeiten in der Publikation “Familie und Integration: Bilingualität als ein wichtiger Schlüssel zur Integration von Kindern in Familie und Gesellschaft” (Leyendecker, 2012)	28
Abbildung 2. Grafische Darstellung des LDA-Modells als Plate-Diagramm.	38
Abbildung 3. Einfluss des Dirichlet-Hyperparameters α auf Wahrscheinlichkeitsdichte der multinomialverteilten Dokument-Thema-Wahrscheinlichkeiten θ_d	40
Abbildung 4. Illustration des Einflusses einer Kovariate “Studienmethodik” auf die Wort- Thema-Wahrscheinlichkeiten	45
Abbildung 5. Einfluss einer Kovariate “Studienmethodik” auf die Prävalenzen	46
Abbildung 6. Zwei Beispiele der Wahrscheinlichkeitsdichten zweier logistisch- normalverteilter Prävalenzverteilungen	48
Abbildung 7. Die Zellen in diesem Levelplot entsprechen den mittleren Dokument-Thema- Wahrscheinlichkeiten der Themen getrennt nach Klassifikationskategorien	70
Abbildung 8. Verteilung des optimalen Wertes für k nach 100 Wiederholungen.....	78
Abbildung 9. Zeitlicher Verlauf der Prävalenz des Themas zu Traumatisierung bei Geflüchteten	80
Abbildung 10. Regressionsbasierte erwartete Prävalenzdifferenzen nach Studienmethodik..	81
Abbildung 11. Verteilung der Themen auf die Hauptklassifikationen.....	83
Abbildung 12. Positionen aller Themen auf den Dimensionen „Prävalenzdifferenz” und „Konvergenz”	93
Abbildung 13. Prävalenzverlauf und Trendlinien der beiden Themen mit zeitlichen konstanten Trendunterschieden.....	96
Abbildung 14. Boxplot der absoluten Prävalenzdifferenzen zwischen Männern und Frauen aller Themen.....	97
Abbildung 15. Ansicht der populären Themen für das Jahr 2017	104
Abbildung 16. Ansicht der “Hot Topics” im Zeitraum 1980–2017	105
Abbildung 17. Ansicht “Alle Themen”	106
Abbildung 18. Ansicht zum Vergleich von erwarteten und beobachteten Verläufen	107
Abbildung 19. Heuristischer Entscheidungsbaum zur Wahl einer geeigneten Topic- Modeling-Variante bei der Identifikation von Forschungsthemen	121

Tabellenverzeichnis

Tabelle 1. Ausschnitt aus dem Klassifikationsschema der Fachdatenbank PSYINDEX mit seinen drei hierarchischen Ebenen.	20
Tabelle 2. Themen und Wort-Dokument-Wahrscheinlichkeiten eines Aufsatzes	27
Tabelle 3. Dokument-Thema-Wahrscheinlichkeit eines Aufsatzes zur mathematischen Schulleistung von minderjährigen Geflüchteten mit emotionalen Traumata.....	27
Tabelle 4. Term-Dokument-Matrix (TDM) des Beispiels aus Abschnitt 2.2.1	31
Tabelle 5. Zusammenfassung der zentralen Merkmale von Topic Modeling und damit verbundene Implikationen.....	35
Tabelle 6. Entscheidungen, die bei LDA vor der Modellinferenz getroffen werden müssen.	43
Tabelle 7. Zentrale Unterscheidungsmerkmale automatisierbarer Ansätze zur Identifikation von psychologischen Forschungsthemen.	53
Tabelle 8. Übersicht der inhaltlichen und methodischen Schwerpunkte der drei Studien.....	62
Tabelle 9. Die Top 10 der Brennpunktthemen im Topic Model.	66
Tabelle 10. Die Top 10 der Brennpunktkategorien bei Verwendung des Klassifikationssystems (in höchstmöglicher Auflösung).....	68
Tabelle 11. Themen mit höchster Wahrscheinlichkeit, von Frauen behandelt zu werden.....	94
Tabelle 12. Themen mit höchster Wahrscheinlichkeit, von Männern behandelt zu werden. ..	94
Tabelle 13. Themen mit den geringsten Unterschieden zwischen Frauen und Männern.	95
Tabelle D1. Themen bei Verwendung von Abstracts als Textquelle	169
Tabelle D2. Auflistung aller Themen mit der Anzahl erfolgreicher Reproduktionen.....	170

Glossar

Dokument-Thema-Wahrscheinlichkeit	Die Wahrscheinlichkeit, dass Dokument d das Thema k behandelt.
Fachdatenbank	Eine Datenbank von Nachweisen wissenschaftlicher Fachliteratur.
Forschungsthema	Ein Gegenstand wissenschaftlicher Untersuchungen, welcher durch eine substanzielle Anzahl inhaltlich kongruenter Fachpublikationen charakterisiert ist.
Freitext	Alle Texte, deren Wörter nicht vollständig in einem zugehörigen Thesaurus definiert sind.
Held-Out-Likelihood	Ein Maß das beschreibt, wie gut das Modell auf neue Textdaten generalisiert.
Hyperparameter	In der Bayes-Statistik ein Parameter der A-priori-Verteilung. Im Kontext von maschinellem Lernen bedeutet dies, dass sein Wert vor der Modellinferenz festgelegt wird.
Klassifikationssystem	Ein System von inhaltlichen Kategorien, denen Publikationen zugeordnet werden.
Kookkurenz	Das gemeinsame Auftreten zweier Wörter.
Korpus	Eine Sammlung von Texten (das Textkorpus, Neutrum).
Modellinferenz	Bei Topic Modeling die algorithmenbasierte Berechnung der Wort- bzw. Dokument-Thema-Wahrscheinlichkeiten entsprechend eines spezifizierten Modells.
Modellwahl	Die Auswahl des Modells, welches die thematische Struktur des Korpus am besten beschreibt. Die Anzahl der Themen ist der wichtigste Modellparameter, der von der analysierenden Person gewählt werden muss.
Prävalenz (bzgl. Themen)	Die durchschnittliche Wahrscheinlichkeit eines Themas im Korpus. Sie wird über die mittlere Dokument-Thema-Wahrscheinlichkeit berechnet.
Semantische Kohärenz	Ein Maß das beschreibt, ob die wahrscheinlichsten Wörter im Thema auch zusammen in den Dokumenten auftreten.
Stemming	Das Kürzen von Wörtern auf den Wortstamm.

Stoppwörter	Wörter, die wenig inhaltliche Relevanz haben (wie z. B. “die”, “ist”, “und”) und in der Regel bei der Vorverarbeitung der Texte entfernt werden.
Text Mining	Das algorithmenbasierte Extrahieren von Information bzw. Mustern aus Textdaten.
Thesaurus	Ein kontrolliertes Vokabular mit Beziehungen zwischen den Begriffen und Erläuterungen bzw. Definitionen zu den Begriffen.
Vokabular	Alle vorhandenen, unterschiedlichen Wörter in einem Korpus.
Wort-Thema-Wahrscheinlichkeit	Die Wahrscheinlichkeit, dass Wort w für Thema k relevant ist.

Liste der in dieser kumulativen Dissertation enthaltenen Publikationen

Studie 1:

Bittermann, A. & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, 226(1), 3–13. <https://doi.org/10.1027/2151-2604/a000318>

Studie 2:

Bittermann, A. & Klos, E. M. (2019). Ist die psychologische Forschung durchlässig für aktuelle gesellschaftliche Themen? Eine szientometrische Analyse am Beispiel Flucht und Migration mithilfe von Topic Modeling. *Psychologische Rundschau*, 70(4), 239–249. <https://doi.org/10.1026/0033-3042/a000426>

Studie 3:

Bittermann, A., Greiner, N. & Fischer, A. (2020). Unterscheiden sich die Forschungsinteressen von Frauen und Männern in der Psychologie? Eine Analyse von PSYINDEX-Einträgen über einen Zeitraum von 50 Jahren. *Psychologische Rundschau*, 71(2), 103–110. <https://doi.org/10.1026/0033-3042/a000482>

1. Zur Relevanz der Identifikation psychologischer Forschungsthemen

Ergänzend zu akademischen Leistungsindikatoren wie der Qualität und Quantität von Publikationen oder dem Einwerben von Drittmitteln wird im Vereinigten Königreich gefordert, dass öffentlich finanzierte Wissenschaft Ergebnisse mit gesellschaftlichem Nutzen hervorbringt. In der Implementierung der sogenannten “impact agenda” wird explizit ein Ertrag der Forschung unter anderem für Gesellschaft, Wirtschaft und Kultur (eben außerhalb der Wissenschaft) gefordert, damit finanzielle Mittel bereitgestellt werden (Jones, Manville & Chataway, 2017). Daher werden wissenschaftliche Einrichtungen seit 2014 im Rahmen des “Research Excellence Frameworks”¹ auch hinsichtlich ihres *Impacts*, also ihres Ertrags und ihres Nutzens, bewertet. Bei der letzten Evaluierung 2014 haben Einrichtungen höherer Bildung (Higher Education Institutions) Impact-Fallstudien eingereicht, mit denen der Impact ihrer Forschung belegt werden sollte. Um die Impact-Bereiche britischer Forschung zu erfassen, wurden die insgesamt 6679 Fallstudien mit automatisierten Textanalyseverfahren analysiert (King’s College London and Digital Science, 2015). Dabei wurden 60 sogenannte “impact topics” identifiziert, die von psychischer Gesundheit über Krebserkrankungen bis Softwareentwicklung reichen und als Transferbereiche wissenschaftlicher Erkenntnisse bezeichnet werden können.

Auch in Deutschland sieht der Wissenschaftsrat (2016, S. 5) den “Transfer von Wissen in Gesellschaft, Kultur, Wirtschaft und Politik” als eine “Kernaufgabe und mit Forschung, Lehre und wissenschaftlichen Infrastrukturanangeboten eine der wesentlichen Leistungsdimensionen wissenschaftlicher Einrichtungen”. In diesem Zusammenhang wurde die Psychologie vom Wissenschaftsrat (2018, S. 82–84) an ihre “natürliche Verantwortung” gegenüber der Gesellschaft erinnert, verbunden mit der Forderung einer “Psychologie des Transfers” und einer “stärkeren Öffnung gegenüber der Gesellschaft”. Angesichts der aktuellen “Großen [sic] gesellschaftlichen Herausforderungen” (Wissenschaftsrat, 2015) wurde betont, dass unter anderem Klimawandel, Migration und demografischer Wandel psychologische Erkenntnisse zu ihrer Bewältigung erfordern (Wissenschaftsrat, 2018, S. 83). Fragestellungen zu diesen Bereichen könnten etwa sein: Wie lässt sich nachhaltiges Umweltverhalten fördern? Welche besonderen Fachkompetenzen erfordert eine interkulturelle Behandlung emotionaler Traumata? Wie kann die Stressbewältigung älterer

¹ <https://www.ref.ac.uk>

Arbeitskräfte unterstützt werden? Auf solche gezielten Fragen Antworten zu finden ist eine Aufgabe adäquater Literaturrecherche. Doch wie können potenzielle Transferbereiche und darin enthaltene Erkenntnisse identifiziert werden, die bislang kaum bekannt waren oder übersehen wurden? Wie lässt sich also die Bandbreite psychologischer Beiträge aufdecken?

Hierzu kommt der metawissenschaftlichen Betrachtung der Psychologie sozusagen aus der “Vogelperspektive” (Ioannidis, Fanelli, Dunne & Goodman, 2015), also mit Methoden der Wissenschaftsforschung und Szientometrie, eine entscheidende Rolle zu. Durch eine Analyse der Themen wissenschaftlicher Fachliteratur kann unter anderem festgestellt werden, was aktuell stark oder bisher kaum beforscht wird, wo potenzielle Forschungslücken und -desiderate vorliegen oder welche historische Entwicklung hinsichtlich der Erforschung eines Themas ausgemacht werden kann. Somit ist die Identifikation von Themen und Trends auch für Forschung und Lehre von Interesse. Weiterhin kann untersucht werden, mit welchem Schwerpunkt ein Thema behandelt wird. So ist beispielsweise angesichts von traumatischen Erfahrungen und Belastungen Geflüchteter in erster Linie eine klinisch-psychologische Perspektive auf Flucht und Migration zu erwarten, jedoch kann die Psychologie mit ihren Disziplinen wie der Sozialpsychologie (Fremdenfeindlichkeit, Vorurteile) oder der Pädagogischen Psychologie (Unterricht mit und Schulerfolg von Migrant*innen) potenziell sehr vielfältig beitragen. Dies empirisch nachzuweisen erfordert inhaltliche Auswertungen entsprechender Fachpublikationen. Um diese Vielfalt unvoreingenommen auszuloten, kommen explorative Verfahren zur Themenidentifikation in Betracht, da sie im Gegensatz zu gezielten Suchen nach vorab definierten Themen datengeleitet und unabhängig vom Vorwissen der analysierenden Person die Inhalte der Forschungslandschaft aufdecken können.

Angesichts der rasant wachsenden Möglichkeiten, Informationen zu speichern und auszutauschen (Hilbert & López, 2011) und einem auch damit einhergehenden Zuwachs an wissenschaftlichen Publikationen in der Psychologie seit Mitte des 20. Jahrhunderts (Krampen, 2015) sind neue Methoden zur Analyse solcher Datenmengen mit hohem Informationsgehalt nötig. Die Anwendung traditioneller Inhaltsanalysetechniken und qualitativer Methoden ist in Anbetracht von mehr als einer Dreihunderttausend psychologischer Fachpublikationen allein aus dem deutschen Sprachraum² unökonomisch bis unmöglich. Statistische Methoden für “Big Data” – also Daten, die nach Laney (2001) groß im Sinne von

² Auf Anfrage umfasst PSYINDEX 357000 Nachweise im Januar 2020.

Umfang, Produktionsgeschwindigkeit und Heterogenität sind – bieten für moderne szientometrische Untersuchungen vielversprechende Möglichkeiten. Speziell zur inhaltlichen Erschließung großer Textmengen sind Techniken des *Text Minings* anwendbar (z. B. Wiedemann, 2013). Ziel dieser computergestützten Verfahren ist es, algorithmenbasiert Information bzw. Muster aus Textdaten zu extrahieren. Text Mining kann vor allem für explorative Untersuchungen eingesetzt werden, um automatisiert Informationen zu gewinnen, von denen man vorher noch nicht wusste, dass sie enthalten sind (Feldman & Dagan, 1995). Aber auch eine hypothesenprüfende Untersuchung kausaler Zusammenhänge ist möglich (Egami, Fong, Grimmer, Roberts & Stewart, 2018).

Die vorliegende Dissertation prüft daher die Anwendung von *Topic Modeling* – einer automatisierten Inhaltsanalysemethode aus dem maschinellen Lernen – zur Identifikation psychologischer Forschungsthemen. Dabei werden zwei Varianten von Topic Modeling in drei unterschiedlichen Anwendungsszenarien eingesetzt, um Stärken und Einschränkungen zu eruieren – vor allem im Hinblick auf digitale Informationsprodukte von forschungsbasierten Infrastruktureinrichtungen in der Psychologie. Aufbauend auf der wegweisenden Arbeit von Griffiths und Steyvers (2004) zur Ermittlung von Forschungsthemen in wissenschaftlichen Publikationen wird konkret für die Psychologie geprüft, wie ein automatisiertes Extrahieren von Forschungsthemen zur Beantwortung verschiedener Fragestellungen beitragen kann, die von fachlichem und auch gesellschaftlichem Interesse sind. Somit soll ein Beitrag geleistet werden, potenzielle Transferbereiche aufzudecken sowie die psychologische Forschungslandschaft inhaltlich so zu erschließen, dass sie auch einer breiteren Öffentlichkeit zugänglich gemacht werden kann (vgl. Kleiner, 2019). Hierfür wird eine App entwickelt, die niederschwellig über psychologische Forschungsthemen informiert und auf passende Fachliteratur verweist.

Die Arbeit ist wie folgt strukturiert: Zunächst werden in Abschnitt 2 verschiedene Arten von Ansätzen zur Erfassung und Analyse von Forschungsthemen unterschieden und hinsichtlich ihrer Eignung zur automatisierten Identifikation psychologischer Forschungsthemen verglichen. Dabei werden Forschungslücken herausgearbeitet, die in Abschnitt 3 für die vorliegende Dissertation konkretisiert werden. Dabei wird auch das übergeordnete methodische Vorgehen dargelegt. Die Abschnitte 4 bis 6 stellen drei empirische Studien vor, wobei über die Originalmanuskripte (Anhänge 1–3) insofern hinausgegangen wird, als dass auf die methodischen Schwerpunkte besonders eingegangen wird. Abschnitt 7 widmet sich

einer Forschungsthemen-App, die aus den Erkenntnissen der drei Studien heraus entwickelt wurde. In Abschnitt 8 erfolgt eine allgemeine Diskussion, in der zunächst der wissenschaftliche Ertrag der Arbeit resümiert wird. Anschließend werden an drei Beispielen Anwendungsmöglichkeiten von Topic Modeling für digitale Produkte von Infrastruktureinrichtungen in der Psychologie aufgezeigt. Im Rahmen einer kritischen Betrachtung von Einschränkungen werden entsprechende Implikationen für weitere Forschung aufgezeigt. Schließlich wird ein Fazit in Form einer zusammenfassenden Bewertung gezogen. Dabei wird eine Entscheidungshilfe für Topic Modeling zur Identifikation von Forschungsthemen vorgelegt.

2. Methoden zur Erfassung und Analyse von Forschungsthemen

Bevor verschiedene Methoden zur Erfassung von Forschungsthemen vorgestellt werden, soll zunächst die Bedeutung des Wortes “Forschungsthema” definiert werden. Ein Forschungsthema wird im Rahmen dieser Arbeit als Gegenstand wissenschaftlicher Untersuchungen (Dudenredaktion, o. J.) verstanden, welcher durch eine substantielle Anzahl inhaltlich kongruenter Fachpublikationen charakterisiert ist. Das bedeutet, dass ein Forschungsthema eine abstrahierte, möglichst semantisch kohärente Zusammenfassung von Inhalten zusammengehöriger Fachliteratur ist. Inhaltlich ähnliche Publikationen adressieren somit das gleiche Thema. Der Kürze wegen werden im inhaltlichen Zusammenhang dieser Dissertation die Begriffe “Forschungsthema” und “Thema” synonym verwendet.

2.1 Traditionelle Ansätze in der Psychologie

Um festzustellen, womit sich die psychologische Fachliteratur befasst, was vergangene oder auch besonders aktuelle Themen sind, existieren unterschiedliche Ansätze, die grob in die zwei Arten “erfahrungsgeleitet” und “datengeleitet” eingeteilt werden können. Damit sind keine völlig ausschließlichen Ansätze gemeint (auch eine erfahrungsgeleitete Bestimmung kann sich auf Daten berufen), vielmehr soll der jeweilige Schwerpunkt verdeutlicht werden. Beide Arten werden im Folgenden anhand von Beispielen beschrieben, wobei der Fokus auf Anwendungen in der Psychologie liegt.

2.1.1 Erfahrungsgeleitete Bestimmung

Bei diesem Ansatz bestimmen Personen anhand ihrer Expertise und/oder ihres fachlichen Überblicks, was als ein (aktuelles) Thema bezeichnet werden kann. So wurde etwa die Auswahl der “Hot Topics” für den 51. Kongress der Deutschen Gesellschaft für Psychologie (DGPs) 2018 durch eine Befragung der Fachgruppen getroffen (DGPs, 2018). Beispiele dieser Themen sind “Open Science”, “Fake News”, “Big Data” oder “Radikalisierung”. Zur Erläuterung jedes Themas wurde von den entsprechenden Expertinnen und Experten ein erläuternder Text formuliert, der Inhalt und Relevanz beschreibt. Als ein anderes Beispiel kann die Zusammenstellung des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID) zu aktuellen Brennpunktthemen³ genannt werden, bei der

³ Das Angebot ist im Januar 2020 noch verfügbar, wird aber nicht mehr aktualisiert.

vordefinierte Suchanfragen in der psychologischen Fachdatenbank PSYNDEX zu denjenigen Themen angeboten werden, die “besonders nachgefragt sind” (ZPID, 2020)⁴. Beispiele solcher Themen sind “Neue Kommunikationsmedien” oder “Burnout und beruflicher Stress”. Ziel dieses Angebots ist es, ein breiteres Publikum außerhalb psychologischer Fachkreise anzusprechen. Dazu wurden die Themen durch einen manuellen Abgleich von Anfragen in der (mittlerweile eingestellten) Suchmaschine *PsychSpider* (Weichselgartner & Baier, 2007) und aktuellen Medienberichten identifiziert. Dabei wurde subjektiv bestimmt, was als aktuell relevanter Medienbericht zählt.

Eine erfahrungsgeleitete Bestimmung hat den Vorteil, besonders aktuelle und klar umrissene Forschungsthemen definieren zu können. Allerdings ist das Vorgehen bei ihrer Bestimmung nicht völlig transparent. Dies geht zu Lasten mangelnder Objektivität und Nachvollziehbarkeit bei ihrer Bestimmung und der Unmöglichkeit einer Automatisierung. Analog zur evidenzbasierten Medizin, nach der Expertenwissen bzw. Meinungen angesehener Autoritäten auf unteren Ebenen der Evidenzhierarchie angesiedelt sind (Burns, Rohrich & Chung, 2012), ist eine empirische Bestimmung zu bevorzugen, wenn Forschungsthemen möglichst gesichert und objektiv bestimmt werden sollen. So sollte ein Forschungsthema auch nachweislich substanzielle Anteile und/oder Anstiege von Fachpublikationen und/oder schriftlichen Diskussionen aufweisen.

2.1.2 Datengeleitete Bestimmung

Mit “datengeleitet” soll im Folgenden ein Vorgehen charakterisiert werden, welches Forschungsthemen intersubjektiv auf Grundlage empirischer und systematischer Datenerhebungen identifiziert. Ein datengeleiteter Ansatz schließt mit objektiven Methoden von Publikationsinhalten oder -metadaten auf übergeordnete Themen. Als Beispiele hierfür in der Psychologie können Maller (1934) oder die szientometrischen Untersuchungen von Krampen (z. B. Krampen & Perrez, 2015) genannt werden: Forschungsthemen wurden hier auf Grundlage der in psychologischen Fachdatenbanken nachgewiesenen Publikationsmetadaten bestimmt, genauer gesagt, anhand ihrer Klassifikationen (in vielen Datenbanken “Subject Headings” genannt). In den Analysen von Krampen wurden die Klassifikationen entsprechend des *Thesaurus of Psychological Index Terms* (Tuleya, 2007)

⁴ <https://psyndex.de/index.php?wahl=PSYNDEX&uwahl=Brennpunktthemen> (16.01.2020)

der *American Psychological Association* (APA)⁵ verwendet. Die Klassifikationskategorien wurden von der APA für ihre Fachdatenbank *PsycInfo*⁶ entwickelt und dienen der thematischen Sacherschließung der Publikationsnachweise. Tabelle 1 zeigt einen Ausschnitt aus diesem Klassifikationsschema in deutscher Übersetzung (Gerards, Gerards, Kuhberg-Lasson, Singleton & Trierweiler, 2014, S. 10–11). Diese Klassifikationen sind somit eine abstrahierte Form der Inhalte und werden in klassifikationsbasierten szientometrischen Untersuchungen mit Themen gleichgesetzt (z. B. Gnamb, Hanfstingl & Leidenfrost, 2006; König, Fell, Kellnhöfer & Schui, 2015; Krampen & Perrez, 2015). Um Prävalenz und Trends der Themen zu bestimmen, werden die Häufigkeiten der entsprechend klassifizierten Publikationen bestimmt.

Tabelle 1. Ausschnitt aus dem Klassifikationsschema der Fachdatenbank PSYINDEX mit seinen drei hierarchischen Ebenen.

2100 Allgemeines
2140 Geschichte und theoretische Systeme
2200 Psychometrie, Statistik und Methodik
2220 Tests und Testen
2221 Sensorisches und motorisches Testen
2222 Entwicklungstests
...
2240 Statistik und Mathematik
2260 Forschungsmethoden und Versuchsplanung
2300 Allgemeine Psychologie
2320 Wahrnehmung
2323 Visuelle Wahrnehmung
2326 Auditive Wahrnehmung, Sprachwahrnehmung
...

⁵ Eine Übersicht über alle Klassifikations-Codes kann hier eingesehen werden:

<https://www.apa.org/pubs/databases/training/class-codes>

⁶ <https://www.apa.org/pubs/databases/psycinfo/>

Der Einfachheit dieses Ansatzes stehen allerdings verschiedene Nachteile gegenüber: Erstens ist das Klassifikationssystem vorab definiert worden, indem von Mitarbeitenden der Datenbank die thematische Breite der Fachdatenbank zum Entwicklungszeitpunkt dieses Systems abgebildet wurde. Das bedeutet, dass einige der zurzeit aktuellen Themen noch gar nicht berücksichtigt werden konnten und daher durch die Klassifikation nicht adäquat repräsentiert sind. Dies führt in der Konsequenz dazu, dass sie nicht oder schlechter gefunden werden können. Zweitens können die Klassifikationstitel selbst auf unteren Ebenen zu breit und abstrakt sein, um spezifische Themen zu erfassen. So würde beispielsweise “E-Learning”⁷ unter “3530 Curricula, Bildungsprogramme, Lehrmethoden” subsumiert – zusammen mit anderen Lehr- und Lernmethoden. Ein drittes Problem ergibt sich aus der Tatsache, dass einige Publikationen mehrere Themen abdecken, in Fachdatenbanken (speziell PsycInfo der APA) jedoch häufig nur eine Klassifikation erhalten. Selbst bei multipler Zuordnung zu verschiedenen Klassifikationen bleibt nicht klar, zu welchen Anteilen sich eine Publikation mit diesen Themen befasst.

Als ein weiteres klassifikationsbasiertes, datengeleitetes Vorgehen kann die Untersuchung von Cascio und Aguinis (2008) im Bereich der Arbeits- und Organisationspsychologie gezählt werden. Mit einer qualitativen Inhaltsanalyse wurden die Themen hier nicht auf Datenbankebene, sondern für die Fachzeitschriften *Journal of Applied Psychology* und *Personnel Psychology* bestimmt. Dazu wurde auf Grundlage von Lehrbuchinhalten eine Taxonomie von 15 breiteren Kategorien mit 50 spezifischeren Unterkategorien entwickelt, anhand derer die Artikel schließlich kodiert wurden. Die Entwicklung erfolgte in einem iterativen Prozess durch Kodierung von Artikeln und einer daraus abgeleiteten Aktualisierung der Taxonomie. Auf diese Weise wurden die Kategorien aus den Daten gebildet, was hinsichtlich thematischer Detailliertheit und Aktualität als Vorteil gegenüber vorab definierten Fachdatenbank-Klassifikation zu sehen ist. Allerdings ist der zeitliche Aufwand offenkundig: Ausgewertet wurden 5780 Artikel, die zwischen 1963 und 2007 veröffentlicht wurden. Damit ist dieses Vorgehen nicht beliebig skalierbar. Außerdem schließt eine manuelle Kodierung eine automatisierte Identifikation von Forschungsthemen aus.

Eine Bestimmung der Themen auf der detailreicheren Ebene von Publikationstiteln wurde von Webster, Jonason und Schember (2009) für “Hot Topics” der Evolutionspsychologie

⁷ In Studie 1 stellte sich via Topic Modeling “E-Learning” als ein eigenes Thema heraus (vgl. Thema 372 in <https://bit.ly/2nI1uVb>).

vorgenommen. Dabei wurden für verschiedene Zeiträume die 25 häufigsten Wörter in den Titeln als Themen aufgefasst (z. B. “Attractiveness, Sexual, Facial, Sex, Men, Female [...]” als ein Thema zu physischer Attraktivität). Analog ging Arik (2013) bei der Auswertung von Wörtern in Titeln, Abstracts und Schlagwörtern im *Turkish Journal of Psychology* vor. Dieser Ansatz der Zusammenfassung besonders häufiger Wörter bietet damit die Möglichkeit, einen detaillierteren Einblick in das Thema zu erlangen als mit breiten Klassifikationskategorien. Allerdings kann auf diesem Wege nur ein Thema pro Zeitraum gefunden werden. Außerdem sagt die Zusammenfassung der insgesamt häufigsten Wörter nichts darüber aus, ob sie auch so zusammen in einzelnen Dokumenten erscheinen (semantische Kohärenz). Daher können einfache Frequenzanalysen “die Sinndimension von Texten nur sehr beschränkt erschließen” (Schaal, Dumm & Fleuß, 2019, S. 275).

Eine andere datengeleitete Identifikation von Forschungsthemen präsentieren Lindahl, Stenling, Lindwall und Colliander (2015) für den Bereich der Sportpsychologie. Mittels *Singular Value Decomposition* (SVD), dem Kern der *Latent Semantic Analysis* (LSA; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990), wurden 1140 Artikel auf zentrale Konzepte reduziert. Bei SVD werden wie bei einer Hauptkomponentenanalyse (*Principal Component Analysis*, PCA⁸) unkorrelierte Komponenten aus einer Matrix mit allen im Textkorpus⁹ vorhandenen Wörtern in den Zeilen und den Artikeln in den Spalten extrahiert. Aus der somit entstandenen Artikel-Komponenten-Matrix wurden die paarweisen Ähnlichkeiten zwischen den Artikeln ermittelt. Anhand der Ähnlichkeiten wurden die Artikel anschließend mittels Clusteranalyse zusammengefasst, so dass thematische ähnliche Artikel gemeinsamen Clustern zugeordnet werden konnten. Die Themen wurden schließlich nach manueller Sichtung der Artikel in den Clustern abgeleitet. Wampold und White (1985) nutzten ebenfalls Clusterverfahren für die Ermittlung von Themen der Beratungspsychologie, allerdings auf Basis von Zitationen. Studien, die ähnliche andere Arbeiten zitieren, wurden einem gemeinsamen Cluster zugeordnet, anhand dessen schließlich das zugrundeliegende Thema bestimmt wurde. Zu den Vorteilen eines Ansatzes mit Clusteranalyse zählen die Identifikation von Themen aus den aktuellen Daten (im Gegensatz zu Fachdatenbank-Klassifikationen) sowie die Möglichkeit, verschiedene Themen pro Zeitintervall zu ermitteln (im Gegensatz zur Verwendung der insgesamt häufigsten Wörter). Dennoch weist auch diese

⁸ Die PCA kann als Spezialfall der SVD bezeichnet werden (Madsen, Hansen & Winther, 2004). Allerdings ist die PCA bei der Berechnung großer Matrizen sehr ineffizient (Kosinski, Wang, Lakkaraju & Leskovec, 2016).

⁹ Entsprechend der Text-Mining-Terminologie wird in dieser Arbeit eine Sammlung von Texten als Korpus (Neutrum) bezeichnet.

Methode einen entscheidenden Nachteil auf: Ein Artikel kann mit den verwendeten “harten” Clusterverfahren¹⁰ nur zu einem Cluster gehören. Allerdings ist es sinnvoll anzunehmen, dass ein Dokument mehrere Themen adressieren kann (Maier et al., 2018), sonst wären beispielsweise Publikationen, die neuropsychologische Aspekte von psychischen Störungen behandeln, entweder der Neuropsychologie oder den psychischen Störungen zugeordnet, was für eine szientometrische Abbildung von Forschungsthemen ein unzureichendes Resultat darstellt.

Die Nachteile der in diesem Abschnitt beschriebenen Ansätze können wie folgt zusammengefasst werden (siehe auch Tabelle 7 in Abschnitt 2.3): Mit klassifikationsbasierten Ansätzen können nur diejenigen Themen gefunden werden, zu denen eine Kategorie existiert. Eine reine Auszählung der häufigsten Wörter (mit entsprechend nur einem Thema pro Zeitraum und unklarer Zusammengehörigkeit auf Dokumentenebene) ist vor allem bei thematisch heterogenen Korpora von Nachteil. Hier ist es wichtig, verschiedene und semantisch kohärente Themen identifizieren zu können. Außerdem sollte die Möglichkeit bestehen, dass Dokumente mehrere Themen beinhalten können.

2.2 Topic Modeling

Um den Nachteilen der oben beschriebenen datengeleiteten Ansätzen zu begegnen, muss ein Verfahren automatisiert verschiedene Themen aus den Daten extrahieren können. Dabei soll berücksichtigt werden, dass ein Dokument mehrere Themen adressieren kann. Hierfür kommen unüberwachte (“unsupervised”) Verfahren des Text Minings in Betracht, also Algorithmen, die Themen ohne menschliche Annotation oder Kodierung der Texte inferieren. Genau darauf ist *Topic Modeling* ausgerichtet, ein Verfahren aus dem maschinellen Lernen und der Computerlinguistik¹¹ mit dem Ziel, Gruppen zusammengehöriger Wörter zu bilden. Diese Gruppen werden als Themen (“topics”) bezeichnet. Dabei werden vom Algorithmus solche Wörter als zusammengehörig gesehen, die dazu neigen, gemeinsam in einem Textdokument aufzutreten. Beispielsweise kann angenommen werden, dass in verschiedenen Publikationen zur Arbeitszufriedenheit Wörter wie “Arbeit”, “Zufriedenheit”, “Unternehmen”, “Angestellte”, “Mitarbeitende” verwendet werden. Wenn diese Wörter in

¹⁰ Bei sogenannten “fuzzy” oder “soft” Clusterverfahren (Bezdek, 1981; Dunn, 1973) können Elemente auch zu mehreren Clustern gehören. Auf deren Nachteile bei der Analyse von Textdaten wird in Abschnitt 2.3 eingegangen.

¹¹ Häufig wird auch in deutschsprachiger Literatur der englische Begriff “Natural Language Processing” verwendet.

vielen Publikationen gemeinsam auftreten, werden sie vom Algorithmus als eine Gruppe erachtet, aus denen die analysierende Person das zugrundeliegende Thema ableiten kann (in diesem Fall eben “Arbeitszufriedenheit”).

Im Folgenden wird zunächst das Grundprinzip von Topic Modeling allgemein beschrieben, bevor zentrale Merkmale dieser Methode aufgeführt werden. Danach werden zwei Varianten von Topic Models¹² vorgestellt, die im Rahmen dieser Dissertation eingesetzt wurden (zur Begründung dieser Auswahl siehe Abschnitt 3.2). Abschließend wird auf die für den methodischen Aufbau der vorliegenden Dissertation wichtigen Begriffe der Validität und Reliabilität im Kontext von Topic Modeling eingegangen.

2.2.1 Grundprinzip

Ausgangspunkt von Topic Modeling ist die Annahme, dass Textdokumente verschiedene Themen zu unterschiedlichen Anteilen behandeln können. Dabei wird angenommen, dass die *manifesten* Wörter des Textes aufgrund einer *latenten* Verteilung von Themen gewählt wurden (Blei, 2012, S. 78). Der Entstehungsprozess eines Textes kann daher wie folgt illustriert werden: Eine Autorin oder ein Autor überlegt zunächst die Themen, von denen der Text handeln soll. Nachdem entschieden wurde, mit welchen Anteilen die Themen im Text vertreten sein sollen, wird der Text geschrieben. Dabei neigen thematisch zusammengehörige Wörter dazu, auch miteinander im Text vorzukommen. Damit einhergehend haben bestimmte Gruppen von Wörtern eine höhere Wahrscheinlichkeit zu Thema X zu gehören als zu Thema Y.

Beispielsweise soll ein Aufsatz über die mathematische Schulleistung von minderjährigen Geflüchteten mit emotionalen Traumata geschrieben werden. Der Text behandelt zwei Themen: Die mathematische Schulleistung (Thema 1) und emotionale Traumata bei Geflüchteten (Thema 2). Der Schwerpunkt soll auf der Schulleistung liegen, das heißt, Thema 1 soll mehr Raum im Text einnehmen. Trotzdem wird auch darauf eingegangen, welche Besonderheiten diese spezielle Gruppe von Schülerinnen und Schülern aufweisen. Davon ausgehend wird der Text geschrieben, wobei Wörter wie “Mathematik”, “Rechnen”, “Leistung” oder “Schule” usw. mit höherer Wahrscheinlichkeit gewählt werden, wenn Thema

¹² Für Übersichten zu weiteren Varianten von Topic Models sei auf Jelodar et al. (2019), Mulunda, Wagacha und Muchemi (2018) sowie Sharma, Kumar und Chand (2017) verwiesen.

1 behandelt wird, während Wörter wie “Flucht”, “Trauma” oder “minderjährig” wahrscheinlicher beim Schreiben über Thema 2 verwendet werden.

Beim Lesen des fertigen Aufsatzes sind nur noch die Wörter im Text bekannt (manifest), während die Themen und ihre Verteilung im Text nicht mehr direkt ersichtlich sind (latent). Ziel des Topic-Modeling-Algorithmus ist es daher, von *kookkurierenden* Wörtern (die also häufig miteinander in Texten verwendet werden) auf die zugrunde liegenden Themenverteilungen in den Texten zu schließen und für jedes Thema die jeweilige Verteilung der wahrscheinlichsten Wörter zu finden¹³. Das Resultat von Topic-Model-Analysen sind daher zwei Verteilungen: (1) Wort-Thema-Wahrscheinlichkeiten für die Wörter, die die Themen charakterisieren, und (2) Dokument-Thema-Wahrscheinlichkeiten für die Themenverteilung in den Dokumenten. Wie der Algorithmus auf diese Verteilungen schließt, wird in den Abschnitten 2.2.3 bzw. 2.2.4 genauer beschrieben. Zunächst sollen diese Verteilungen an einem Beispiel illustriert werden.

Ein Korpus soll aus drei Dokumenten bestehen, die der Einfachheit halber nur die zentralen Begriffe beinhalten:

Dok 1: “Mathematik Schule Schüler Rechnen Lernen Leistung”

Dok 2: “Flucht Krieg Trauma Migration Stress Minderjährig”

Dok 3: “Mathematik Rechnen Leistung Schule Lernen Flucht Trauma Minderjährig”

Da menschlichen Leserinnen und Lesern die Bedeutung von Wörtern bekannt ist, lässt sich leicht erkennen: Dokument 1 behandelt nur das Rechnen in der Schule, Dokument 2 nur Flucht und Trauma. Das dritte Dokument ist der oben genannte Aufsatz zur mathematischen Schulleistung minderjähriger Geflüchteter, der zwei verschiedene Themen behandelt. Das Topic Model allerdings kennt keine Wortbedeutungen und sucht nach Themen allein anhand von Wörtern, die gemeinsam in Dokumenten verwendet werden.

¹³ Mohr und Bogdanov (2013, S. 6) bezeichnen dieses Vorgehen daher treffenderweise als “reverse-engineering the intents of the author(s) in producing the corpus”, also eine Nachkonstruktion des Schaffensprozesses aus dem fertigen Text.

Tabelle 2 zeigt die durch Topic Modeling generierten Themen mit den entsprechenden Wort-Thema-Wahrscheinlichkeiten. Jedes Thema ist zunächst eine Liste mit Wahrscheinlichkeiten aller verschiedenen Wörter im Korpus (Vokabular). Allerdings sind diese Wahrscheinlichkeiten für jedes Thema unterschiedlich, sodass die Wörter für jedes Thema nach ihren entsprechenden Wahrscheinlichkeiten sortiert sind. Daher repräsentieren die jeweils ersten Wörter im Thema den Inhalt am besten. In diesem Beispiel ist schnell ersichtlich, dass die ersten sechs Wörter der Themen zur Interpretation herangezogen werden können¹⁴. Die Summen aller Wahrscheinlichkeiten ergeben spaltenweise (als für jedes Thema) 1. Die Werte sind umso höher, je häufiger die Begriffe im Korpus vorkommen (das Wort "Schüler" etwa erscheint nur einmal, während "Mathematik" in zwei Dokumenten enthalten ist). Dabei muss betont werden, dass die Titel oder "Labels" der Themen nicht vom Topic Model erzeugt werden. Sie werden (ggf. im Konsens verschiedener Personen) anhand der wahrscheinlichsten Wörter und nach Inspektion der für die Themen repräsentativsten Dokumente formuliert. Qualitätskriterien für die Erstellung von Thementiteln existieren bislang noch nicht.

In Tabelle 3 sind die Dokument-Thema-Wahrscheinlichkeiten aufgeführt. Hier lässt sich leicht erkennen, dass bei Topic Modeling Dokumente als eine Mischung aus Themen beschrieben werden können: Dokument 3 adressiert Thema 1 mit etwas höherer Wahrscheinlichkeit als Thema 2. Die Spaltenmittelwerte, also die durchschnittlichen Dokument-Thema-Wahrscheinlichkeiten, repräsentieren die *Prävalenz* des Themas. Damit kann abgebildet werden, wie wahrscheinlich es ist, dass ein Thema im Korpus adressiert wird. Thema 1 ist in diesem Beispiel das dominierende Thema, wird also insgesamt mit höherer Wahrscheinlichkeit behandelt (obgleich der Unterschied hier nicht besonders groß ist). Die Zeilensummen in Tabelle 3 ergeben stets 1: Die Anteile der Themen im Dokument können daher in Prozent vom Gesamtumfang des Dokuments interpretiert werden.

¹⁴ In realen Beispielen ist dies meist weniger einfach. In aller Regel werden die fünf bis zehn wahrscheinlichsten Wörter berichtet.

Tabelle 2. Themen und Wort-Dokument-Wahrscheinlichkeiten¹⁵. Jedes Thema hat eine individuelle Verteilung der Wahrscheinlichkeiten.

Thema 1: Mathematische Schulleistung		Thema 2: Flucht und Trauma	
Mathematik	0.18	Flucht	0.22
Schule	0.18	Trauma	0.22
Rechnen	0.18	Minderjährig	0.22
Lernen	0.18	Krieg	0.11
Leistung	0.18	Migration	0.11
Schüler	0.09	Stress	0.11
Flucht	0	Mathematik	0
Krieg	0	Schule	0
Trauma	0	Schüler	0
Migration	0	Rechnen	0
Stress	0	Lernen	0
Minderjährig	0	Leistung	0

Anmerkung: Die Thementitel sind selbst gewählt; sie werden nicht vom Algorithmus erstellt.

Tabelle 3. Dokument-Thema-Wahrscheinlichkeit eines Aufsatzes zur mathematischen Schulleistung von minderjährigen Geflüchteten mit emotionalen Traumata. Jedes Dokument hat eine individuelle Verteilung der Wahrscheinlichkeiten.

Dokument	Thema 1: Mathematische Schulleistung	Thema 2: Flucht und Trauma
Dok 1	1	0
Dok 2	0	1
Dok 3	0.63	0.37
Mittelwert (Prävalenz)	0.54	0.46

Anmerkung: Die Prävalenz ist die durchschnittliche Wahrscheinlichkeit eines Themas im Korpus.

¹⁵ Der R-Code dieses Beispiels lautet: `topicmodels::LDA(dtm, k = 2, control = list(alpha = 0.01, delta = 0.01, seed = 3), method = "Gibbs")`

Die Verteilung der Dokument-Thema-Wahrscheinlichkeit eines realen Beispiels zeigt Abbildung 1. Hierbei handelt es sich um den Beitrag von Leyendecker (2012) mit dem Titel “Familie und Integration: Bilingualität als ein wichtiger Schlüssel zur Integration von Kindern in Familie und Gesellschaft”. Es ist leicht ersichtlich, dass das Topic Model die zwei zentralen Themen identifizieren konnte: Von den 19 Themen im Modell¹⁶ hat Thema 8 zu Sprachentwicklung¹⁷ den höchsten Anteil, gefolgt von Thema 15 zu Integration¹⁸. Sprachentwicklung hat deswegen eine höhere Wahrscheinlichkeit, da mehr Begriffe im Abstract der Publikation für dieses Thema wahrscheinlich waren. Die Daten dieses Beispiels stammen aus Studie 2 dieser Dissertation (Abschnitt 5).

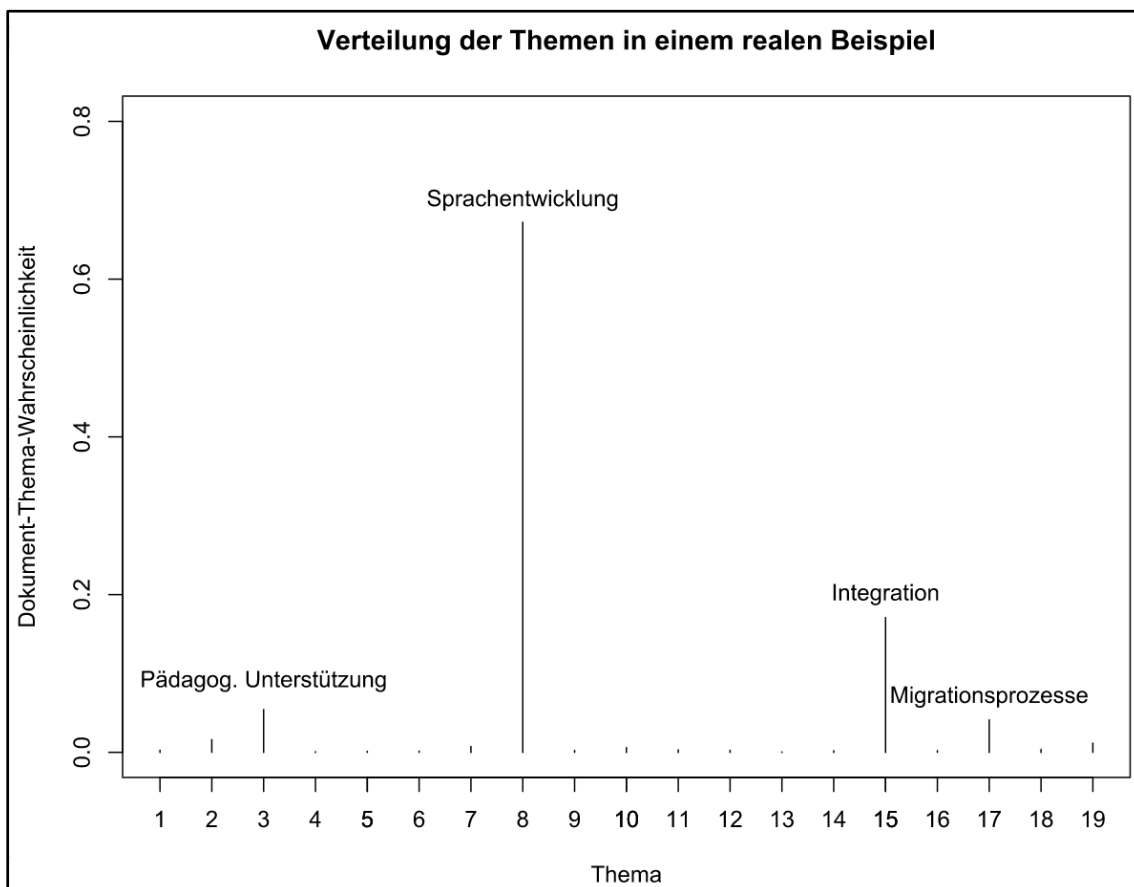


Abbildung 1. Verteilung der Dokument-Thema-Wahrscheinlichkeiten (in einem Topic Model mit 19 Themen) in der Publikation “Familie und Integration: Bilingualität als ein wichtiger Schlüssel zur Integration von Kindern in Familie und Gesellschaft” (Leyendecker, 2012). Die selbst gewählten Labels der vier wahrscheinlichsten Themen sind hinzugefügt.

¹⁶ Die Themen wurden anhand der Abstracts inferiert. Das Abstract des Beispiels kann auf PubPsych eingesehen werden: <https://bit.ly/33w4eVk>

¹⁷ Die fünf wahrscheinlichsten Wörter (mit à posteriori Zusammenfassung von Wörtern des gleichen Wortstamms; Schofield & Mimno, 2016) sind: “Kinder, Eltern, Migrationshintergrund, Deutsche_Sprache, Sprachentwicklung”

¹⁸ “Integration, Gesellschaft, Kultur, Erziehung, Diskussion”

2.2.2 Zentrale Merkmale und ihre Implikationen

Die Grundidee von Topic Modeling ist, wie im letzten Absatz beschrieben, von manifesten, also beobachtbaren Wörtern in Texten, auf die für ihr Zustandekommen ursächlichen latenten Themen zu schließen. Dafür werden zwei Wahrscheinlichkeitsverteilungen geschätzt: (1) Die Wahrscheinlichkeiten der Wörter, zu den Themen zu gehören, sowie (2) die Wahrscheinlichkeiten der Themen, in den Dokumenten adressiert zu werden. Somit ist Topic Modeling ein *probabilistisches Verfahren* (Blei, 2012). Die Themen werden als “latente Faktoren in einem Bayesschen Wahrscheinlichkeitsmodell” modelliert (Waldherr et al., 2019, S. 6). Das bedeutet: Es werden keine Aussagen über Anzahlen und Häufigkeiten von Dokumenten für ein Thema gemacht, sondern Wahrscheinlichkeiten angegeben: Ein Thema wird nicht x -mal adressiert, sondern hat eine Wahrscheinlichkeit von x % adressiert zu werden. Dies erfordert ein Umdenken im Vergleich zu klassifikationsbasierten Ansätzen, die frequentistisch sind und damit leichter interpretierbare Aussagen machen.

Topic Modelle fassen diejenigen Wörter zusammen, die zu *Kookkurenz* neigen, also häufig zusammen innerhalb von Dokumenten verwendet werden (Blei, Ng & Jordan, 2013). Die Idee, kookkurierende Wörter zu thematischen Begriffsclustern¹⁹ zusammenzufassen, gründet in der “distributional hypothesis“ der Linguistik (Harris, 1954). Demnach haben Wörter, die im gleichen Kontext erscheinen, ähnliche Bedeutungen. Durch ein solch relationales Verständnis von Wortbedeutungen (de Saussure, 1959/2011) kann mittels Topic Modeling dem Problem der Polysemie, also dem Vorhandensein mehrerer Bedeutungen eines Wortes, Rechnung getragen werden. Beispielsweise kann mit dem Wort “Läufer” ein Sportler, eine Figur im Schach oder ein Teppich gemeint sein. Im Kontext anderer Begriffe wird die Bedeutung sofort ersichtlich: “Marathon, Läufer, Rekord, Kipchoge”. Dieser Ansatz lässt sich in Zusammenhang mit der “Gebrauchstheorie der Bedeutung”²⁰ des Sprachphilosophen Ludwig Wittgenstein (1953/2003) sehen (z. B. Navarro-Colorado, 2018; Schmiedel, Müller & vom Brocke, 2019).

Aus dem Kookkurenz-Ansatz ergibt sich die Tatsache, dass das Topic Model *keine inhaltliche Hierarchisierung von Themen* vornimmt. Alle Themen werden zunächst gleichberechtigt auch als “Themen” bezeichnet; eine begriffliche Differenzierung im Sinne einer semantischen

¹⁹ Wobei nochmals betont wird, dass Gruppen oder Cluster von Wörtern erst durch die auswertende Person entstehen. Es gibt keine allgemeine Lösung, wieviele der wahrscheinlichsten Wörter des Themas zur Interpretation herangezogen werden.

²⁰ PU § 43: “Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache.”

Hierarchisierung von Themen (z. B. “Angststörungen” als Subthema oder Aspekt des Themas “Psychische Störungen) wird nicht vorgenommen. Solche hierarchischen Beziehungen zwischen den Themen können bei Bedarf mit Clusteranalysen der Wort-Thema-Wahrscheinlichkeiten aufgedeckt werden (Schöch, 2015). Die Ähnlichkeit von Themenpaaren kann außerdem über Korrelation der Wort-Thema-Wahrscheinlichkeitsvektoren untersucht werden, sofern dies für die Fragestellung von Relevanz ist.

Ein weiteres Merkmal charakterisiert Topic Modeling als ein Verfahren zur *Dimensionsreduktion* (Kosinski et al., 2016): Eine große Anzahl von Wörtern wird auf wenige Themen reduziert. Damit Textdaten statistisch ausgewertet werden können, werden sie zunächst in eine numerische Form übertragen. Dazu wird das Vokabular im Korpus bestimmt, also alle vorhandenen, unterschiedlichen Wörter. Das Vokabular des Beispiels in Abschnitt 2.2.1 umfasst die Wörter (in Reihenfolge des Erscheinens):

Vokabular = {Mathematik, Schule, Schüler, Rechnen, Lernen, Leistung, Flucht,
Krieg, Trauma, Migration, Stress, Minderjährig}

Jedes Wort wird anschließend als diskretes Merkmal aufgefasst mit den Häufigkeiten im Dokument als Ausprägung. Jedes Dokument kann somit als Vektor von Worthäufigkeiten repräsentiert werden:

Dok 1 = {1,1,1,1,1,1,0,0,0,0,0,0}

Dok 2 = {0,0,0,0,0,0,1,1,1,1,1,1}

Dok 3 = {1,1,0,1,1,1,1,0,1,0,0,1}

Für Dokument 1 bedeutet dies, dass die ersten sechs Wörter des Vokabulars (“Mathematik” bis “Leistung”) jeweils einmal enthalten sind, während die restlichen sechs Wörter keinmal verwendet wurden. Diese numerische Repräsentation kann auch in einer sogenannten *Dokument-Term-Matrix* (DTM) mit Dokumenten in den Zeilen und den Wörtern dargestellt werden. Tabelle 4 zeigt zwecks besserer Darstellung die transponierte DTM als *Term-Dokument-Matrix* (TDM) an. Die Dimensionalität der Texte ist in diesem Beispiel überschaubar: Jedes Dokument wird auf zwölf Dimensionen abgebildet (die zwölf Wörter bzw. Terms). Anders sieht es in realen Beispielen aus: So umfasst etwa die DTM in Studie 1 dieser Dissertation 314573 (Dokumente) × 6073 (Wörter) Zellen. In jedem Fall wird die

hochdimensionale Datenstruktur durch Topic Modeling auf wenige Themen reduziert²¹ (wobei “wenige” relativ zur Vokabulargröße gesehen werden muss). In Abschnitt 2.2.1 wurde ein Topic Model mit zwei Themen vorgestellt, das heißt, das Korpus wurde von zwölf auf zwei Dimensionen reduziert (vgl. Tabellen 2 und 3).

Tabelle 4. Term-Dokument-Matrix (TDM) des Beispiels aus Abschnitt 2.2.1. Enthalten sind die Häufigkeiten der jeweiligen Wörter in den Dokumenten.

Wort	Dokument 1	Dokument 2	Dokument 3
Mathematik	1	0	1
Schule	1	0	1
Schüler	1	0	0
Rechnen	1	0	1
Lernen	1	0	1
Leistung	1	0	1
Flucht	0	1	1
Krieg	0	1	0
Trauma	0	1	1
Migration	0	1	0
Stress	0	1	0
Minderjährig	0	1	1

Anmerkung: TDM ist die transponierte Dokument-Term-Matrix (DTM).

Wichtig ist zu betonen, dass die optimale Anzahl an Dimensionen (Themen), auf die reduziert wird, nicht vom Topic Model selbst bestimmt wird. Dies ist ein wichtiger Parameter der *Modellwahl* und konkret bedeutet das, dass die Anzahl an Themen, die das Topic Model umfassen soll, von der analysierenden Person festgelegt werden muss – und zwar bevor der Algorithmus die Themen aus den Texten inferiert. Dies entspricht der Analogie von Abschnitt 2.2.1, nach bei der Entstehung eines Textes zunächst überlegt wird, wie viele und welche Themen der Text umfassen soll. Da bei Vorliegen des fertigen Textes diese Überlegungen

²¹ Hier besteht zunächst eine Gemeinsamkeit zum Beispiel mit der Hauptkomponentenanalyse; auf Unterschiede wird in Abschnitt 2.3 eingegangen.

nicht mehr bekannt sind, muss für das gesamte Korpus eine passende Anzahl k an Themen festgelegt werden, deren Inhalte der Algorithmus bestimmen kann. Die richtige Anzahl k an Themen im Korpus zu bestimmen ist nicht trivial und es gibt keine einheitliche Lösung dafür. Generell gilt: Je kleiner k , desto gröber und allgemeiner werden die Themen. Damit gehen potenziell wichtige Details verloren und die Themen können inhaltlich zu heterogen werden. Je größer k ist, desto spezifischer und potenziell ähnlicher werden die Themen, da immer feinere Nuancen von Themen separat aufgeführt werden (z. B. aus dem Thema “Trauma” die Subthemen “Traumabehandlung” und “Traumaentstehung”). Dies kann allerdings zu Themen führen, die nur von einer sehr geringen Zahl an Dokumenten adressiert werden oder die so stichprobenspezifisch sind, dass im Sinne eines “Overfittings” eine Generalisierung erschwert wird. Banks, Woznyj, Wesslen und Ross (2018) empfehlen, Modelle mit bis zu 100 Themen zu testen (z. B. in Zehnerschritten), um sich genaueren Modellkandidaten anzunähern. Daneben bietet sich die Recherche von Topic-Model-Anwendungen mit vergleichbaren Korpora an (z. B. Analysen von wissenschaftlichen Fachpublikationen), um den Bereich möglicher Werte von k festzulegen.

Ein üblicher Ansatz zur Bestimmung des optimalen k ist es, verschiedene Topic Models mit unterschiedlichen Werten für k zu rechnen und die Modellpassung anhand von statistischen Metriken zu evaluieren. Vielmals wird hierfür die Held-Out-Likelihood²² herangezogen (empfohlen bei Kosinski et al., 2016), ein Maß das beschreibt, wie gut das Modell auf neue Textdaten generalisiert. Dazu wird für jedes k das Korpus in einen Trainings- und einen Testdatensatz (Held-Out-Sample) geteilt. Die Held-Out-Likelihood ist umso höher, je besser sich die Dokumente im Testdatensatz mit den Themen des Trainingsdatensatzes beschreiben lassen. Modell der Wahl ist schließlich dasjenige mit der höchsten Held-Out-Likelihood.

Maier et al. (2018) betonen, dass die Modellwahl sowohl anhand statistischer Metriken als auch durch eine qualitative Inspektion der resultierenden Themen erfolgen sollte. Als statistisches Maß empfehlen sie außerdem die semantische Kohärenz der Themen, welche beschreibt, wie oft die wahrscheinlichsten Wörter im Thema auch zusammen in den Dokumenten auftreten (Mimno, Wallach, Talley, Leenders & McCallum, 2011). Hierbei berufen sich Maier et al. (2018) auf die Studie von Chang, Gerrish, Wang, Boyd-Graber und Blei (2009), nach der Modelle mit hoher Held-Out-Likelihood häufig schlechter mit

²² In manchen Studien wird die vergleichbare *Perplexity* als Maß verwendet, die algebraisch äquivalent zum inversen geometrischen Mittel der pro-Wort-Likelihood ist (Blei et al., 2003). Daneben wurde noch eine Reihe anderer Metriken vorgeschlagen, siehe hierzu z. B. Murzintcev (2019).

menschlicher Interpretation übereinstimmen. Die qualitative Prüfung der Themen, also ein Sichten hinsichtlich Interpretierbarkeit, Detailgrad und Relevanz für die Fragestellung, sollte idealerweise im Konsens unterschiedlicher Personen erfolgen.

Tabelle 4 macht noch eines deutlich: Die Reihenfolge der Wörter in den Dokumenten geht in der DTM bzw. TDM verloren. Topic Modelle ignorieren daher die Positionen der Wörter im Text, was auch als *“bag-of-words-model”* bezeichnet wird. Die zu einem Wort zugehörigen Negationen oder Adjektive sind dadurch nicht mehr erkennbar. Dem kann durch sogenannte (automatisch oder manuell erstellte) *N-Grammen* begegnet werden, welche zusammengehörige, aufeinanderfolgende Wörter zusammenführen (so etwa *“kein_Erfolg”* statt *“kein”* und *“Erfolg”*).

Topic Modelle sind *unsupervised*, das heißt, das Modell muss nicht vorab trainiert werden und der Inferenzprozess (das Lernen der Themen aus den Daten) läuft vollständig automatisiert ab. Dies ist durchaus im Sinne der Exploration großer Textkorpora und vor allem dann von Bedeutung, wenn keine Vorkenntnisse über die inhaltliche Struktur des Korpus vorhanden sind. Die Automatisierung macht die Ergebnisse zwar objektiver als qualitative Inhaltsanalysetechniken, die auf manuelle Kodierung beruhen. Allerdings sind die Ergebnisse von Topic Modellen bei großen Textmengen schwer zu validieren (Hecking & Leydesdorff, 2019). Daher kommt der Validität und Reliabilität der Themen besondere Bedeutung zu (siehe Abschnitt 2.2.6). Vor allem die Hauptergebnisse und überraschende Befunde sollten einer genauen Betrachtung unterzogen werden (Schmiedel et al., 2019).

Schließlich ist Topic Modeling ein Verfahren aus dem *maschinellen Lernen*, was bedeutet, dass genügend Daten vorhanden sein müssen, damit der Algorithmus *“lernen”* kann. Die Datenmenge bezieht sich bei Topic Modeling einerseits auf die Anzahl der Dokumente, andererseits auf deren Länge (Wortanzahl). Für gut interpretierbare Themen empfiehlt Nguyen (2015) mindestens 1000 Dokumente im Korpus bei einer Dokumentlänge von 100 Wörtern (für ein Modell mit 100 Themen). Dabei ist allerdings entscheidend, um welche Art von Text es sich handelt. So ist bei wissenschaftlichen Fachartikeln der Anteil der zentralen Inhalte am Gesamtumfang im Abstract höher als im Volltext. Im Falle von standardisierten Schlagworten (wie den PSYNDEX Terms; ZPID, 2016), die per Definitionem ausschließlich die zentralen Inhalte abbilden, führen auch im Schnitt sechs bis sieben Begriffe pro Dokument zu sinnvollen Ergebnissen (vgl. PsychArchives-ESM 1 zu Studie 3 in Anhang D).

Tabelle 5 fasst die zentralen Merkmale und damit verbundene Implikationen für die Anwendung von Topic Modeling zusammen. Konkrete Parameter, welche Entscheidungen beim Anwender erfordern, werden in Tabelle 6 in Abschnitt 2.2.3 aufgeführt. Es kann konstatiert werden, dass nicht die Existenz von Inhalten in Textkorpora “bewiesen” wird. Vielmehr gibt das Modell Wahrscheinlichkeitsschätzungen zu Themen, die sorgfältig interpretiert werden müssen. Auch wenn es ein unüberwachtes Verfahren aus dem maschinellen Lernen ist – viele Entscheidungen sind zu treffen, die potenziell Einfluss auf das Ergebnis und daraus gezogene Schlussfolgerungen haben: Die Anzahl der Themen, Hyperparameter der Dirichlet-Verteilung (siehe Abschnitt 2.2.3), Schritte der Datenvorverarbeitung (*preprocessing*) wie Kürzen auf den Wortstamm (*stemming*) oder Lemmatisierung²³, das Entfernen von Stoppwörtern (*stopwords*; Wörter, die wenig inhaltliche Relevanz haben wie z. B. "die", "ist", "und") oder Häufigkeitsschwellen für das Entfernen besonders häufiger bzw. seltener Wörter im Korpus. Ergänzend zu einer guten Kenntnis des Korpus und der jeweiligen Forschungsfrage können die Best-Practice-Empfehlungen von Banks et al. (2018) sowie von Maier et al. (2018) Orientierung geben. Maier et al. (2018) veranschaulichen die Folgerichtigkeit ihrer Empfehlungen anhand von empirischen Daten aus einem laufenden Forschungsprojekt.

²³ Hierbei werden Flexionsformen eines Wortes auf eine Grundform gebracht, z. B. werden “ist” und “sind” durch “sein” ersetzt.

Tabelle 5. Zusammenfassung der zentralen Merkmale von Topic Modeling und damit verbundene Implikationen.

Merkmal	Implikation
Probabilistisches Verfahren	Ein Thema wird mit einer bestimmten wahrscheinlich adressiert. Es werden keine Aussagen über die Häufigkeit von Dokumenten zu einem Thema gemacht.
Inferenz durch Wort-Kookkurenz	Themen sind Gruppen von Wörtern, die häufig gemeinsam in Texten vorkommen. Damit kann die konkrete Bedeutung von Polysemen ersichtlich werden.
Keine dem Modell inhärente Hierarchie der Themen	Die Themen können inhaltlich ähnlich sein, oder Subthemen bzw. Facetten anderer Themen repräsentieren. Ihre Beziehungen können mit Korrelationen oder Clusterverfahren untersucht werden.
Dimensionsreduktion	Das gesamte Vokabular im Korpus wird auf wenige Themen reduziert, wobei die Anzahl der Themen vorab festgelegt werden muss.
Modellwahl	Für die Bestimmung der für das Korpus optimalen Anzahl an Themen muss eine Reihe von Modellen gerechnet und evaluiert werden.
Bag-of-words-model	Das Topic Model ignoriert die Reihenfolge von Wörtern im Dokument, wodurch zum Beispiel Negationen und Adjektive nicht mehr eindeutig einem Wort zuordenbar sind. Ggf. ist es hilfreich, N-Gramme zu bilden.
Unsupervised model	Die Validität und Reliabilität der inferierten Themen muss beachtet werden.
Machine learning	Das Korpus sollte mindestens 1000 Dokumente umfassen, damit interpretierbare Themen vom Algorithmus "gelernt" werden können.

2.2.3 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) wurde von Blei, Ng und Jordan (2003) vorgestellt. LDA kann als die am meisten verbreitete Topic-Modeling-Variante bezeichnet werden (Mohr & Bogdanov, 2013) und ist, wie im Folgenden gezeigt wird, für Forschungsziele der vorliegenden Arbeit geeignet. Daher soll beschrieben werden, wie bei LDA die Themen aus unstrukturierten Texten identifiziert werden. Für einen Abriss der Entwicklung von LDA, insbesondere als eine Generalisierung von *Probabilistic Latent Semantic Indexing* (Hofmann, 1999), sei auf Blei et al. (2003) und Bikienga (2018) verwiesen.

Der Name des Verfahrens leitet sich zunächst aus der Annahme ab, dass Themen latent sind. Sie können also nicht direkt beobachtet werden, sondern werden aus den ermittelten Wort-Thema-Wahrscheinlichkeiten (den wahrscheinlichsten Wörter eines Themas) geschlossen. Das “Dirichlet” (nach dem Mathematiker Johann Peter Gustav Lejeune Dirichlet, 1805–1859) in LDA verweist auf die Dirichlet-Verteilung, anhand derer die Wort-Thema-Wahrscheinlichkeiten einerseits und die Dokument-Thema-Wahrscheinlichkeiten andererseits bestimmt werden. Die Dirichlet-Verteilung ist eine multivariate Wahrscheinlichkeitsverteilung, deren Hyperparameter²⁴ Einfluss darauf nehmen, ob ein Wort bzw. Dokument für wenige oder viele Themen wahrscheinlich ist (unten dazu mehr). Die Form dieser Verteilung bestimmt, welchem Thema ein Wort zugewiesen wird (“allocation”) bzw. wie viele Themen ein Dokument adressiert.

Im Folgenden wird dargestellt, wie die Themen aus den Textdaten generiert werden (vgl. hierzu Blei, 2012, Blei et al., 2003 sowie Blei & Lafferty, 2009). Dabei soll vor allem auf die Rolle der Dirichlet-Verteilung fokussiert werden, da bei einer Anwendung von LDA die Wahl der Dirichlet-Hyperparameter unmittelbaren Einfluss auf das Ergebnis und seine Interpretation haben kann. Daher sollen die Hintergründe zu den Entscheidungen, welche die analysierende Person zu treffen hat, erläutert werden. Wie in den vorherigen Abschnitten beschrieben ist das Ziel von Topic Modeling, zwei Wahrscheinlichkeitsverteilungen zu ermitteln:

²⁴ Ein Hyperparameter ist in der Bayes-Statistik ein Parameter der A-priori-Verteilung. Im Kontext von maschinellem Lernen bedeutet dies, dass sein Wert vor der Modellinferenz festgelegt wird. Mehr dazu weiter unten im Text.

- (1) Wort-Thema-Wahrscheinlichkeiten (um die Wörter der Themen zu bestimmen; hier mit β bezeichnet²⁵)
- (2) Dokument-Thema-Wahrscheinlichkeiten (um die Themenanteile im Dokument bzw. ihre Prävalenz zu bestimmen; hier θ mit bezeichnet)

Ein Thema k wird als Verteilung β_k über ein festes Vokabular modelliert. Das bedeutet, dass ein Thema eine Verteilung von Wahrscheinlichkeiten aller verschiedenen Wörter im Korpus ist. Dies entspricht Tabelle 2 in Abschnitt 2.2.1. Da ein Dokument mehrere Themen adressieren kann, wird für jedes Dokument d eine Verteilung θ_d modelliert (also die Wahrscheinlichkeiten der Themen für Dokument d). Dies entspricht Tabelle 3 in Abschnitt 2.2.1. Die Wörter bzw. Inhalte der Themen können somit aus β_k , ihre Anteile in den Dokumenten und darüber auch ihre Prävalenz im Korpus aus θ_d abgeleitet werden.

Die beiden Wahrscheinlichkeitsverteilungen β_k und θ_d werden in einem generativen Prozess ermittelt. Topic Modelle werden als generative Modelle für Dokumente bezeichnet, da ein probabilistischer Prozess spezifiziert wird, der die Genese der Dokumente modelliert (Steyvers & Griffiths, 2007). Zugrunde liegt die Annahme, dass erst die Themen, dann die Dokumente entstehen. Abbildung 2 zeigt das LDA-Modell als Plate-Diagramm, welches sich wiederholende Prozesse und Abhängigkeiten der Variablen illustriert. Der im Plate-Diagramm grafisch dargestellte generative Prozess der Textentstehung wird bei LDA wie folgt angenommen (vgl. Blei & Lafferty, 2009):

- I. Für jedes Thema k (aus der Anzahl K aller Themen), wähle zufällig eine Wort-Thema-Verteilung β_k aus einer Dirichlet-Verteilung $\text{Dir}(\eta)$ mit Hyperparameter η .
- II. Für jedes Dokument d (aus der Anzahl D aller Dokumente),
 - A. wähle zufällig eine Dokument-Thema-Verteilung θ_d aus einer Dirichlet-Verteilung $\text{Dir}(\alpha)$ mit Hyperparameter α .
 - B. Für jede Wortposition d,n (aus der Anzahl N aller Wörter in Dokument d)
 1. ordne ein Thema $Z_{d,n}$ aus einer Multinomialverteilung von θ_d zu und
 2. ordne ein für Thema $Z_{d,n}$ spezifisches Wort $W_{d,n}$ aus einer Multinomialverteilung von $\beta_{Z_{d,n}}$ zu.

²⁵ Die verwendeten Symbole variieren sehr stark in der Literatur; in diesem Abschnitt erfolgt die Bezeichnung analog zu Blei & Lafferty (2009).

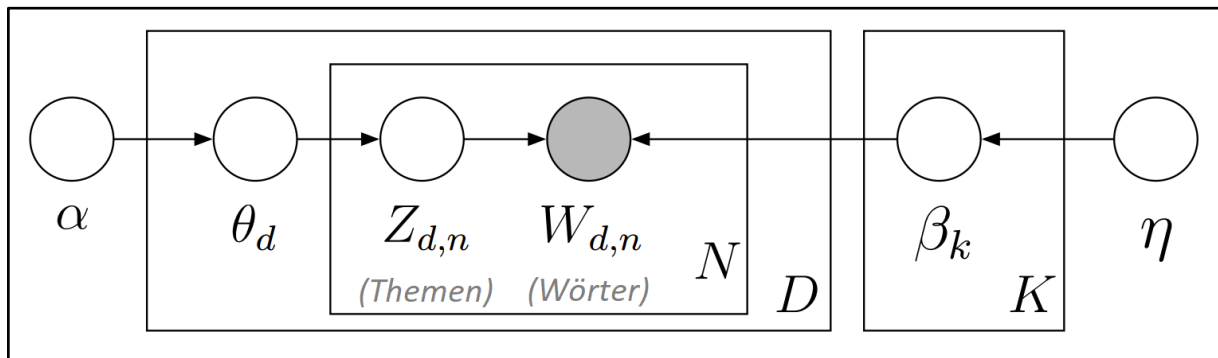


Abbildung 2. Grafische Darstellung des LDA-Modells als Plate-Diagramm (modifiziert nach Blei & Lafferty, 2009, S. 74, Figure 2). Sich wiederholende Prozesse werden in Rechtecken (“plates”) zusammengefasst. Pfeile beschreiben Abhängigkeiten. Alle Variablen sind latent (weiße Kreise), ausgenommen die Wörter $W_{d,n}$ in den Dokumenten (grau). θ_d und β_k sind Wahrscheinlichkeitsverteilungen, α und η sind Dirichlet-Hyperparameter. Weitere Erläuterungen im Text.

Die Rolle der Dirichlet-Verteilung kann mit einer Würfel-Analogie veranschaulicht werden²⁶: Dem generativen Prozess zufolge wird bei LDA angenommen, dass man, um ein Wort $W_{d,n}$ an Position d,n zu schreiben, zufällig ein Thema $Z_{d,n}$ ²⁷ für diese Position auswählt (vgl. oben Schritt II B1). Dann wird zufällig ein Wort aus einer Liste $\beta_{Z_{d,n}}$ derjenigen Wörter ausgewählt, die dieses Thema ausmachen (Schritt II B2). Diese beiden zufälligen Auswahlen können als Würfelwürfe gedacht werden: *Wörterwürfel* für die Verteilungen der Wort-Thema-Wahrscheinlichkeiten β_k (also die Themeninhalte) und *Themenwürfel* für die Verteilungen der Dokument-Thema-Wahrscheinlichkeiten θ_d (also die Themenverteilungen). Die Wörterwürfel (einer für jedes Thema k) haben so viele Seiten wie Wörter im Vokabular. Die Themenwürfel (einer für jedes Dokument d) haben so viele Seiten wie Themen im Modell. Um ein Wort zu schreiben würfelt man zunächst das Thema $Z_{d,n}$ aus, über das man gerade schreiben möchte (mit dem Themenwürfel für θ_d) und anschließend greift man zum entsprechenden Wörterwürfel für dieses Thema und würfelt das Wort aus, das man schreiben möchte (mit dem Wörterwürfel für β_k). Die Würfel repräsentieren somit die Multinomialverteilungen von β_k (vgl. Tabelle 2 in Abschnitt 2.2.1) bzw. θ_d (vgl. Tabelle 3 in Abschnitt 2.2.1). Dabei sind verschiedenste Multinomialverteilungen denkbar, wie etwa Gleichverteilung (bzgl. β_k : alle Wörter sind für ein Thema gleich wahrscheinlich) oder schiefe Verteilungen (bzgl. β_k : nur wenige Wörter sind für ein Thema gleich wahrscheinlich).

²⁶ Zu diesem und dem nachfolgenden Absatz hat dankenswerterweise Andreas Fischer (Nürnberg) beigetragen.

²⁷ Mit $Z_{d,n}$ wird also ein für die Position d,n bestimmtes Thema aus der Liste aller k bezeichnet ($k = Z_{d,n}$).

Die entscheidende Frage lautet nun: Welche dieser Multinomialverteilungen liegen den Wörter- bzw. Themenwürfeln zugrunde? Anders ausgedrückt: Wie wahrscheinlich es ist, einen Inhalts- oder Themenwürfel zu ziehen, auf dem die Seiten gleichwahrscheinlich (Laplace-Würfel) oder unterschiedlich wahrscheinlich (“gezinkt”) sind? Bei der Analyse von Texten sind gezinkte Themenwürfel zu bevorzugen, da nicht alle Dokumente auch stets alle Themen adressieren. Beispielsweise werden bei einer Publikation in einer klinisch-psychologischen Fachzeitschrift nicht alle Themen aus allen Bereichen der Psychologie (wie etwa Personalauswahl oder Unterrichtsgestaltung) behandelt, sondern nur wenige (klinisch-psychologische) Themen. Auch die Wörterwürfel sind gezinkt: Nur wenige Wörter dürfen für ein Thema wahrscheinlich sein, damit die Themen inhaltlich unterscheidbar sind.

Die Dirichlet-Verteilungen bestimmten nun, wie wahrscheinlich es ist, einen gezinkten Wörter- bzw. Themenwürfel zu ziehen. Entscheidenden Einfluss auf die Würfelwahl (gezinkt/Laplace) haben die beiden Dirichlet-Hyperparameter α und η . Mit niedrigem α ist der Themenwürfel gezinkt und wird sehr wahrscheinlich Seiten haben, die unterschiedlich wahrscheinlich sind. Mit $\text{Dir}(\alpha)$ kann somit der Annahme Rechnung getragen werden, dass ein Dokument d verschiedene Themen k in unterschiedlichem Ausmaß $\theta_{d,k}$ adressieren kann, was eine sinnvolle Annahme bei der Analyse von Texten ist. Mit dem Hyperparameter α wird also die Form der Dirichlet-Verteilung $\text{Dir}(\alpha)$ beeinflusst, welche wiederum eine A-priori-Aussage darüber macht, wie die θ_d verteilt sind: Hohe Werte von α bedeuten, dass ein Dokument viele Themen adressieren kann, Werte von $\alpha < 1$ bedeuten, dass ein Dokument nur wenige Themen adressiert. Für die Wort-Thema-Wahrscheinlichkeiten mit A-priori-Dirichlet-Verteilung $\text{Dir}(\eta)$ gilt analog: Hohe Werte von η bedeuten, dass viele Wörter für ein Thema wahrscheinlich sind, Werte von $\eta < 0$ bedeuten, dass nur wenige Wörter für ein Thema wahrscheinlich sind. Letzteres führt zu inhaltlich stärker getrennten Themen.

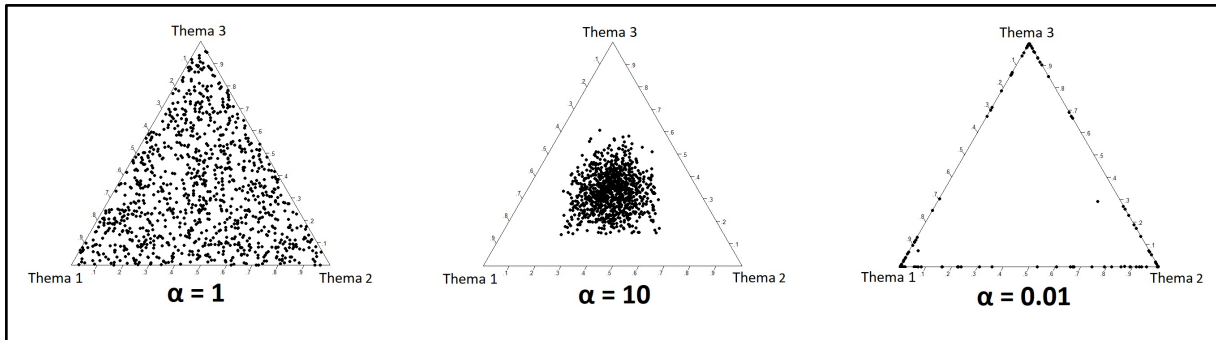


Abbildung 3. Einfluss des Dirichlet-Hyperparameters α auf Wahrscheinlichkeitsdichte der multinomialverteilten Dokument-Thema-Wahrscheinlichkeiten θ_d in 1000 Dokumenten (Punkte) für das Beispiel von drei Themen (Ecken)²⁸. Bei drei Themen repräsentiert ein Dreieck alle möglichen Anteils kombinationen dieser Themen in einem Dokument. Liegt z. B. ein Punkt in der Ecke unten links, hat Thema 1 eine Wahrscheinlichkeit von nahezu 100 %, im Dokument behandelt zu werden, während die anderen beiden Themen eine Wahrscheinlichkeit von 0 % haben. Liegt ein Punkt in der Mitte, sind alle Themen mit ca. 33 % gleich wahrscheinlich im Dokument. Bei $\alpha = 1$ werden keine Wahrscheinlichkeitsverteilungen bevorzugt. Hohe Werte von α begünstigen Gleichverteilungen (Mitte des Dreiecks). Sehr niedrige Werte von α machen Verteilungen wahrscheinlicher, bei denen nur ein bis zwei Themen im Dokument adressiert werden (die Punkte häufen sich in den Ecken bzw. liegen auf den Kanten). Die Dirichlet-Verteilung gibt somit an, wie wahrscheinlich die möglichen Multinomialverteilungen θ_d sind.

Abbildung 3 illustriert den Einfluss von α auf die Verteilungen der Dokument-Thema-Wahrscheinlichkeiten θ_d an einem Beispiel von drei Themen (für die Verteilungen der β_k verhält es sich analog). Jeder Punkt steht für eine Multinomialverteilung θ_d , also der Wahrscheinlichkeiten der drei Themen in Dokument d . Die Kanten des Dreiecks stehen für die Wahrscheinlichkeitsverteilungen zweier Themen (Achsen mit dem Wertebereich von 0 bis 1). Liegt beispielsweise ein Punkt auf der Mitte der unteren Kante des Dreiecks, bedeutet dies, dass das obere Thema mit einer Wahrscheinlichkeit von 0 %, die beiden anderen Themen mit jeweils 50 % adressiert werden. Liegt ein Punkt genau auf einer Ecke, hat dieses Thema eine Wahrscheinlichkeit von 100 %, die anderen jeweils 0 %. Liegt ein Punkt in der Mitte des Dreiecks, sind alle drei Themen gleich wahrscheinlich (33.33 %). In der Würfelanalogie würde letzteres bedeuten, dass die Themenwürfel Laplace-Würfel sind, während an anderen Stellen die Würfel “gezinkt” sind: Manche Seiten/Themen sind wahrscheinlicher als andere. Die Form der Dirichlet-Verteilung nimmt nun Einfluss darauf, wo sich die Punkte häufen, das heißt, welche Verteilungen für θ_d wahrscheinlicher werden. Bei $\alpha = 1$ werden keine Wahrscheinlichkeitsverteilungen bevorzugt: Die Punkte verteilen sich

²⁸ erstellt mit dem R-Paket “DirichletReg” und dem Code (für $\alpha = 1$; analog für andere Werte von α):
`plot(DR_data(rdirichlet(1000, alpha = c(1,1,1))), a2d = list(colored = FALSE, c.grid = FALSE))`

gleichmäßig über das Dreieck (also über alle möglichen Verteilungen). Steigt α , sammeln sich die Punkte in der Mitte. Dadurch sind die meisten Verteilungen der θ_d nahezu gleichverteilt. Bei niedrigen Werten von α sammeln sich die Punkte in den Ecken und auf den Kanten, das heißt, es werden vor allem solche Verteilungen der θ_d bevorzugt, bei denen nur ein bis zwei Themen eine hohe Wahrscheinlichkeit haben, während das dritte Thema eine Wahrscheinlichkeit von 0 % aufweist. Somit kann Einfluss darauf genommen werden, ob die Dokumente im Korpus viele oder nur wenige Themen adressieren sollen. Diese Entscheidung ist am sinnvollsten vor dem Hintergrund der Fragestellung und der konkreten Art der Texte entscheidend: Bei einem Korpus mit Übersichtsarbeiten zur gesamten Psychologie ist es sinnvoll anzunehmen, dass die Artikel mehrere Themen (z. B. Disziplinen der Psychologie) adressieren (höheres α , jedoch < 1), während bei einem Korpus mit Dissertationen zu spezifischen Fragestellungen ein niedrigeres α sinnvoll ist. Anhaltspunkte und Erfahrungswerte für plausible Größenordnungen der Parameter finden sich in der Literatur (z. B. Tang, Meng, Nguyen, Mei & Zhang, 2014).

Die Hyperparameter α und η sind zusammen mit der Anzahl an Themen k vorab festzulegen. Wie werden nun im Rahmen der Modellinferenz die Wort- bzw. Dokument-Thema-Wahrscheinlichkeiten geschätzt, bzw.: “What is the hidden structure that likely generated the observed collection?” (Blei, 2012, S. 79). Zur Ermittlung der konkreten Werte von β_k und θ_d wird der generative Prozess nun umgekehrt: Von den beobachtbaren Wörtern $W_{d,n}$ soll auf β_k und θ_d geschlossen werden. Allerdings können ihre Verteilungen nicht direkt berechnet werden (Griffiths & Steyvers, 2004). Stattdessen werden die Themenzuordnungen eines jeden Wortes geschätzt, anhand derer schließlich β_k und θ_d bestimmt werden. Hierfür existieren verschiedene Arten von lernenden Algorithmen (Blei & Lafferty, 2009), wobei *Gibbs-Sampling*²⁹ (Griffiths & Steyvers, 2004) der gebräuchlichste ist (Wehrheim, 2018). Sein Ablauf kann wie folgt beschrieben werden (vgl. Chen, 2011):

1. Zunächst wird jedem Wort in jedem Dokument zufällig ein Thema zugewiesen (sog. *random initialization*). Dadurch entstehen bereits Verteilungen von β_k und θ_d , wiewgleich noch keine sehr guten.
2. Diese Verteilungen werden für die vorliegenden Daten iterativ optimiert, indem für jedes Dokument d wie folgt vorgegangen wird:
 - a. Für jedes Wort W in Dokument d und für jedes Thema k , berechne:

²⁹ Der Algorithmus ist nach dem Physiker Josiah Willard Gibbs benannt.

- i. den Anteil der Wörter im Dokument d , die momentan dem Thema k zugeordnet sind, und
 - ii. den Anteil der Zuordnungen zu diesem Thema in allen Dokumenten bei diesem Wort W .
 - b. Nun wird davon ausgegangen, dass alle Themenzuordnungen korrekt sind, außer für das aktuelle Wort W . Dessen Zuordnung wird aktualisiert mit der (durch die Dirichlet-Verteilung beeinflussten) Wahrscheinlichkeit, dass Thema k dieses Wort hat entstehen lassen.
3. Je öfter dieser Aktualisierungsprozess wiederholt wird, desto besser werden die Zuordnungen, aus denen schließlich die finalen Verteilungen von β_k und θ_d berechnet werden. Wie oft diese Aktualisierung stattfindet, wird von der analysierenden Person durch den Gibbs-Kontrollparameter "Anzahl der Iterationen" festgelegt.

Für die Anwendung von LDA ergibt sich aus diesem Ablauf folgende Einsicht: Durch die zufällige Zuweisung zu Beginn des Gibbs-Samplings ist der Algorithmus nicht deterministisch, das heißt, verschiedene Ausführungen können zu unterschiedlichen Ergebnissen führen (Maier et al., 2018) mit unter Umständen nur lokalen Optima (Awati, 2015). Dies hat zur Folge, dass der Algorithmus nicht mehr konvergiert (die Verteilungen ändern sich nicht mehr), obwohl die gefundene Lösung nur für die gewählten zufälligen Startwerte passt. Konsequenz daraus ist ein mehrfaches Ausführen des Algorithmus, wobei für jede Ausführung ein *random seed* gesetzt wird. Dabei handelt es sich um einen beliebigen ganzzahligen Wert, welcher den Zufallszahlengenerator der Software reproduzierbar macht. Dies bedeutet – im Gegensatz zu traditionellen frequentistisch-statistischen Verfahren der Psychologie – dass nur durch den random seed die Ergebnisse und Themen reproduziert werden können. Tabelle 6 fasst die für die LDA-Anwendungspraxis wesentlichen Entscheidungen zusammen, die von der analysierenden Person à priori getroffen werden müssen.

Tabelle 6. Entscheidungen, die bei LDA vor der Modellinferenz getroffen werden müssen.

Entscheidung bzgl.	Bedeutung
Anzahl der Themen k	Je kleiner k , desto breiter gefasst sind die Themeninhalte. Je größer k , desto spezifischer sind die Themeninhalte (s. a. Abschnitt 2.2.2).
Dirichlet-Hyperparameter α und η	α nimmt Einfluss darauf, wie viele Themen in einem Dokument adressiert werden ($\alpha < 1$ für wenige Themen). Maier et al. (2018) empfehlen, mehrere Werte von α für verschiedene k zu testen. η nimmt Einfluss darauf, wie viele Wörter für ein Thema wahrscheinlich sind, d. h., wie stark die Themen voneinander abgrenzbar sind ($\eta < 1$ für wenige Wörter).
<i>random seed</i>	Ein beliebiger ganzzahliger Wert, der benötigt wird, um trotz der zufälligen Initialisierung des Algorithmus reproduzierbare Ergebnisse zu erzeugen.
Gibbs-Parameter	Neben der Anzahl der Algorithmus-Iterationen des Gibbs-Samplings (je mehr, desto besser, aber auch desto länger die Dauer) kann unter anderem auch festgelegt werden, wie viele verschiedene Initialisierungen getestet werden. Für jede sollte ein anderer <i>random seed</i> gesetzt werden. Weitere Parameter zum Fine-Tuning erläutert Awati (2015).

2.2.4 Structural Topic Model

Mit dem *Structural Topic Model* (STM) präsentierten Roberts, Stewart, Tingley und Airoidi (2013) eine Erweiterung von LDA, die das Modellieren von Dokument-Metadaten als Kovariaten bzgl. Themenprävalenz und Themeninhalt erlaubt. Durch solch eine Berücksichtigung der Korpusstruktur (daher das “structural” im Namen) kann beispielsweise untersucht werden, ob Dokumenttyp oder Erscheinungsjahr Einfluss auf die Wörter der Themen bzw. auf ihre durchschnittliche Wahrscheinlichkeit im Korpus haben.

Abbildung 4 illustriert den Effekt einer Kovariate auf die Wort-Thema-Wahrscheinlichkeiten. Abgebildet sind die wahrscheinlichsten Wörter eines Themas zu Führungsverhalten und Betriebsführung. Das Thema wurde in einem Korpus von PSYINDEX-Einträgen identifiziert (Bittermann, 2018). Die x-Achse markiert, ob ein Wort in diesem Thema bei empirischen Studien einen höheren Rang einnimmt als bei nicht-empirischen Studien. Wenn die Kovariate keinen Einfluss auf die Wort-Thema-Wahrscheinlichkeiten hat, dann liegen alle Wörter mittig

auf der gestrichelten Linie. Hier zeigt sich, dass das Thema zu Führungsverhalten und Betriebsführung in empirischen Studien eine stärkere Ausrichtung auf Einstellungen der Angestellten und Führungsstil hat, während nicht-empirische Studien (wie etwa theoretische Diskussionen oder Übersichtsartikel) mehr auf Personalmanagement oder Organisationsentwicklung fokussieren. Demnach konzentrieren sich empirische Studien mehr auf Merkmale von Personen (die eben empirisch erhoben wurden), während nicht-empirische Studien das Thema auf einer abstrakteren Ebene (Organisation, Methoden) beleuchten.

Den Einfluss derselben Kovariate "Studienmethodik" auf die Prävalenz aller Themen zeigt Abbildung 5. Auch hier bildet die x-Achse Wahrscheinlichkeitsdifferenzen ab. Jeder Punkt markiert die Prävalenzdifferenz eines Themas bzw. die Prävalenz in empirischen Studien minus der Differenz in nicht-empirischen Studien. Die Linien um einen Punkt zeigen die Konfidenzintervalle. Themen mit einer höheren Wahrscheinlichkeit, empirisch untersucht zu werden, befinden sich auf der rechten Seite. Schließt das Konfidenzintervall eines Themas 0 ein (gestrichelte Linie), so ist der Unterschied nicht signifikant. Vor allem von nicht-empirischen Publikationen wurde Thema 20 zur Anwendung verschiedener Therapieformen behandelt, während das Thema 256 zu Leistung in Schule und Hochschule in empirischen Studien untersucht wurde. Die Ergebnisse eines solchen Topic Models mit Studienmethodik-Kovariate kann hilfreicher Ausgangspunkt für Forschungssynthesen sein: Welche Themen mit hohen Anteilen von Primärstudien existieren, und wurden ggf. noch nicht meta-analytisch untersucht?

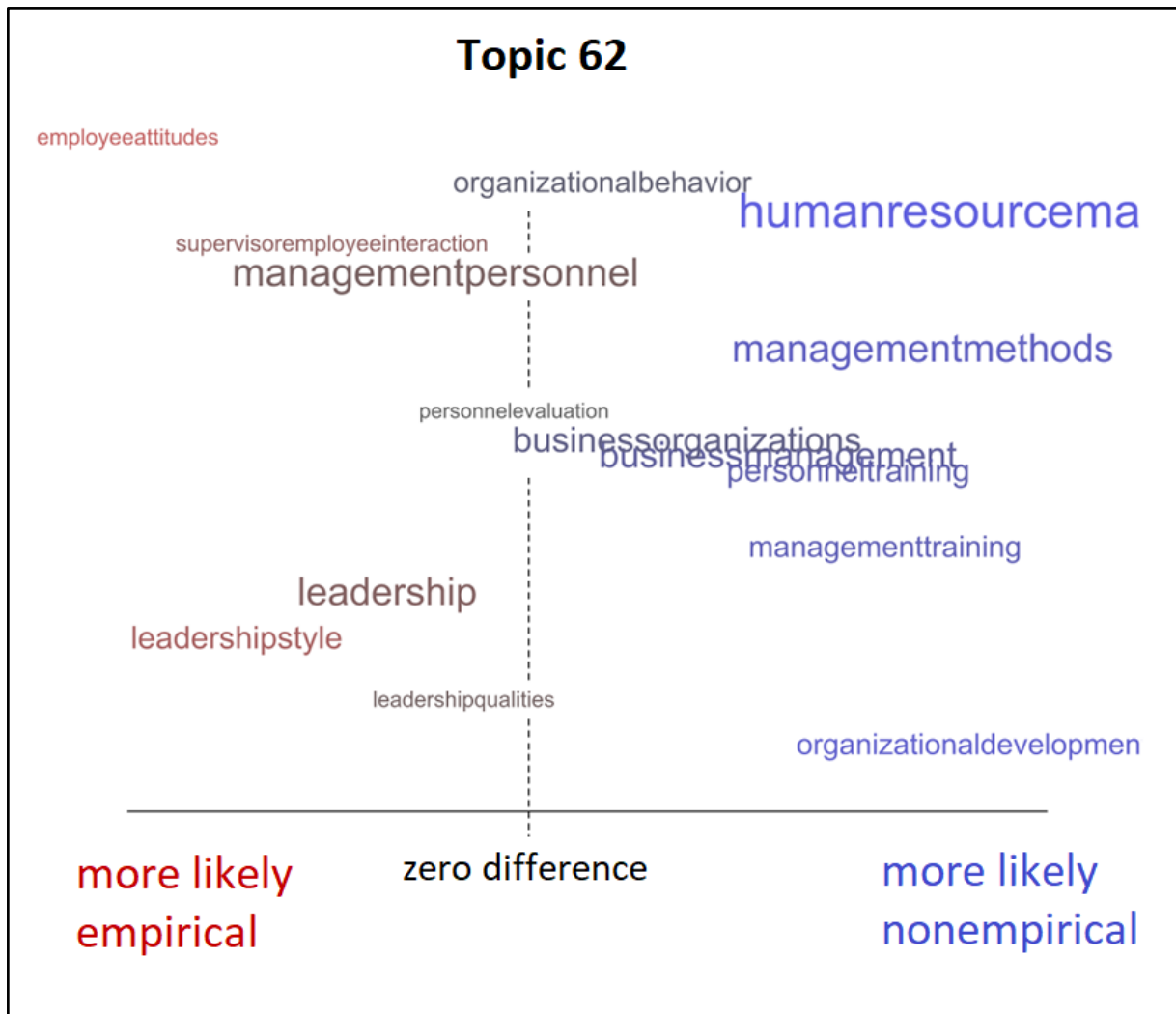


Abbildung 4. Illustration des Einflusses einer Kovariate “Studienmethodik” (empirisch vs. nicht-empirisch) auf die Wort-Thema-Wahrscheinlichkeiten eines Themas zu Führungsverhalten und Betriebsführung (modifiziert nach Bittermann, 2018). Die x-Achse markiert Wahrscheinlichkeitsdifferenzen: Je weiter außen ein Wort steht, desto wahrscheinlicher ist es, dass es in empirischen bzw. nicht-empirischen Studien eine größere Relevanz hat. Die gestrichelte Linie markiert Wahrscheinlichkeitsdifferenzen von Null. In diesem Beispiel zeigt sich, dass empirische Studien mehr auf Merkmale von Personen fokussieren, während nicht-empirische Studien das Thema eher auf Organisationsebene bzw. aus methodischer Sicht adressieren. Die y-Achse hat keine Bedeutung und dient nur einer übersichtlichen Darstellung.

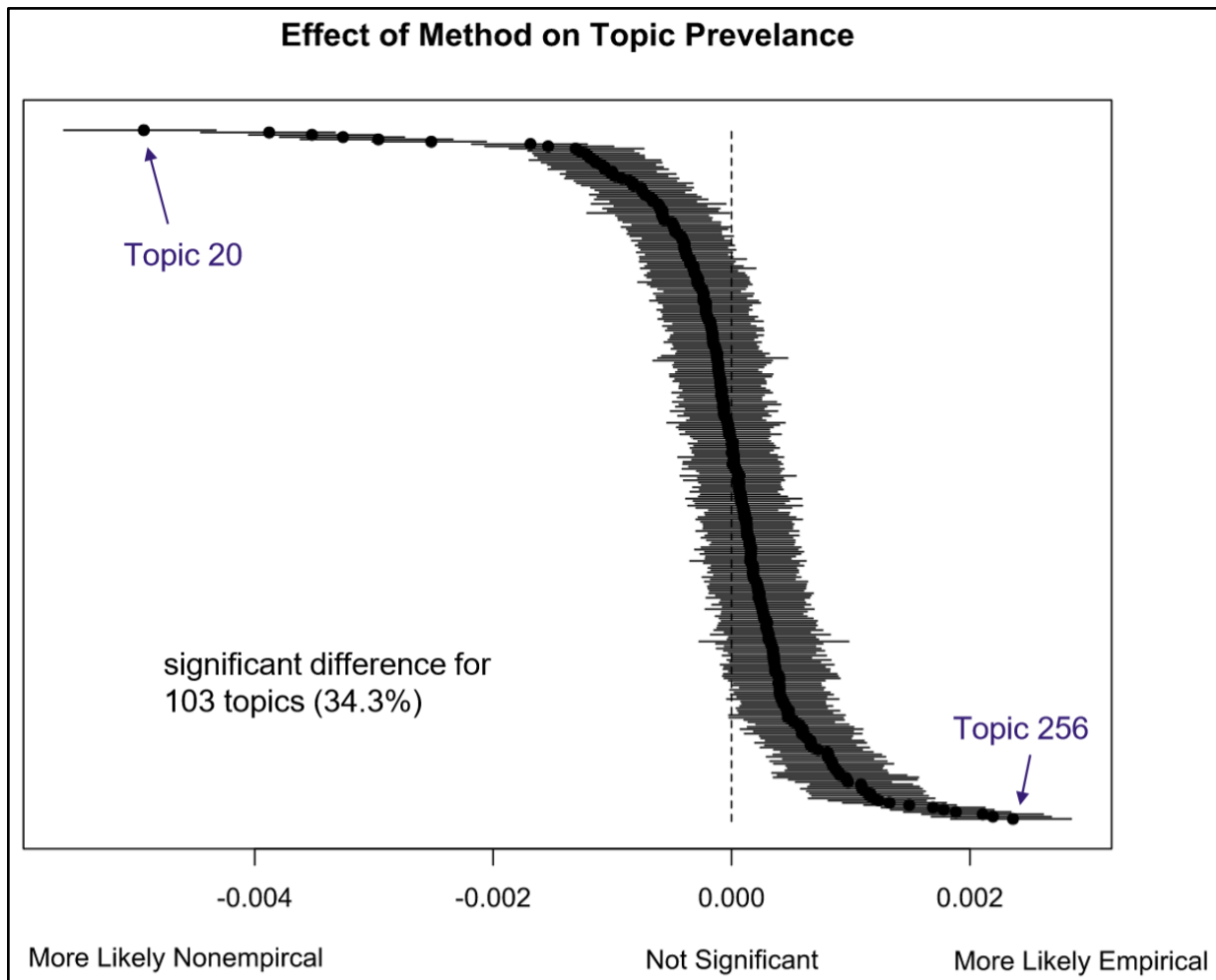


Abbildung 5. Einfluss einer Kovariate “Studienmethodik” (empirisch vs. nicht-empirisch) auf die Prävalenzen aller 300 Themen im Modell (aus Bittermann, 2018). Die x-Achse zeigt Prävalenzdifferenzen (Prävalenz_{empirisch} - Prävalenz_{nicht-empirisch}). Positive Werte bedeuten, dass ein Thema eine höhere Prävalenz bei empirischen Studien hat und vice versa. Jeder Punkt markiert die Prävalenzdifferenz eines Themas, die dazugehörige horizontale Linie das jeweilige 95 %-Konfidenzintervall. Die y-Achse hat keine Bedeutung und dient nur einer übersichtlichen Darstellung. Thema 20 (stärker nicht-empirisch adressiert) betrifft die Durchführung bestimmter Therapieformen, Thema 256 (stärker empirisch untersucht) bezieht sich auf Leistung in Schule und Hochschule.

Die Wahrscheinlichkeitsdifferenzen eines Themas werden im STM durch lineare Regressionen bestimmt. Um den Effekt auf einer Kovariate auf die Prävalenz zu bestimmen, sieht die Regressionsgleichung für das Beispiel der Studienmethodik wie in Gleichung (1) dargestellt aus (analog für die Wort-Thema-Wahrscheinlichkeiten):

$$\text{Prävalenz}_{dk} \sim \beta_0 + \beta_1 \times \text{Studienmethodik}_d + \varepsilon_d \quad (\text{Gl. 1})$$

Die Prävalenz für Thema k in Dokument d setzt sich also aus der Konstante β_0 sowie der Ausprägung der binären Variable “Studienmethodik” multipliziert mit dem Regressionskoeffizienten β_1 sowie einem Fehlerterm ε_d zusammen. Prinzipiell können auch mehrere (dummycodierte) Merkmale sowie deren Interaktionen in das Modell aufgenommen werden.

Neben dem Einfluss von Kovariaten weist STM noch einen Unterschied zu LDA auf: Die Themenprävalenzen können korreliert sein (Roberts et al., 2014). Entsprechend der impliziten Unabhängigkeitsannahme von Faktoren in der Dirichlet-Verteilung geht LDA davon aus, dass die Prävalenzen der Themen voneinander unabhängig sind (Blei & Lafferty, 2009).

Allerdings kann angenommen werden, dass in einem Artikel zum Thema “psychische Störungen im Kindesalter” ein zusätzliches klinisch-psychologisches Thema zu Therapieformen wahrscheinlicher ist, als ein Thema zur Organisationsentwicklung. Blei und Lafferty (2007) adressierten dies mit dem *Correlated Topic Model* (CTM), bei dem anstelle einer Dirichlet-Verteilung eine logistische Normalverteilung zugrunde gelegt wird. Durch diese Verteilungen können Korrelationen zwischen den Prävalenz-Verteilungen der Themen abgebildet werden, was mit einer Dirichlet-Verteilung nicht möglich ist. Abbildung 6 illustriert zwei logistische Normalverteilungen von Prävalenzen. Die logistische Normalverteilung kann Prävalenzverteilungen begünstigen, welche entweder nur Thema 1 (und 3) oder Thema 2 (und 3) oder nur beide zusammen (oder Thema 3) wahrscheinlich machen. Solche Formen kann eine Dirichlet-Verteilung nicht annehmen (vgl. Abbildung 3 in Abschnitt 2.2.3).

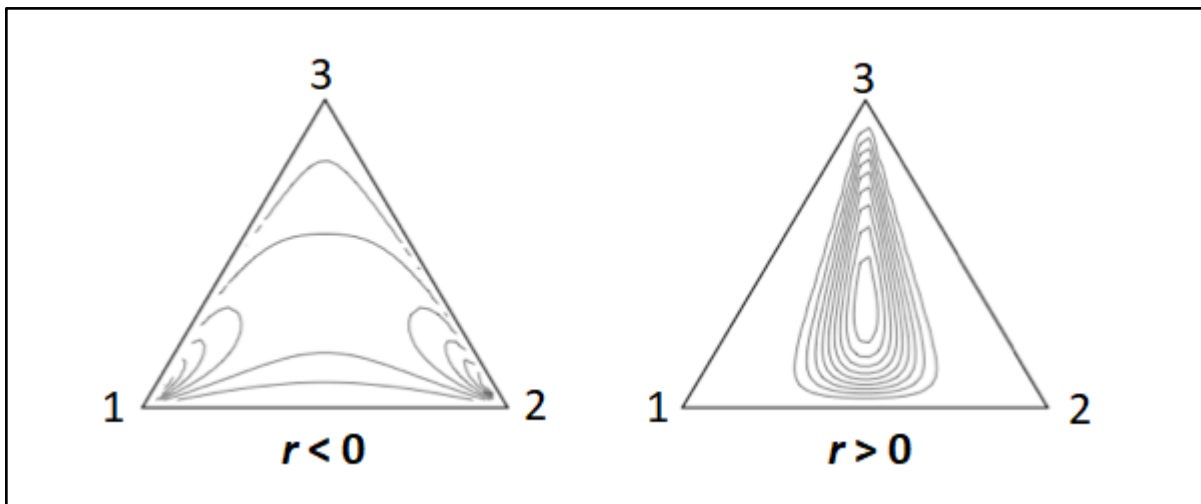


Abbildung 6. Zwei Beispiele der Wahrscheinlichkeitsdichten zweier logistisch-normalverteilter Prävalenzverteilungen (modifiziert nach Blei & Lafferty, 2007; vgl. Abbildung 3 in Abschnitt 2.2.3). Wie bei einer topografischen Karte sind hohe Dichten durch Höhenlinien markiert (wo also viele Prävalenzverteilungen liegen). Jede Ecke steht für ein Thema (1–3), die Dreiecke repräsentieren den Möglichkeitsraum aller Prävalenzverteilungen der drei Themen. Liegen die Prävalenzverteilungen in einer Ecke, ist nur dieses Thema wahrscheinlich. Liegen die Prävalenzverteilungen zwischen zwei Themen auf einer Kante, so sind nur diese beiden Themen wahrscheinlich. In der Mitte des Dreiecks sind alle Themen gleich wahrscheinlich. Entsprechend bevorzugt die Verteilung links entweder Thema 1 (und evtl. 3) oder Thema 2 (und evtl. 3), d. h., die Prävalenzen der Themen 1 und 2 sind negativ korreliert ($r < 0$). Umgekehrt zeigt die Verteilung rechts eine positive Korrelation ($r > 0$) zwischen Thema 1 und 2 (entweder beide zusammen, oder Thema 3).

Auf dem CTM aufbauend wird im STM die logistische Normalverteilung von Prävalenzen durch logistisch-normalverteilte Prävalenz-Regressionsmodelle ersetzt. Durch das Aufnehmen von Kovariaten in diese Regressionsmodelle mit Prävalenz als Kriterium (vgl. oben Gl. 1) können die Verteilungen der Prävalenzen zwischen den Dokumenten entsprechend variieren (Roberts et al., 2013). Das bedeutet also, dass nicht für jedes Dokument die gleiche Dirichlet-Verteilung bzw. logistische Normalverteilung zugrunde liegt, sondern eine individuelle logistische Normalverteilung in Abhängigkeit der Kovariate. Auch den Verteilungen der Wort-Thema-Wahrscheinlichkeiten liegt keine Dirichlet-Verteilungen zugrunde, sondern multinomiale logistische Regressionen, welche den Einfluss von Kovariaten erlauben. Die Berücksichtigung von Korrelationen der Themenprävalenzen in STM bzw. CTM hat allerdings den Effekt, dass eher breite und unspezifische Themen resultieren, auch bei einer großen Anzahl von Themen im Modell (Wiedemann & Niekler, 2018, S. 12). Steht also die Suche nach möglichst detaillierten Themen im Fokus, ist LDA (oder eine der Varianten) zu bevorzugen. Im Kontext psychologischer Forschungsthemen lässt sich ableiten, STM dann zu verwenden, wenn das Korpus bereits thematisch sehr

eingegrenzt ist und die sich am deutlichsten unterscheidenden Themen von Interesse sind (vgl. Studie 2, Abschnitt 5).

Der generative Prozess bei STM ist in den Grundannahmen mit LDA vergleichbar, allerdings weist die Modellinferenz wichtige Unterschiede auf: Anstelle von Gibbs-Sampling mit nicht-deterministischer zufälliger Initialisierung (siehe Abschnitt 2.2.3) wird ein *variational expectation-maximization (VEM) algorithm* (Grimmer, 2011) mit deterministischer *spectral initialization* (Arora et al., 2013) verwendet. Der VEM-Algorithmus schätzt im sog. “E-Schritt” (“expectation”) die erwarteten Werte und optimiert im “M-Schritt” (“maximization”) die Verteilungsparameter mittels Maximum-Likelihood-Methode. Dies wird solange wiederholt, bis keine weitere Verbesserung der Likelihood erzielt werden kann. Während des M-Schrittes werden die Hyperparameter der logistischen Normalverteilung (bzgl. Themenprävalenz) bzw. der multinomialen Regression (bzgl. Themeninhalt) aus den Daten geschätzt (Roberts et al., 2014). Während also bei LDA die α - und η -Hyperparameter vorab definiert werden müssen³⁰, werden bei STM die Hyperparameter aus den Daten bestimmt. Bei der Algorithmus-Initialisierung werden die Wörter zu Beginn nicht zufällig den Themen zugewiesen, sondern auf Grundlage einer Wort-Kookkurrenz-Matrix zugeordnet. Dies führt zu einer nachverfolgbaren Modellinferenz und dem Effekt, dass der STM-Algorithmus bei jeder Ausführung zum selben Ergebnis führt. Deshalb ist hier das Setzen eines *random seeds* nicht nötig.

Zusammenfassend kann festgehalten werden, dass STM eine Erweiterung von LDA darstellt, welche den Einfluss von Kovariaten auf Prävalenz und Themeninhalt untersuchen lässt. Damit sind auch experimentelle Untersuchungsdesigns denkbar, etwa ob eine Intervention Einfluss auf den Wortgebrauch in Umfrage-Freitexten hat (Roberts et al., 2014). Bei der Anwendung von STM sind weniger Entscheidungen vorab zu treffen als bei LDA (vor allem keine Dirichlet-Hyperparameter), was je nach Korpus und Fragestellung einerseits erleichternd (weniger Modelle zu evaluieren), andererseits einschränkend sein kann (Annahmen über das Ausmaß der thematischen Vielfalt in den Texten können nicht berücksichtigt werden). Außerdem kann die Berücksichtigung der Korrelation zwischen

³⁰ Der Vollständigkeit halber sei erwähnt, dass dies vor allem für LDA-Implementierungen in R zutrifft. Das *Gensim*-Paket (Rehurek & Sojka, 2010) für Python erlaubt ein Schätzen von alpha aus den Daten (siehe <https://radimrehurek.com/gensim/models/ldamodel.html>).

Themenprävalenzen einerseits zu schärfer abgegrenzten Themen, andererseits zum Übersehen spezifischer Themen führen.

2.2.5 Themen-Validität und Themen-Reliabilität

Die Wahl des besten Topic Models für ein bestimmtes Korpus im Sinne der adäquatesten Abbildung seiner thematischen Struktur stellt eine große Herausforderung dar, die in zwei Schritte unterteilt werden kann: (1) Zusammenstellung der Modellkandidaten und (2) Auswahl des besten Modells. Für den ersten Schritt sind vor allem die Anzahl der Themen und – bei LDA – die Dirichlet-Hyperparameter wichtige Entscheidungen. Hierfür existieren zwar keine eindeutigen Lösungen, jedoch können Best-Practice-Empfehlungen (Banks et al., 2018; Maier et al., 2018) und vergleichbare Studien in der Literatur Orientierung geben. Statistische Evaluationskriterien (z. B. Held-Out-Likelihood, semantische Kohärenz, siehe Abschnitt 2.2.2) geben hilfreiche Anhaltspunkte für den zweiten Schritt, der Auswahl des Modells. Allerdings können statistische Kennwerte allein nicht garantieren, dass die Ergebnisse vor dem Hintergrund des konkreten Untersuchungsgegenstandes sinnvoll zu interpretieren sind. Maier et al. (2018) empfehlen deshalb ein zusätzliches Sichten der Themen unterschiedlicher Modelle hinsichtlich Detailgrad, Relevanz und Interpretierbarkeit. Der Detailgrad ergibt sich aus der Anzahl an Themen (je mehr Themen, desto detaillierter können sie sein). Relevant sind Themen dann, wenn sie in Zusammenhang mit der Fragestellung stehen. Irrelevant wäre beispielsweise ein Thema, das nur aus Funktionswörtern oder häufigen Wörtern wissenschaftlicher Abstracts besteht (“Studie, Untersuchung, Ergebnis, Diskussion”). Die Interpretation der Themen ist allerdings eine größere Herausforderung.

Topic Models stellen mit den Wort-Thema-Wahrscheinlichkeiten und den Dokument-Thema-Wahrscheinlichkeiten zwei Verteilungen zur Verfügung, mit denen die “semantische Validität” (Maier et al., 2018, S. 99) der Themen bestimmt werden kann. Aus den Wort-Thema-Wahrscheinlichkeiten können die wahrscheinlichsten Wörter entnommen werden, die das Thema charakterisieren. Für beispielsweise die fünf Wörter "Backfisch", "Schön", "Groß", "Gefallen", und "Gesehen" scheint es zunächst naheliegend, dass es sich um ein Thema zur Präferenz von Lebensmitteln handelt. Das Zuweisen eines solchen Titels oder Labels zu den Themen allein anhand der Wortcluster ist allerdings anfällig für Fehlinterpretationen, wenn die Bedeutung der einzelnen Wörter nicht korpusweit definiert ist (wie etwa anhand eines Thesaurus). Hilfreich ist eine grundlegende Kenntnis des Korpus,

beispielsweise hinsichtlich Zeit und Kontext der Textentstehung, denn Begriffe können ihre Bedeutungen im Verlauf der Zeit ändern. Wenn das Korpus Dokumente umfasst, die vor dem Ende des 20. Jahrhunderts entstanden sind, ist allein anhand dieser fünf Wörter nicht auszuschließen, dass mit “Backfisch” ein weiblicher Teenager gemeint ist. Deshalb sollten zur Themeninterpretation weiterhin die repräsentativsten Dokumente für das Thema eingesehen werden. Diese Dokumente sind diejenigen mit der höchsten durchschnittlichen Dokument-Thema-Wahrscheinlichkeit im Korpus. Das Sichten einzelner Dokumente in ihrer ursprünglichen Form (z. B. das gesamte Abstract mit Titel) erleichtert das Verständnis der Wortcluster.

Maier et al. (2018, S. 107) betonen, dass Interpretierbarkeit ein wichtiges, aber kein hinreichendes Kriterium für Themen-Validität ist. Zusätzlich muss sichergestellt werden, dass die ermittelten Themen auch den intendierten untersuchten Gegenstand repräsentieren (Konstruktvalidität). Dabei weisen sie auf Möglichkeiten der externen Validierung hin, etwa anhand der zeitlichen Korrespondenz von thematisch relevanten Ereignissen mit Publikationsverläufen (z. B. Evans, 2014) oder Übereinstimmung mit einem Code-System (z. B. Jacobi, Van Atteveldt & Welbers, 2015). So ist bzgl. zeitlicher Korrespondenz anzunehmen, dass ein Thema zu Terror sich in einer Zunahme entsprechender Publikationen nach dem 11. September 2001 widerspiegelt. Als Code-System bietet sich im Kontext psychologischer Forschungsthemen das APA-Klassifikationssystem an (Tuleya, 2007; siehe Abschnitt 2.1.2). Ein Thema, dessen Wörter sich als “Störungsbild der Depression” interpretieren lassen, sollte vor allem von solchen Publikationen adressiert worden sein, die mit “3211 Affektive Störungen” klassifiziert worden sind.

Interpretierbare und valide Themen alleine sind jedoch von geringem Nutzen, wenn sie nicht zuverlässig reproduziert werden können. Dies ist vor allem bei LDA wichtig (vgl. Abschnitt 2.2.3), da verschiedene Modellinferenzen (Algorithmus-Wiederholungen) aufgrund der zufälligen Initialisierung des Algorithmus zu unterschiedlichen Ergebnissen führen können. Die Zuverlässigkeit, mit der ein Thema auch bei verschiedenen Inferenzen gefunden werden kann, bezeichnen Maier et al. (2018) als Themen-Reliabilität. Sie empfehlen hierfür die Metrik von Niekler (2016), welche die Ähnlichkeit der Themenpaare zweier Modellinferenzen quantifiziert. Dazu wird die Kosinusähnlichkeit der jeweils wahrscheinlichsten Wörter der Themenpaare ermittelt. Die Kosinusähnlichkeit ist ein Maß für die Ähnlichkeit zweier Vektoren mit dem Wertebereich von -1 bis 1, wobei 1 bei

vollständiger Übereinstimmung der Vektoren (die gleichen Elemente an der gleichen Stelle) erreicht wird. Ein Thema kann nach Niekler (2016) dann als reliabel erachtet werden, wenn es z. B. bei mindestens acht von zehn Modellinferenzen eine Kosinusähnlichkeit von mindestens .8 aufweist.

Zusammenfassend soll nach Maier et al. (2018) die Überprüfung der semantischen Validität in Verbindung mit statistischen Maßen wie der semantischen Kohärenz (Mimno et al., 2011) und ggf. einer externen Validierung gewährleisten, dass interpretierbare Themen resultieren, die auch den intendierten Untersuchungsgegenstand abbilden. Um sicherzustellen, dass diese Themen nicht zufällig zustande gekommen sind, sollte bei LDA die Reliabilität der Themen über verschiedene Modellinferenzen hinweg überprüft werden. Da die Modellinferenz bei STM deterministisch ist, ist bei dieser Variante eine Analyse der Themen-Reliabilität nicht notwendig.

2.3 Diskussion der automatisierbaren Ansätze und Forschungslücken

Im Folgenden sollen die in den letzten Abschnitten vorgestellten Ansätze im Hinblick auf die automatisierte Identifikation psychologischer Forschungsthemen bewertend verglichen werden. Der Fokus auf Automatisierbarkeit ergibt sich aus dem Anspruch, auch große und stetig wachsende Textkorpora analysieren zu können, was mit traditionellen qualitativen Inhaltsanalysetechniken nicht durchführbar ist. Aus den bisherigen Studien zu Forschungsthemen der Psychologie werden Forschungslücken abgeleitet, welche Ausgangspunkt für die Abschnitt 3.1 aufgeführten Ziele der Dissertation aufgegriffen werden.

Tabelle 7. Zentrale Unterscheidungsmerkmale automatisierbarer Ansätze zur Identifikation von psychologischen Forschungsthemen.

	Topic Modeling	Klassifikations-analyse	Frequenzanalyse	Singular Value Decomposition mit Cluster-analysen
Beispielanwendungen in der Psychologie:	Oh et al. (2017)	Krampen & Perrez (2015)	Webster et al. (2009)	Lindahl et al. (2015)
potenzieller Detailgrad:	Hoch	Gering	Gering bis mittel	Hoch
Themenableitung:	Bottom-up	Top-down	Bottom-up	Bottom-up
Statistik-Zweig:	Bayes	Frequentistisch	Frequentistisch	Frequentistisch
Zentrale Stärke:	Spezifisch für Textdaten ausgerichtete Dimensionsreduktion auf Wortcluster	Einfache Interpretation durch definierte Kategorien*	Einfache Analyse	Identifikation unkorrelierter Komponenten in Dokument-Term-Matrizen
Zentrale Schwäche(n):	Probabilistische Interpretation der Ergebnisse	Es können keine anderen Themen als die vorab definierten Kategorien gefunden werden, Kategorien tlw. sehr breit	Nur ein Thema pro Jahr / Kategorie, Themeninhalte sehr heterogen	Dokumente können nur einem Thema zugeordnet werden

Anmerkung: Es werden nur die Hauptstärken bzw. -schwächen berichtet, die den jeweiligen Ansatz von den anderen unterscheiden (für ausführlichere Beschreibungen vgl. Abschnitt 2.1.2).

* Die Inhalte der Kategorien in PSYINDEX werden in den “Erläuterungen zum Klassifikationssystem (Scope Notes)” erläutert (Gerards et al., 2014, S. 34–57).

Alle Ansätze haben zunächst das gemeinsame Ziel, wissenschaftliche Fachpublikationen nach ihrer inhaltlichen Ähnlichkeit zu gruppieren, um daraus die gemeinsamen Themen abzuleiten. Anhand der in Tabelle 7 zusammengefassten Unterscheidungsmerkmale “potenzieller Detailgrad”, “Themenableitung” und “Statistik-Zweig” können die zentralen Stärken und Schwächen der Ansätze verglichen und für die Identifikation psychologischer Forschungsthemen bewertet werden. Als am wenigsten geeignet können einfache Frequenzanalysen (z. B. Webster et al., 2009) beurteilt werden. Hier werden die insgesamt

häufigsten Wörter in eine Rangfolge gebracht und als Thema aufgefasst. Obwohl damit Wörter aus Volltexten analysiert werden können, leidet der Detailgrad der gefundenen Themen an der mangelnden Trennung von zusammengehörigen Wortgruppen: Die häufigsten Begriffe eines Jahres ergeben eine inhaltlich zu heterogene Gruppe und es ist nicht ersichtlich, welche Wörter mit welchen auch zusammen in den Publikationen verwendet wurden. Zwar wäre eine Kombination mit dem klassifikationsbasierten Ansatz von Krampen und Perrez (2015) denkbar, so dass die häufigsten Wörter getrennt nach Klassifikation gebildet werden. Doch auch hier ergibt sich das Problem der inhaltlichen Breite von Klassifikationskategorien und enthaltenen Subkategorien, die einfache Frequenzanalysen nicht identifizieren können. So umfasst beispielsweise die Kategorie “2910 Soziale Strukturen” des PSYINDEX-Klassifikationsschemas unter anderem soziale Schichten, gesellschaftliche Machtstrukturen, Demografie sowie die Folgen von Migration (Gerards et al., 2014, S. 41). Mit Klassifikationsansätzen können also nur solche Themen gefunden werden können, die vorab definiert wurden. Dieser Top-Down-Zugang zu Forschungsthemen verhindert eine flexible Anpassung an sich schnell ändernde Forschungstrends und -schwerpunkte.

Potenziell detaillierte und verschiedene Themen, die bottom-up aus den Daten ermittelt werden, ermöglichen Kombinationen aus *Singular Value Decomposition* (SVD) mit Clusterverfahren (z. B. Lindahl et al., 2015). Dabei werden ähnliche Dokumente in gemeinsame Gruppen zusammengefasst, aus denen dann die zugrundeliegenden Themen abgeleitet werden. “Hartes” Clustern führt allerdings zum entscheidenden Nachteil, dass ein Dokument nur einem Thema zugeordnet werden kann. Zwar können bei sogenannten “fuzzy” oder “weichen” Clusterverfahren (Bezdek, 1981; Dunn, 1973) Elemente auch zu mehreren Clustern gehören. Allerdings setzen auch diese Clusterverfahren dichte (“dense”) Datenmatrizen voraus, was jedoch bei numerischen Repräsentationen von Textdaten nicht der Fall ist (Schubert, 2017). Sogenannte Dokument-Term-Matrizen (DTM; vgl. Tabelle 4 in Abschnitt 2.2.2) enthalten für jedes Wort im Korpus eine Spalte mit Häufigkeiten. Da nicht jedes Dokument alle Wörter im Gesamtvokabular des Korpus verwendet, enthalten viele Zellen in der DTM keine Werte (“sparse matrices”).

Topic Modeling hingegen ist spezifisch für Textdaten entworfen und kann als eine Art Hauptkomponentenanalyse für diskrete Daten gedacht werden (Blei, 2012, S. 80). Im direkten Vergleich erwies sich Topic Modeling (LDA) der Hauptkomponentenanalyse hinsichtlich semantischer Kohärenz der Themen überlegen (Hecking & Leydesdorff, 2019). Allerdings

erzielt Topic Modeling als Verfahren der Bayes-Statistik probabilistische Ergebnisse, deren Interpretation und Validierung (siehe Abschnitt 2.2.5) herausfordernder ist als die von Häufigkeiten. Wenn jedoch Möglichkeiten der Untersuchung von Themen-Validität und Reliabilität (Maier et al., 2018) berücksichtigt werden, ist Topic Modeling den anderen hier vorgestellten Ansätzen deutlich überlegen: Die Themen können bottom-up aus den Daten abgeleitet werden, wobei Dokumente mehrere Themen adressieren können. Die spezifisch für Textdaten berechneten Wort-Thema- und Dokument-Thema-Wahrscheinlichkeiten ermöglichen ein detailliertes Explorieren des Korpus. Da Topic Modeling somit einen induktiven Ansatz mit quantifizierbaren Resultaten kombiniert (Maier et al., 2018), bietet sich dieses Verfahren vor allem für explorative Untersuchungen an, das heißt, wenn die potenzielle Vielfalt der Themen noch nicht oder nur eingeschränkt bekannt ist.

In der Psychologie wurde Topic Modeling bislang für verschiedene andere Fragestellungen eingesetzt. Zum Beispiel verwendeten Griffiths, Steyvers und Tenenbaum (2007) dieses Verfahren zur Vorhersage von Wortassoziationen und die Auswirkungen von semantischer Assoziation und Mehrdeutigkeit auf eine Vielzahl von Sprachverarbeitungs- und Gedächtnisaufgaben. Steyvers und Griffiths (2008) zeigten durch den Einsatz von Topic Modeling, dass sowohl das menschliche Gedächtnis als auch die Informationsbeschaffung ähnlichen Rechenanforderungen ausgesetzt sind. Topic Modeling wurde unter anderem auch für die Modellierung von Paar- und Familientextdaten (Atkins et al., 2012), zur Verbesserung der Vorhersage von Neurotizismus und Depressionen (Resnik, Garron & Resnik, 2013) oder zur Untersuchung von Signalen zur psychischen Gesundheit auf Twitter verwendet (Coppersmith, Dredze & Harman, 2014).

Somit ist das Verfahren in der Psychologie zwar nicht unbekannt, doch für die Analyse von psychologischen Forschungsthemen kaum erprobt. Zu Beginn dieser Dissertation lag nur die Studie von Oh, Stewart und Phelps (2017) als Ausnahme hierfür vor. In dieser Untersuchung wurden 64 Themen in 3603 zwischen 1963 und 2015 veröffentlichten Artikeln der Fachzeitschrift *Journal of Counseling Psychology* ermittelt. Als Topic-Modeling-Variante wurde LDA gewählt; eine Untersuchung der Reliabilität der damit gefundenen Themen wurde jedoch nicht vorgenommen. Interpretiert wurden die Themen anhand der jeweils wahrscheinlichsten Wörter und der Titel der repräsentativsten Dokumente ("semantische Validität" sensu Maier et al., 2018; siehe Abschnitt 2.2.5). Möglichkeiten der externen Validierung wurden nicht aufgezeigt.

Als eine wegweisende Studie zur Identifikation von Forschungsthemen mit Topic Modeling ist die disziplinübergreifende Arbeit von Griffiths und Steyvers (2004) zu nennen. Sie ermittelten „Hot Topics“ durch eine Analyse von 28154 Abstracts, die zwischen 1991 und 2001 in den *Proceedings der National Academy of Sciences* (PNAS) der Vereinigten Staaten von Amerika veröffentlicht wurden. Ein Thema wurde als "hot" definiert, wenn es einen steigenden linearen Trend der Prävalenz zeigte. Die Themen wurden mit einem Kategoriensystem der PNAS validiert, welches verschiedene wissenschaftliche Disziplinen umfasst (z. B. Agrarwissenschaften, Biologie, Psychologie oder Sozialwissenschaften). Es konnte gezeigt werden, dass die Inhalte der Artikel mit diesen breiten Kategorien übereinstimmen. Die Autoren kommen zu dem Schluss, dass ein Ermitteln der Themen, die der Struktur solcher Datensätze zugrunde liegen, der erste Schritt ist, um deren Inhalt zu visualisieren und sinnvolle Trends zu entdecken.

Angesichts des disziplinenübergreifenden Korpus bei Griffiths und Steyvers (2004) bleibt es eine offene Frage, ob ein Übertragen ihres Ansatzes auch zu einer Identifikation und Validierung von Forschungsthemen in der Psychologie möglich ist. Dazu müsste ein analoges Kategoriensystem verwendet werden, welches hinreichend differenziert für einen Vergleich mit Themeninhalten ist. Zusätzlich zu einer Überprüfung der semantischen Validität wie bei Oh et al. (2017) könnten die Themen somit auch extern validiert werden, was das Vertrauen in die Ergebnisse erhöhen würde.

Weiterhin waren die Korpora bei beiden Studien sprachlich homogen. Für eine Untersuchung von allen Beiträgen aus der gesamten Psychologie müsste der Tatsache Rechnung getragen werden, dass nicht alle Publikationen bzw. Abstracts in englischer Sprache verfügbar sind. Umgekehrt gilt für Untersuchungen im deutschsprachigen Raum, dass angesichts einer steigenden Zahl englischsprachiger Publikationen (Schui & Krampen, 2016) effektive Lösungen für den Umgang mit Mehrsprachigkeit gefunden werden müssten.

Mehrsprachigkeit kann bei Topic Modeling zu verschiedenen Themen trotz gleicher Inhalte führen (z. B. englisches "Therapy"-Thema und deutsches "Therapie"-Thema) bzw. die semantische Kohärenz der Themen (Mimno et al., 2011) senken. Dies ist beispielsweise bei kumulativen Dissertationen der Fall, wenn in einem deutschsprachigen Abstract die englischen Titel der Einzelbeiträge aufgeführt sind. Zwar existieren Topic-Modeling-Varianten für mehrsprachige Textquellen (z. B. Mimno, Wallach, Naradowsky, Smith & McCallum, 2009; Vulić, De Smet, Tang, & Moens, 2015), jedoch stellen diese hohe

Anforderungen der Datenvorbereitung: Thematisch passende Dokumente müssen sprachübergreifend vorab gepaart werden. Dies stellt einerseits eine hohe Hürde für eine Automatisierung dar und kann durch ungleiche Verteilungen von Sprachen in bestimmten Themenbereichen erschwert werden: In der Psychologie aus dem deutschsprachigen Raum wird beispielsweise in den Bereichen der Neuropsychologie, Elektrophysiologie und Allgemeinen Psychologie hauptsächlich auf Englisch publiziert (siehe PsychArchives-ESM 1 zu Studie 3 in Anhang D). Damit kann konstatiert werden, dass noch keine Möglichkeit aufgezeigt wurde, Publikationen unabhängig von der Abstract-Sprache mit automatisierbaren Topic-Modeling-Varianten zu analysieren.

Schließlich ist festzuhalten, dass in der Psychologie Effekte von Dokumentmetadaten, wie Studientyp oder Autoreninformation, bisher völlig unberücksichtigt blieben. Bisherige Fragestellungen konzentrierten sich auf die reine Identifikation von Themen. Durch die Berücksichtigung von Dokument- oder Autorenvariablen in entsprechenden Topic-Modeling-Varianten eröffnen sich Möglichkeiten für differenziertere Untersuchungen. Beispielsweise wäre im Rahmen von Forschungssynthesen die Information interessant, zu welchen Themen schon sehr viel oder bislang kaum empirisch geforscht wurde.

3. Gegenstand der vorliegenden Dissertation

3.1 Forschungsziele und Datengrundlage

Ausgehend von den in Abschnitt 2.3 dargelegten Forschungslücken untersucht die vorliegende Dissertation anhand dreier Anwendungsszenarien, wie Forschungsthemen der gesamten Psychologie valide und automatisiert identifiziert werden können. Dazu werden vier übergeordnete Forschungsziele verfolgt:

- (1) Der Vergleich von Topic Modeling mit etablierten klassifikationsbasierten Ansätzen zur Themenidentifikation,
- (2) die Prüfung der Eignung eines Datenbank-Klassifikationssystems der Psychologie zur externen Validierung von Themen,
- (3) die Erprobung von Möglichkeiten zum Umgang mit Mehrsprachigkeit von psychologischen Fachpublikationen, und
- (4) Entwicklung einer App zur automatisierten Identifikation von Forschungsthemen in der Psychologie als Anwendungsfall von Topic Modeling.

Zur externen Validierung von speziell psychologischen Themen wird vor allem ein Klassifikationssystem psychologischer Fachliteratur (Tuleya, 2007; ZPID, 2016) herangezogen. Dem Ansatz von Griffiths und Steyvers (2004) folgend ist ein Thema dann extern valide, wenn seine Inhalte den inhaltlich kongruenten Kategorien des Klassifikationssystems entsprechen (vgl. Abschnitt 2.2.5). Damit soll sichergestellt werden, dass das Thema auch das erfasst, was untersucht werden soll. Die drei Studien sollen grundlegende Erkenntnisse hervorbringen, die in die Entwicklung einer nutzerfreundlichen Forschungsthemen-App (siehe Abschnitt 7) einerseits und einer Entscheidungsheuristik zur Auswahl einer geeigneten Topic-Modeling-Variante (siehe Abschnitt 8.4) andererseits einfließen. Außerdem sollen auch Anknüpfungspunkte für digitale Produkte von forschungsbasierten Infrastruktureinrichtungen in der Psychologie am Beispiel des ZPID aufgezeigt werden.

Neben diesen methodischen Zielsetzungen werden anhand der Analyse von Fachpublikationen innerhalb von drei Studien auch inhaltliche Fragestellungen adressiert, die vor allem für die deutschsprachige Psychologie von Relevanz sind. Studie 1 konzentriert sich auf Brennpunktthemen bzw. “Hot Topics” der Psychologie. Dass die Ermittlung von Brennpunktthemen der Psychologie von aktueller Relevanz ist, zeigen die 2016 eingeführten “Hot Topic Sessions” auf DGPs-Kongressen der Deutschen Gesellschaft für Psychologie (Klostermann, 2016). In Abschnitt 1 wurde die Forderung von Wissenschaftsrat (2018, S. 83) wiedergegeben, dass die Psychologie zur Bewältigung von Herausforderungen der Gesellschaft beitragen muss. Vor diesem Hintergrund wird Topic Modeling in Studie 2 verwendet, um unter anderem zu untersuchen, ob und mit welchen Inhalten psychologische Beiträgen den Themenbereich “Flucht und Migration” behandeln. Studie 3 greift eine Diskussion zu geschlechtsspezifischen Forschungspräferenzen in der wissenschaftlichen Psychologie auf (z. B. Gundlach, Tröster & Moschner, 1999), indem methodischen Einschränkungen bisheriger Untersuchungen durch eine nahezu Vollerhebung von Dissertationen und einer Auswertung mit Topic Modeling begegnet wird.

Als Datengrundlage wird die Fachdatenbank³¹ *PSYNDEX* herangezogen. Diese Referenzdatenbank für psychologische Fachliteratur wird vom Leibniz-Zentrum für Psychologische Information und Dokumentation (*ZPID*)³² produziert und umfasst den gesamten deutschen Sprachraum, womit Fragestellungen mit direktem Bezug zur Psychologie sowie zu gesellschaftlichen Themen der deutschsprachigen Länder möglich sind. Außerdem stellt *PSYNDEX* aus den folgenden Gründen auch für die vorliegende Methodenerprobung eine günstige Ausgangslage dar (vgl. im Folgenden auch Bosnjak & *ZPID*, 2018, S. 12):

- (1) Mit über einer Dreitmillion³³ Publikationsnachweisen kann ein Korpus gebildet werden, das die Grenzen qualitativer inhaltsanalytischer Ansätze mit manueller Kodierung überschreitet und das Potenzial von Text-Mining-Methoden wie Topic Modeling ausnutzt.
- (2) Es werden die gesamte Psychologie sowie ihre angrenzenden Nachbarggebiete abgedeckt, wodurch eine hohe thematische Vielfalt vorliegt. Dadurch kann der Einsatz

³¹ An dieser Stelle sei noch einmal kurz festgehalten, dass mit “Fachdatenbank” eine Datenbank von Literaturnachweisen gemeint ist, während ein “Klassifikationssystem” eine Ordnung von inhaltlichen Kategorien ist, in die die Literaturnachweise eingeordnet werden.

³² <https://leibniz-psychology.org>

³³ Auf Anfrage umfasst *PSYNDEX* 357000 Nachweise im Januar 2020.

von Topic Modeling sowohl für sehr heterogene als auch thematisch eingegrenzte Subkorpora geprüft werden.

- (3) Alle Nachweise sind ungeachtet der Originalsprache sprachlich einheitlich mit einem standardisierten Vokabular von Schlüsselbegriffen (ZPID, 2016) indexiert. Wenn die Verwendung eines solchen Vokabulars für Topic Modeling zu validen Themen führt, wäre dies eine Möglichkeit, dem Problem der Mehrsprachigkeit zu begegnen (vgl. Abschnitt 2.3).
- (4) Die Datenbank sichert ein hohes Qualitätsniveau unter anderem dadurch, dass geschulte Psychologinnen und Psychologen nach definierten Kriterien Abstracts überprüfen und ggf. überarbeiten sowie inhaltliche Metadaten annotieren (unter anderem die standardisierten Schlüsselbegriffe und Klassifikationen). Entsprechend wurde PSYNDEX in der letzten Evaluierung durch die Leibniz-Gemeinschaft mit “sehr gut” bewertet (Senat der Leibniz-Gemeinschaft, 2018, S. 2). Dies ist insofern besonders wichtig, als dass die Interpretation und Validität der gefundenen Themen von der Qualität der analysierten Texte abhängt (sogenannte „Garbage-In-Garbage-Out“-Problematik; Johnston, Jones & Kruger, 2019, S. 220).
- (5) PSYNDEX ist strukturell kompatibel mit PsycInfo, der von der *American Psychological Association* produzierten Fachdatenbank. Damit können die gewonnenen Einsichten zur Anwendung von Topic Modeling auch auf diese Datenbank und somit auf die internationale Psychologie ausgeweitet werden.

3.2 Anmerkungen zum methodischen Vorgehen und zu den Schwerpunkten der drei Studien

Die Auswahl von LDA und STM als Topic-Modeling-Varianten ergibt sich zunächst durch ihre jeweiligen Stärken: LDA erlaubt durch Hyperparameter-Tuning eine spezifische Anpassung an die vorliegenden Daten, während STM explizit den Einfluss von Dokumentvariablen zu modellieren erlaubt. Hinzu kommt der Vorteil, dass entsprechende Software zu beiden Varianten sehr gut dokumentiert ist und in der Anwendungs-Community intensiv diskutiert³⁴ wird, was die Software einerseits vertrauenswürdiger (Fehlerbehebung),

³⁴ Eine Suche auf der Softwareentwickler-Plattform “stackoverflow.com” erzielt im Januar 2020 für “[topic-modeling] lda” 500, für “[topic-modeling] stm” 44 Treffer. Zum Vergleich: Eine populäre Variante für sich dynamisch ändernde Themeninhalte über die Zeit, *Dynamic Topic Modeling* (Blei & Lafferty, 2006), erzielt mit

andererseits einfacher anzuwenden macht. Schließlich existieren für diese beiden Varianten Best-Practice-Empfehlungen (Banks et al., 2018; Maier et al., 2018).

Methodisch verfolgen die drei Studien einen stringenten Prozess des sukzessiven Setzens neuer Schwerpunkte. In Studie 1 soll zunächst als “Proof-of-concept” die Überlegenheit von Topic Modeling zum bislang etabliertesten Ansatz zur Identifikation psychologischer Forschungsthemen gezeigt werden. Dabei wird die Themen-Validität durch Verwendung des APA-Klassifikationssystems (Tuleya, 2007; ZPID, 2016) geprüft. Diese Möglichkeit der externen Validierung wird in Studie 2 durch den Vergleich von Thementrends mit externen Ereignissen ergänzt. Daneben wird der Effekt von Dokument-Metadaten auf die Themenprävalenz untersucht: Die Studienmethodik der Publikation (im Sinne von “empirisch” versus “nicht-empirisch”) wird als Dokumentvariable in das Modell aufgenommen. Studie 3 rückt die Themen-Reliabilität in den Vordergrund. Diese betrifft nur LDA (siehe Abschnitt 2.2.5) und soll als Fortführung zu Studie 1 quantifiziert werden. Dies ist bei Studie 3 von besonderer Bedeutung, da hier eine Unterschiedsfragestellung (wie unterscheiden sich Frauen und Männer in ihren Forschungsinteressen) im Fokus steht, für deren Beantwortung robuste Ergebnisse von besonderer Relevanz sind. Für diese Fragestellung wird auch der Einfluss einer Autorenvariable untersucht. Durch die in jeder Studie gewonnenen Einsichten entsprechend der gesetzten methodischen Schwerpunkte wird die Modellwahl (LDA-Hyperparametern, Themenanzahl) optimiert. Dies führt zunehmend belastbaren Ergebnissen. Zeitliche Trends stellen bei allen drei Studien einen methodischen Schwerpunkt dar, wobei in Studie 1 Trendanstiege (“Hot Topics”), bei Studie 2 Publikationsspitzen (hohe Prävalenz zu einem bestimmten Zeitpunkt) und bei Studie 3 Trend-Dyaden (Prävalenz des Themas bei Frauen versus Männern) analysiert werden. Tabelle 8 gibt einen Überblick der inhaltlichen und methodischen Schwerpunkte der drei Studien.

Entsprechend der methodischen Ausrichtung dieser Dissertation fokussieren die Kurzdarstellungen der drei Studien in den Abschnitten 4 bis 6 vor allem auf die methodischen Aspekte und Hintergründe. Um die Fortentwicklung der Methodik über die drei Studien hinweg zu verdeutlichen, wird das methodische Vorgehen teilweise ausführlicher berichtet als in den Originalpublikationen. Außerdem werden die Ergebnisse den methodischen

“[topic-modeling] dynamic topic model” 11 Treffer, eine Variante für kurze Texte, *Biterm Topic Model* (Yan, Guo, Lan & Cheng, 2013), erbringt mit “[topic-modeling] btm” 4 Treffer.

Schwerpunkten entsprechend diskutiert. Ergänzende inhaltliche Ergebnisdarstellungen und Diskussionen können den Originalpublikationen in den Anhängen 1–5 entnommen werden.

Tabelle 8. Übersicht der inhaltlichen und methodischen Schwerpunkte der drei Studien.

Studie	Topic Model	Inhaltlicher Schwerpunkt	Methodische Schwerpunkte
1	LDA	Brennpunktthemen der Psychologie	<ul style="list-style-type: none"> ● Proof-of-concept: Topic Modeling vs. Klassifikationssystem zur Identifikation von Themen in der Psychologie ● Themen-Validität via Code-System ● Trendanalyse
2	STM	Verhältnis zwischen gesellschaftlichen Herausforderungen und den Themen der psychologischen Fachliteratur	<ul style="list-style-type: none"> ● Themen-Validität via zeitlicher Kongruenz ● Verbesserte Modellwahl ● Modellierung einer Dokumentvariable: Studienmethodik ● Analyse von Publikationsspitzen
3	LDA	Forschungsinteressen von Psychologinnen und Psychologen	<ul style="list-style-type: none"> ● Themen-Reliabilität ● Optimierung der LDA-Hyperparameter und der Modellwahl ● Modellierung einer Autorenvariable: Geschlecht ● Vergleich von Trend-Dyaden

Anmerkungen: Zu den Begriffen “Validität” und “Reliabilität” im Kontext von Topic Modeling siehe Abschnitt 2.2.5. LDA = Latent Dirichlet Allocation (Blei et al., 2003), STM = Structural Topic Model (Roberts et al., 2013).

4. Studie 1: “How to Identify Hot Topics in Psychology Using Topic Modeling”

Bittermann, A. & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, 226(1), 3–13. <https://doi.org/10.1027/2151-2604/a000318>

4.1 Hintergrund

Um festzustellen, welche Themen als “Hotspots”, also Themen von besonderer Bedeutung in forschungsaktiven Bereichen (Erdfelder & Bošnjak, 2016), erachtet werden können, existieren unterschiedliche Ansätze (siehe Abschnitt 2.1.2), wobei die Verwendung von Datenbank-Klassifikationen den etabliertesten Ansatz darstellt (z. B. Gnambs et al., 2006; König et al., 2015; Krampen & Perrez, 2015). Daher galt es in dieser Studie zu überprüfen, ob Topic Modeling im Vergleich zum klassifikationsbasierten Vorgehen einen Mehrwert im Sinne von Detailgrad und Themenabdeckung bietet. Das Klassifikationssystem wird außerdem als Möglichkeit zur externen Validierung von Themen, die mit Topic Modeling identifiziert wurden, untersucht.

4.2 Fragestellungen

- (1) Welche Trends psychologischer Forschungsthemen können mit Topic Modeling identifiziert werden?
- (2) Wie verhält sich Topic Modeling zum Klassifikationsansatz hinsichtlich Detailgrad und Überschneidung der Themeninhalte?

4.3 Methode

Die methodischen Schwerpunkte umfassen (1) Proof-of-concept, (2) externe Themen-Validierung und (3) Trendanalyse.

4.3.1 Daten

Datengrundlage bildeten die standardisierten Schlagwörter von $N = 314573$ aller in der psychologischen Fachdatenbank PSYINDEX nachgewiesenen Publikationen aus den Jahren 1980 bis 2016. Die Verwendung des standardisierten Vokabulars des APA-Thesaurus

(Tuleya, 2007) in Form der “PSYINDEX Terms” (ZPID, 2016) kann vor allem bei einem Korpus, das die thematische Bandbreite der gesamten Psychologie abdeckt, als optimaler Kompromiss aus Detailgrad und Zusammenfassung von Inhalten angesehen werden, wodurch viele Probleme der Datenaufbereitung von Freitexten (etwa Synonyme, Flexionen, Stoppwörter usw.) entfallen. Im Vergleich zu freien Schlagwörtern, die von den Autorinnen und Autoren selbst gewählt wurden, hat dieses Vokabular weitere Vorteile: (1) Da diese Schlagwörter von geschultem wissenschaftlichen PSYINDEX-Personal vergeben werden, ist zunächst sichergestellt, dass die tatsächlichen zentralen Inhalte der Publikationen adäquat und vollständig abgebildet werden. (2) Außerdem ist die Bedeutung jedes Schlagwortes im Thesaurus definiert. Dadurch wird eine unterschiedliche Verwendung gleicher Begriffe vermieden, so dass keine latente semantische Heterogenität innerhalb eines Themas entsteht. Beispielsweise unterscheidet der Thesaurus zwischen “Adaptation” (physiologische Prozesse) und “Adjustment” (psychologische, soziale oder emotionale Prozesse), was von Autorinnen und Autoren in ihrer eigenen Verschlagwortung womöglich unterschiedlich verstanden wurde. (3) Schließlich sind die standardisierten Schlagwörter für sämtliche PSYINDEX-Einträge verfügbar, ungeachtet der Sprache der Publikation. Da Themen anhand von Kookkurenzen (Blei et al., 2003), das heißt, ihres gemeinsamen Auftretens in Dokumenten, identifiziert werden, würde ein mehrsprachiges Korpus zu sprachlich getrennten Themen führen, selbst wenn die gleichen Inhalte betroffen sind. Als Nachteil eines standardisierten Vokabulars muss seine Endlichkeit genannt werden, wodurch feinere inhaltliche Nuancen nicht abgebildet werden können. Dies ist besonders bei thematisch eingegrenzten Korpora wichtig, etwa, wenn Thesaurus-Schlagwörter einen Themenbereich abstecken, der inhaltlich tiefer untersucht werden soll (siehe Studie 2 zum Bereich “Flucht und Migration”). Für ein Korpus aus der gesamten Psychologie kann die inhaltliche Bandbreite von über 6500 Thesaurusbegriffen jedoch als adäquat erachtet werden. In PsychArchives-ESM 1 zu Studie 3 (siehe Anhang D) wird dies noch ausführlicher dargelegt.

4.3.2 Topic Modeling und Themen-Validität (Schwerpunkte 1 und 2)

LDA mit Gibbs-Sampling wurde angewandt, wobei die Gibbs-Parameter und die LDA-Hyperparameter aus der Literatur abgeleitet wurden (Awati, 2015; Tang et al., 2014). Zur Bestimmung der Anzahl der Themen k wurde, wie von Kosinski et al. (2016) beschrieben, dasjenige Modell gewählt, dessen Held-Out-Likelihood am höchsten war (siehe Abschnitt 2.2.2). Zusätzlich wurden die Resultate (also die gefundenen Brennpunkthemen) der getesteten Modelle verglichen. Die semantische Validität der Themen (Maier et al., 2018)

wurde durch Vergleich der wahrscheinlichsten Wörter mit den repräsentativsten Dokumenten sichergestellt.

Um die Themen mit dem Klassifikationssystem vergleichen zu können (erster und zweiter methodischer Schwerpunkt), war folgende Überlegung der Ausgangspunkt: Wenn die Themen differenzierter als Klassifikationen sind, dann wurden zu dem Thema zugehörige Dokumente mit verschiedenen bzw. unterschiedlichen Klassifikationen indexiert. Dies ist beispielsweise beim Thema “Traumatisierung bei Geflüchteten” denkbar; dies wäre nur mit einer Kombination der Klassifikationskategorien “2910 Soziale Strukturen” (was laut Erläuterung im Thesaurus Migration einschließt) und “3215 Neurosen und Angststörungen” abbildbar. Allerdings wäre der spezifische Bezug zu Geflüchteten nicht ersichtlich: Eine Studie zur Prävalenz von Angststörungen bei verschiedenen sozialen Schichten würde ebenfalls durch diese beiden Kategorien abgebildet werden, hat jedoch offensichtlich nichts mit dem “Traumatisierung bei Geflüchteten” gemein. Um zu prüfen, inwiefern Themen und Klassifikationen übereinstimmen, wurden analog zu Griffiths und Steyvers (2004) die Dokumente zunächst nach ihrer Klassifikation gruppiert. Anschließend wurden die mittleren Dokument-Thema-Wahrscheinlichkeiten für jede Klassifikation berechnet. Verteilen sich die Dokument-Thema-Wahrscheinlichkeiten auf verschiedene Klassifikationen, so reicht eine Klassifikation nicht aus, um ein Thema abzubilden.

4.3.3 Trends (Schwerpunkt 3)

Dem Vorgehen von Griffiths und Steyvers (2004) folgend, wurden diejenigen Themen als “Hot Topics” aufgefasst, deren Dokument-Thema-Wahrscheinlichkeiten den größten Anstieg über die eingeschlossenen Publikationsjahre aufwiesen (dritter methodischer Schwerpunkt). Dazu wurden für jedes Thema lineare Regressionen mit Kriterium Prävalenz (über die mittlere Dokument-Thema-Wahrscheinlichkeit) und Prädiktor Publikationsjahr gerechnet. Für die Ermittlung der Brennpunktthemen anhand des Klassifikationssystems wurde entsprechend die relative Häufigkeit der Klassifikation an allen Publikationen je Jahr als Kriterium verwendet.

4.4 Ergebnisse

Ein Modell mit $k = 500$ Themen zeigte die beste statistische Modellgüte, wobei auch für andere geprüfte Werte von k die Brennpunktthemen den folgenden Bereichen zuzuordnen waren: Neuropsychologie, Online-Therapie, Migration, Traumatisierung und visuelle Aufmerksamkeit. Die

zehn Themen mit den stärksten Anstiegen zwischen 1980 und 2016 sind in Tabelle 9 für Topic Models und in Tabelle 10 für Klassifikationen aufgelistet. Bei Topic Modeling betreffen die Anstiege die mittlere Dokument-Thema-Wahrscheinlichkeit, bei den Klassifikationen die relative Häufigkeit an allen Publikationen des jeweiligen Jahres. Da es Ziel der Studie mehr, möglichst detaillierte Themen zu identifizieren und die damit einhergehende Anzahl an Themen relativ groß gewählt wurde, gibt es erwartungsgemäß inhaltliche Überschneidungen der Themen. Um die spezifischen Inhalte bzw. Facetten der Themen besser unterscheiden zu können, sind in Tabelle 9 die 15 wahrscheinlichsten Wörter enthalten. Bei der Interpretation der Regressionsgewichte im Topic Model ist zu beachten, dass in einem Modell mit $k = 500$ Themen die durchschnittliche Prävalenz $1/k = 1/500 = 0.2$ Prozent beträgt. Die Regressionsgewichte fallen entsprechend klein aus.

Tabelle 9. Die Top 10 der Brennpunktthemen im Topic Model.

Rang	Thema	Regressions- koeffizient b_1^*
1	364: Functional Magnetic Resonance Imaging, Cerebral Blood Flow, Prefrontal Cortex, Amygdala, Neuroanatomy, Biological Neural Networks, Cingulate Cortex, Brain, Oxygenation, Insula, Rewards, Striatum, Hippocampus, Brain Connectivity, Cognitive Control	0.0026
2	249: Functional Magnetic Resonance Imaging, Cerebral Blood Flow, Brain, Parietal Lobe, Prefrontal Cortex, Neuroanatomy, Frontal Lobe, Temporal Lobe, Oxygenation, Neuroimaging, Magnetic Resonance Imaging, Biological Neural Networks, Occipital Lobe, Visual Cortex, Spectroscopy	0.0020
3	386: Internet, Computer Mediated Communication, Online Therapy, Online Social Networks, Internet Usage, Electronic Communication, Communications Media, Websites, Social Media, Virtual Reality, Computer Assisted Therapy, Cellular Phones, Privacy, Telemedicine, Information Technology	0.0020
4	459: Genes, Polymorphism, Genetics, Serotonin, Genotypes, Dopamine, Alleles, Biological Markers, Phenotypes, Attention Deficit Disorder With Hyperactivity, Susceptibility (Disorders), Neurotransmission, Brain Derived Neurotrophic Factor, Neural Receptors, Tryptophan	0.0016
5	371: Cross-Cultural Differences, Human Migration, Cross-Cultural Communication, Cultural Sensitivity, Cross-Cultural Treatment, Multiculturalism, Expatriates, Transcultural Psychiatry, International Organizations, Cross-Cultural Counseling, Globalization, Multicultural Education, Foreign Workers, Acculturation, Racial And Ethnic Differences	0.0016

Tabelle 9 (Fortsetzung). Die Top 10 der Brennpunktthemen im Topic Model.

Rang	Thema	Regressions- koeffizient b_1^*
6	323: Magnetic Resonance Imaging, Brain, Neuroimaging, Neuroanatomy, Hippocampus, Gray Matter, Brain Size, Tomography, Prefrontal Cortex, White Matter, Amygdala, Cingulate Cortex, Cerebral Cortex, Temporal Lobe, Morphology	0.0016
7	86: Posttraumatic Stress Disorder, Emotional Trauma, Refugees, Trauma, War, Victimization, Torture, Persecution, Survivors, Violence, Injuries, Asylum Seeking, Exposure Therapy, Human Migration, Transgenerational Patterns	0.0015
8	344: Posttraumatic Stress Disorder, Emotional Trauma, Trauma, Eye Movement Desensitization Therapy, Stress Reactions, Intrusive Thoughts, Adjustment Disorders, Acute Stress Disorder, Traumatic Neurosis, Posttraumatic Growth, Complex PTSD, Exposure Therapy, Accidents, Medical Personnel, Metaphor	0.0014
9	365: Attention, Visual Attention, Selective Attention, Visual Search, Distraction, Cues, Reaction Time, Stimulus Parameters, Eye Movements, Attentional Capture, Visual Perception, Stimulus Salience, Visual Stimulation, Attentional Bias, Divided Attention	0.0014
10	95: Emotional Trauma, Posttraumatic Stress Disorder, Trauma, Dissociation, Dissociative Disorders, Early Experience, Dissociative Identity Disorder, Depersonalization, Borderline Personality Disorder, Neurobiology, Introjection, Dissociative Patterns, Amnesia, Psychodynamic Psychotherapy, Depersonalization/Derealization Disorder	0.0014

Anmerkungen: Die Regressionskoeffizienten entsprechen hier dem den durchschnittlichen Anstieg der Prävalenz in Prozent pro Jahr. * Alle $p < 0.0001$.

Tabelle 10. Die Top 10 der Brennpunktkategorien bei Verwendung des Klassifikationssystems (in höchstmöglicher Auflösung).

Rang	Kategorien	Regressionskoeffizient b_1^*
1	2520 Neuropsychology & Neurology	0.17
2	3215 Neuroses & Anxiety Disorders	0.13
3	3365 Promotion & Maintenance of Health & Wellness	0.10
4	3310 Psychotherapy & Psychotherapeutic Counseling	0.09
5	2530 Electrophysiology	0.09
6	2340 Cognitive Processes	0.07
7	3620 Personnel Management & Selection & Training	0.07
8	3311 Cognitive Therapy	0.06
9	3211 Affective Disorders	0.06
10	3660 Organizational Behavior	0.06

Anmerkung: Die Regressionskoeffizienten entsprechen hier dem den durchschnittlichen Anstieg der relativen Häufigkeit in Prozent pro Jahr. * Alle $p < 0.0001$.

Im Vergleich von Tabelle 9 und Tabelle 10 können drei Unterschiede zwischen den beiden Ansätzen festgestellt werden. Erstens weisen die Themen des Topic Models inhaltliche Überschneidungen auf. So lassen sich die Themen 364, 249 und 323 der Neuropsychologie zuordnen, die Themen 86, 344 und 95 der Bereich der Posttraumatischen Belastungsstörung. Allerdings betreffen diese inhaltlich ähnlichen Themen jeweils unterschiedliche Aspekte. So bezieht sich beispielsweise Thema 86 vor allem auf Flucht und Trauma, Thema 344 fokussiert auf Traumatherapie und Thema 95 konzentriert sich auf traumabezogene Störungen und Prozesse. Die Themen gehen also auf unterschiedliche Details ein. Dadurch, dass mehrere Themen zu einem gemeinsamen Bereich auf den vorderen Rängen erscheinen, sind diejenigen Themen, welche beim Klassifikationsansatz zusätzlich vorhanden sind (Tabelle 10), auf unteren Rängen zu finden (z. B. EEG auf Rang 11, Führungsverhalten auf Rang 19 oder Kognitiv-behaviorale Therapie auf Rang 20; in Tabelle 9 nicht eingeschlossen). Zweitens umfassen die Themen meist verschiedene Klassifikationskategorien. Leicht ersichtlich ist dies bei Thema 86, das einerseits Posttraumatische Belastungsstörung (“3215 Neurosen und Angststörungen”) adressiert, andererseits eben die besondere Patientengruppe von Flüchtlingen und Migranten (die laut Thesaurus entsprechende Kategorie “2910 Soziale

Strukturen” ist keine gesonderte “Brennpunktkategorie”). Damit einhergehend sind drittens die Kategorien zu unspezifisch. So ist etwa bei “3215 Neurosen und Angststörungen” nicht ablesbar, dass speziell die Posttraumatische Belastungsstörung und konkret bzgl. Migration/Flucht von steigendem Interesse in der psychologischen Fachliteratur ist.

Eine genauere Betrachtung hinsichtlich inhaltlicher Übereinstimmungen und Unterschiede ermöglicht der Levelplot in Abbildung 7³⁵. Abgebildet sind die mittleren Dokument-Thema-Wahrscheinlichkeiten der Brennpunktthemen aufgeschlüsselt nach den Klassifikationskategorien der Dokumente. Die Schattierungen repräsentieren somit die Verteilungen der Dokument-Thema-Wahrscheinlichkeiten der Themen über die Kategorien und vice versa. Bei Gleichverteilung haben alle Zellen einer Spalte die gleiche Farbe. Hebt sich eine einzelne Zelle von den anderen ab, deutet dies darauf hin, dass ein Zusammenhang zwischen der entsprechenden Klassifikationskategorie und dem entsprechenden Thema besteht. Die Einteilung der Schattierungsskala folgt der Spannweite der beobachteten Dokument-Thema-Wahrscheinlichkeiten (Min = 0.0017, Max = 0.0059). In Abbildung 7 zeigt sich eine grundsätzliche inhaltliche Entsprechung zugunsten eines höheren Detailgrades und einer größeren thematischen Vielschichtigkeit der Themen des Topic Models, denn keines der Brennpunktthemen stimmt mit einer einzigen Klassifikation überein. Stattdessen sind mehrere Klassifikationen nötig, um den Inhalt eines Themas abzubilden. Beispielsweise sind dies beim Thema “Traumatisierung bei Geflüchteten und Folter” (Thema 86) Kategorien aus den Bereichen “2900 Gesellschaftliche Systeme”, “3200 Psychische und physische Störungen”, “3300 Behandlung und Prävention”, “3800 Militärpsychologie” und “4200 Rechtspsychologie und Kriminologie”. Wie Tabelle 9 zeigt auch Abbildung 7, welche verschiedenen Aspekte inhaltlich verwandte Themen ansprechen. Leicht ist dies bei Thema 86 erkennen: Dokumente mit hohen Anteilen dieses Themas mit Bezug zur Posttraumatischen Belastungsstörung wurden häufig auch mit “2900 Gesellschaftliche Systeme” klassifiziert, was den Fokus auf Geflüchtete und Migranten im Vergleich zu den Themen 344 und 95 offenbart.

³⁵ Hinweis: Die entsprechende Abbildung in der Originalpublikation (Anhang A, Figure 2) weist eine Schattierungsskala mit einem zusätzlichen Bereich auf (0.006–0.007), für den jedoch keine Werte beobachtet wurden. Dieser Bereich wurde in Abbildung 7 entfernt; die Aussage der Abbildung bleibt dadurch unberührt.

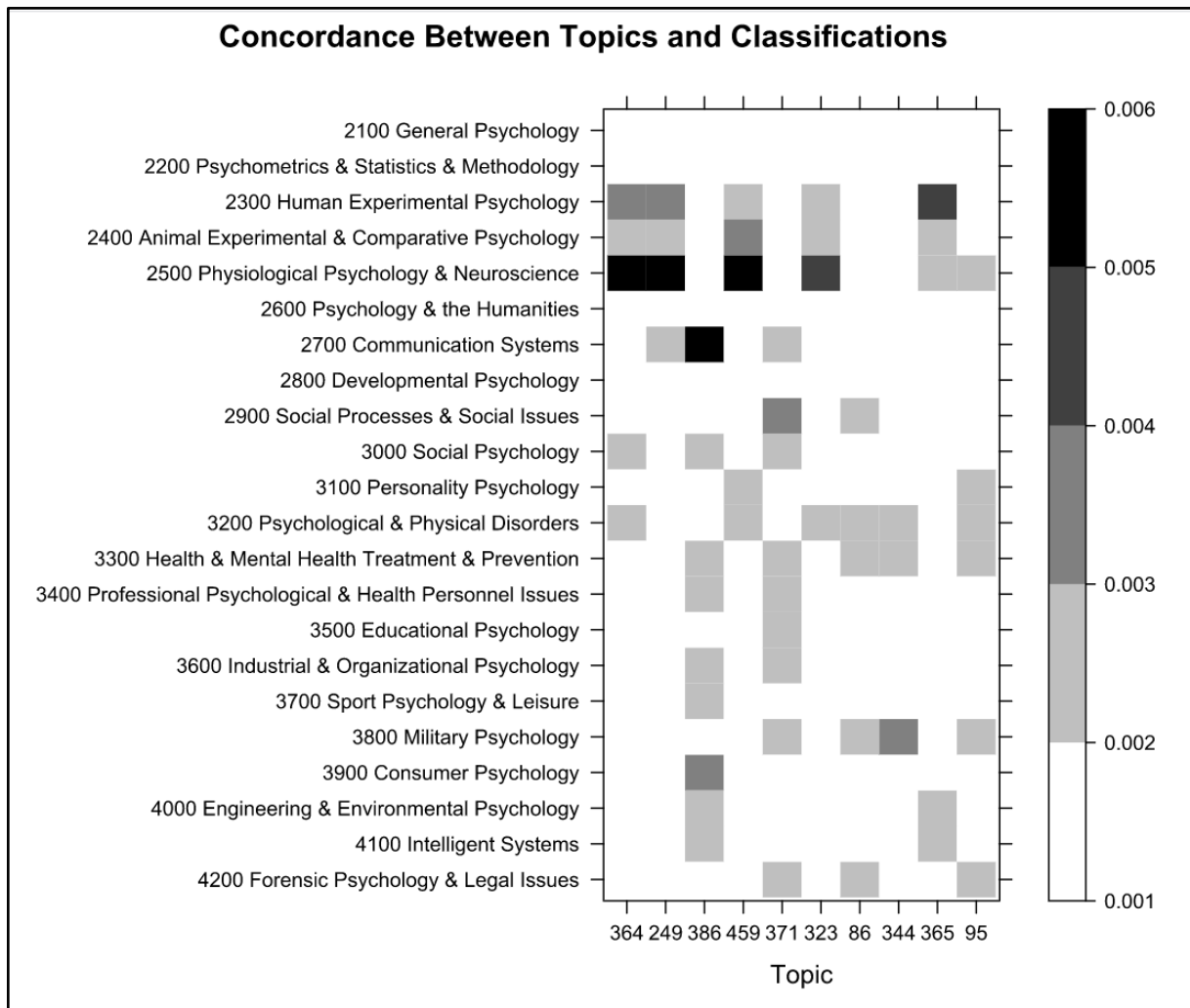


Abbildung 7. Die Zellen in diesem Levelplot entsprechen den mittleren Dokument-Thema-Wahrscheinlichkeiten der Themen getrennt nach Klassifikationskategorie (hier nur die Hauptkategorien; die Unterkategorien sind jeweils eingeschlossen). Je dunkler die Zelle, desto höher die Wahrscheinlichkeit. Z. B. haben Dokumente, die mit “2700 Kommunikationssysteme” klassifiziert wurden, die höchste Wahrscheinlichkeit für Thema 386 (Internet und Online-Therapie). Kein Thema stimmt mit nur einer Kategorie überein. Dies ist ein Hinweis darauf, dass die Themeninhalte nur mit mehreren Kategorien abgebildet werden können. Außerdem kann hier abgelesen werden, dass Themen aus ähnlichen Bereichen (z. B. 86 und 344) unterschiedliche Aspekte erfassen, da ihre Verteilung über die Kategorien nicht identisch ist.

4.5 Diskussion

Der Topic-Modeling-Ansatz zeigte sich dem Klassifikationsansatz zur Identifikation der Brennpunktthemen hinsichtlich Detailgrad und Themenvielfalt überlegen. Die Themen waren nur durch Kombination mehrerer Klassifikationen und dann auch nur annähernd abbildbar. Mit dem Klassifikationsansatz wäre Themen wie “Online-Therapie” (Thema 386), “interkulturelle Behandlung” (Thema 371) und “Traumatisierung von Geflüchteten” (Thema 86) unentdeckt geblieben. Überdies belegte die grundsätzliche thematische Übereinstimmung zwischen Klassifikationen und Themen sensu Griffiths und Steyvers (2004) die externe Validität der Themen, was in Verbindung mit den im Thesaurus definierten PSYNDEX Terms eine zuverlässige Interpretation der Themen ermöglichte. Das Klassifikationssystem ist somit zwar als Methode zur Identifikation von Themen schlechter geeignet als Topic Modeling, ermöglicht jedoch grundsätzlich ein externes Validieren der damit gefundenen Themen. Einschränkend muss konstatiert werden, dass dem Ansatz von Griffiths und Steyvers (2004) folgend die Übereinstimmung der Themen mit dem Klassifikationssystem nicht quantifiziert wurde. Denkbar wäre, mittels Kolmogorov-Smirnov-Anpassungstest die beobachtete Verteilung der Dokument-Thema-Wahrscheinlichkeiten eines Themas über alle Klassifikationskategorien (die Spalten in Abbildung 7) mit einer Gleichverteilung zu vergleichen. Damit könnte die Übereinstimmung der Themen mit den Klassifikationskategorien für jedes Thema und somit für das gesamte Topic Model numerisch abgebildet werden.

Die Anwendbarkeit des hier vorgestellten Ansatzes im Rahmen von Forschungssynthesen ergibt sich aus der Selektion von Primärstudien zum jeweiligen Thema. Mit dem LDA-Ansatz dieser Studie ist die Auswahl entsprechender empirischer Publikation Primärliteratur nur post-hoc möglich. Das heißt, zu einem gewünschten Thema werden die repräsentativsten Publikationen ausgegeben und über eine Datenbank-Suchoberfläche (zum Beispiel www.PubPsych.eu) nach Studienmethodik gefiltert. Das Erkennen, welche Themen besonders viel empirische Evidenz aufweisen oder wo es Forschungslücken und -desiderate gibt, ist also erst über diesen Umweg möglich. Damit ergibt sich als Implikation für Folgestudien, die Studienmethodik einer Publikation (im Sinne von empirisch versus nicht-empirisch), direkt mitzumodellieren (siehe Studie 2, Abschnitt 5).

Methodische Einschränkungen ergaben sich zunächst durch die Begrenztheit eines standardisierten Vokabulars. Der APA-Thesaurus wird zwar regelmäßig aktualisiert und

erweitert (so wurden 2019 über 300 neue Begriffe aufgenommen, unter anderem “Big Data” und “Unsupervised Learning”), allerdings mit einer zeitlichen Verzögerung: Das Sonderheft der *Psychological Methods* zu “Big Data in Psychology” beispielsweise erschien bereits Ende 2016 (Volume 21, Issue 4) und weitere Publikationen noch früher. Daher lässt sich schlussfolgern, dass über das standardisierte Vokabular zwar Forschungsthemen der gesamten Psychologie gut abgebildet werden können, gerade entstehende Brennpunktthemen jedoch übersehen werden könnten. Die Verwendung von Freitexten aus Abstracts oder Volltexten würde direkt eine geeignete Option darstellen, obgleich Schwierigkeiten in der Datenaufbereitung resultieren würden (allen voran die unterschiedlichen Sprachen der Publikationen). Da Fachliteratur bedingt durch den Publikationsprozess eine Latenzzeit von etwa ein bis zwei Jahren aufweist (Krampen, Montada, Müller & Schui, 2004), ist darüber hinaus die Überlegung naheliegend, Diskussionen und Kommentare von Fachleuten und -gesellschaften in sozialen Medien (wie etwa auf Twitter) einzubeziehen, um besonders aktuelle Themen automatisiert zu identifizieren (siehe Abschnitt 7.5.2).

Eine zweite Einschränkung betrifft den Zeitraum zur Bestimmung der Steigung der Regressionsgeraden. Entsprechend Griffiths und Steyvers (2004) wurde der gesamte Zeitraum an Publikationsjahren (1980–2016) als Prädiktor verwendet. Denkbar wäre es, einen kürzeren Zeitraum festzulegen, damit Themen mit deutlicherem Anstieg in den letzten Jahren mehr Gewicht erhalten. Idealerweise könnte die Entscheidung, welcher Zeitraum einbezogen werden soll, vor dem Hintergrund einer konkreten Fragestellung individuell getroffen werden (z. B. “nur die letzten fünf Jahre” oder “nur die 1990er Jahre”; siehe Abschnitt 7).

Als weitere Einschränkungen ist zu nennen, dass die Reliabilität der Brennpunktthemen hier nur stichprobenhaft durch qualitative Inspektion geprüft wurde, nicht anhand statistischer Kennwerte. Zwar zeigten sich die identifizierten Themenbereiche auch bei Modellen mit anderen Themenanzahlen k , jedoch würde die Übereinstimmung nicht quantifiziert. Auch wurden die LDA-Hyperparameter für alle Werte von k konstant gehalten, wodurch nicht ausgeschlossen werden kann, dass Modelle mit anderem k und anderen Hyperparametern einen besseren statistischen Fit erzielt hätten. Die Tatsache, dass die gefundenen Brennpunktthemen (die im Fokus dieser Studie standen), zumindest bei manueller Inspektion modellübergreifend gefunden werden konnten, weist allerdings auf die Robustheit der gefundenen Ergebnisse hin. Nichtsdestotrotz lässt sich daraus ableiten, in zukünftigen LDA-

Anwendungen den Aspekten der Reliabilität und Hyperparameter-Optimierung besondere Aufmerksamkeit zu widmen (siehe Studie 3, Abschnitt 6).

5. Studie 2: “Ist die psychologische Forschung durchlässig für aktuelle gesellschaftliche Themen?”

Bittermann, A. & Klos, E. M. (2019). Ist die psychologische Forschung durchlässig für aktuelle gesellschaftliche Themen? Eine szientometrische Analyse am Beispiel Flucht und Migration mithilfe von Topic Modeling. *Psychologische Rundschau*, 70(4), 239–249.
<https://doi.org/10.1026/0033-3042/a000426>

5.1 Hintergrund

Die Beziehung zwischen gesellschaftlichen Herausforderungen und entsprechenden Beiträgen der Wissenschaft ist von fachpolitischer Relevanz. So löste die Forderung des Wissenschaftsrates (2018) bzgl. einer stärkeren gesellschaftlichen Verantwortungsübernahme der Psychologie, auch und gerade um zur Bewältigung der sogenannten “Großen gesellschaftlichen Herausforderungen” (Wissenschaftsrat, 2015) beizutragen, eine Reihe von Stellungnahmen aus (Antoni, 2019). Anhand einer szientometrischen Untersuchung von PSYNDEX-Daten war es daher angebracht, eine empirische Antwort auf die Frage zu geben, ob sich aktuell gesellschaftlich relevante Themen in psychologischen Fachpublikationen widerspiegeln.

Für einen ersten Überblick wurden Publikationsstatistiken zu den Themen “Klimawandel”, “Flucht und Migration”, “demografischer Wandel”, “Veränderungen der Arbeitswelt im digitalen Zeitalter”, “Mensch-Maschine-Interaktion” und “künstliche Intelligenz” angefertigt und dem Präsidenten der DGPs zur Veröffentlichung zur Verfügung gestellt (Antoni, 2019). Die vorliegende Studie ging darüber hinaus, indem exemplarisch für “Flucht und Migration” genauer untersucht wurde, welche konkreten Inhalte diese Publikationen haben, wieviel empirische Evidenz zu den jeweiligen Themen vorliegt, und aus welche fachlichen Perspektiven konkrete Herausforderungen angegangen wurden. Beispielsweise ist anzunehmen, dass sich eine erfolgreiche Integration von unbegleiteten minderjährigen Geflüchteten nicht auf die Behandlung emotionaler Traumata (Domäne der Klinischen Psychologie) beschränken lässt, sondern hinsichtlich der Sprach- und Identitätsentwicklung auch die Entwicklungspsychologie und die Pädagogische Psychologie betrifft. Für den Zeitraum von 1980 bis 2017 wurden im Bereich Flucht und Migration Themen erwartet, die in Zusammenhang mit den innerbundesrepublikanischen Migrationsbewegungen nach der

deutschen Wiedervereinigung, den fremdenfeindlichen Übergriffen in den 1990er Jahren, sowie den starken Migrationsbewegungen nach Deutschland 2015 und 2016 stehen. Auf eine gezielte Suche nach diesen Themen wurde jedoch verzichtet, um nicht weitere potenzielle Themen zu übersehen. Für ein Explorieren der thematischen Vielfalt wurde daher Topic Modeling angewandt.

5.2 Fragestellungen

- (1) Ist die psychologische Forschung durchlässig für jeweils aktuelle gesellschaftliche und politische Fragestellungen?
- (2) Wird sie dem Anspruch gerecht, eine evidenzbasierte Grundlage für diese Fragestellungen zu schaffen?
- (3) Welchen psychologischen Teildisziplinen können die Forschungsthemen zugeordnet werden?

5.3 Methode

Die methodischen Schwerpunkte dieser Studie betreffen (1) die Modellierung einer Dokumentvariable, (2) eine verbesserte Modellwahl, (3) weitere Möglichkeiten der externen Validierung sowie (4) die Analyse von Publikationsspitzen.

5.3.1 Daten

Das Korpus war hier auf den Bereich “Flucht und Migration” beschränkt, definiert durch eine PSYNDEX-Recherche von Publikationen, die mit mindestens einem der standardisierten Schlagwörter “Refugees”, “Human Migration”, “Immigration”, “Asylum Seeking”, “Political Asylum”, “Expatriates”, “Foreign Workers” und “Migrant Farm Workers” indexiert waren. Die Auswahl dieser Begriffe erfolgte entsprechend der Erläuterungen der PSYNDEX Terms (ZPID, 2016). Die Validität der Korpusauswahl hängt davon ab, ob mit den Begriffen auch tatsächlich Flucht und Migration abgebildet werden kann. Schaal et al. (2019) weisen in diesem Zusammenhang auf zwei mögliche Quellen für spätere Fehlinterpretationen hin: “False Positives” im Sinne einer Verwendung von Begriffen in anderer Bedeutungsabsicht sowie implizite Thematisierungen im Sinne einer Behandlung des Themas mit anderen als den verwendeten Suchbegriffen. Im Gegensatz zu freien Begriffen ist das gesamte standardisierte Schlagwortvokabular von PSYNDEX in einem Thesaurus definiert (ZPID, 2016), wodurch “False Positives” vermieden werden. Außerdem werden diese

standardisierten Schlagworte durch das wissenschaftliche Personal von PSYNDEX den Publikationen zugeordnet, womit sichergestellt wird, dass auch implizite Thematisierungen in Freitexten abgedeckt sind (etwa wann “Migration” nicht explizit als Begriff im Abstract auftaucht, jedoch aus dem Kontext erkennbar ist). Da im Gegensatz zu Studie 1 das Korpus durch standardisierte Schlagwörter bereits eingegrenzt war, musste aus einer inhaltlich tieferen Ebene geschöpft werden, um darüberhinausgehende Informationen extrahieren zu können. Andernfalls hätten die Themen hauptsächlich aus den oben genannten Suchbegriffen bestanden, was den möglichen Erkenntnisgewinn stark eingeschränkt hätte. Eingeschlossen wurden daher $n = 4073$ deutschsprachige Abstracts der Jahre 1980 bis 2017. Die Beschränkung auf eine einheitliche Sprache erfolgte angesichts der Problematik mehrsprachiger Texte (siehe Abschnitt 2.3). Der Anteil ausgeschlossener englischsprachiger Abstracts betrug rund 15 %.

5.3.2 Aufbereitung der Textdaten

Für den Ausschluss von Wörtern, die zwar häufig miteinander vorkommen, aber inhaltlich wenig Beitrag leisten (wie etwa “Studie”, “untersucht”, “Ergebnis”), wurde eine Liste deutschsprachiger Abstracts wissenschaftlicher Publikationen erstellt (siehe ESM 1 zu Studie 2³⁶). Diese Liste wurde in einem iterativen Prozess aus der Identifikation besonders häufiger Wörter im Korpus und der Analyse der Themen vorläufiger Modelle erstellt. Zusätzlich wurden PSYNDEX-Mitarbeitende konsultiert, die in der Bearbeitung von Abstracts versiert sind. Auf ein sogenanntes “Stemming”, also das Kürzen von Wörtern auf ihren Wortstamm wurde aus verschiedenen Gründen verzichtet: Aus inhaltlicher Sicht betrifft dies den Erhalt wichtiger Differenzierungen. So weisen etwa die Wörter „Flüchtlinge“ und „Geflüchtete“ unterschiedliche Konnotationen auf (Hildebrand, 2015). Aus methodischer Sicht sind die Befunde von Schofield und Mimno (2016) zu berücksichtigen, nach denen Stemming negativen Einfluss auf Held-Out-Likelihood und Topic-Stabilität³⁷ haben kann.

5.3.3 Topic Modeling (Schwerpunkte 1 und 2)

Neben der Identifikation der Themen wurde in dieser Studie untersucht, wieviel empirische Evidenz zu den einzelnen Themen vorliegt. Daher war notwendig, eine Dokumentvariable in das Topic Model aufzunehmen (erster methodischer Schwerpunkt). Dies ist im Gegensatz zu

³⁶ <http://dx.doi.org/10.23668/psycharchives.2499>

³⁷ Bei LDA betrifft dies die Topic-Reliabilität, also die Reproduzierbarkeit der Themen bei wiederholter Modellinferenz (Studie 3).

LDA bei der Entwicklung des STM explizit berücksichtigt worden. Daher wurde der Einfluss des dichotomen Merkmals “Studienmethodik” mit den Ausprägungen “empirisch” und “nicht-empirisch” (entsprechend der Dokument-Metadaten in PSYINDEX) auf die Themenprävalenz untersucht.

Die STM-Modellinitialisierung ist zwar deterministisch (siehe Abschnitt 2.2.4), wodurch für eine festgelegte Anzahl an Themen k auch bei wiederholter Modellinferenz (mehrmalige Algorithmusdurchläufe zur Generierung der Themen) identische Ergebnisse resultieren³⁸. Für die *Inferenz der Themen* entfällt also das Setzen eines random seeds (was bei LDA zwecks Reproduzierbarkeit zwingend notwendig ist, siehe Abschnitt 2.2.3). Allerdings betrifft dies nicht die *Bestimmung von k* , also der optimalen Anzahl an Themen, wenn sie über die Held-Out-Likelihood (HL)-Methode (Kosinski et al., 2016) ausgemacht wird: Trainings- und Testdatensatz werden zufällig festgelegt. Für eine Reproduzierbarkeit dieser zufälligen Aufteilung ist hier, wie bei vielen Algorithmen aus dem Maschinellen Lernen, ein random seed nötig. Zusätzlich kann durch Verwendung mehrerer seeds ausgeschlossen werden, dass der gefundene Optimalwert für k nur für eine zufällige, nicht-repräsentative Aufteilung gültig ist. Dieses Vorgehen ist analog zu mehreren Modellinitialisierungen bei LDA (mit dem Gibbs-Parameter “random starts”; Awati, 2015).

Die Modellwahl (hinsichtlich k) wurde daher wie folgt verbessert (zweiter methodischer Schwerpunkt): Zunächst wurde der Wertebereich einer möglichen Themenanzahl festgelegt. In Studie 1 wurde dieser aus der Literatur (Griffiths & Steyvers, 2004) abgeleitet, da die Korpora eine vergleichbare inhaltliche Breite aufwiesen. Hier wurde basierend auf stichprobenhaften Testmodellen³⁹ ein Bereich von fünf bis 40 Themen überprüft. Das Ergebnis ist zunächst ein HL-Wert für jedes k (wie in Studie 1). Da allerdings aufgrund der zufälligen Aufteilung ist Test- und Trainingsdatensatz nicht ausgeschlossen werden kann, dass dieses Ergebnis suboptimal ist, wurde die Bildung dieser beiden Datensätze 100-mal wiederholt (mit 100 verschiedenen seeds). Nach jeder Wiederholung wurde der Wert von k mit der höchsten HL (Modalwert) gespeichert. Die Verteilung des optimalen k ist in Abbildung 8 dargestellt. Am häufigsten war $k = 19$ der Wert mit der höchsten HL, mit Abstand gefolgt von $k = 18$ und $k = 22$ (jeweils bei zehn der 100 Wiederholungen der beste Wert). Da 18 sehr nahe bei 19 liegt, wurde der nächsthäufigste Wert von $k < 19$, also $k = 16$, zusammen mit $k = 19$ und $k = 22$ einer manuellen Überprüfung (Maier et al., 2018)

³⁸ Daher ist die Reliabilität der Themen bei LDA relevant (Studie 3), nicht bei STM.

³⁹ Bei diesen Testmodellen war k sehr klein (2) bis sehr groß (100) sensu Banks et al. (2018). Nach Inspektion der jeweiligen Held-Out-Likelihood-Werte konnte der Bereich auf $k \in [5;40]$ eingegrenzt werden.

unterzogen. Dabei galt es festzustellen, ob die Inhalte der Themen zu fein oder so grob aufgelöst waren. Eine Anzahl von $k = 19$ Themen wurde als geeigneter Kompromiss empfunden, da die Themen einerseits differenziert genug waren, andererseits nicht so spezifisch, dass nur wenige Dokumente einem solchen Thema entsprochen hätten.

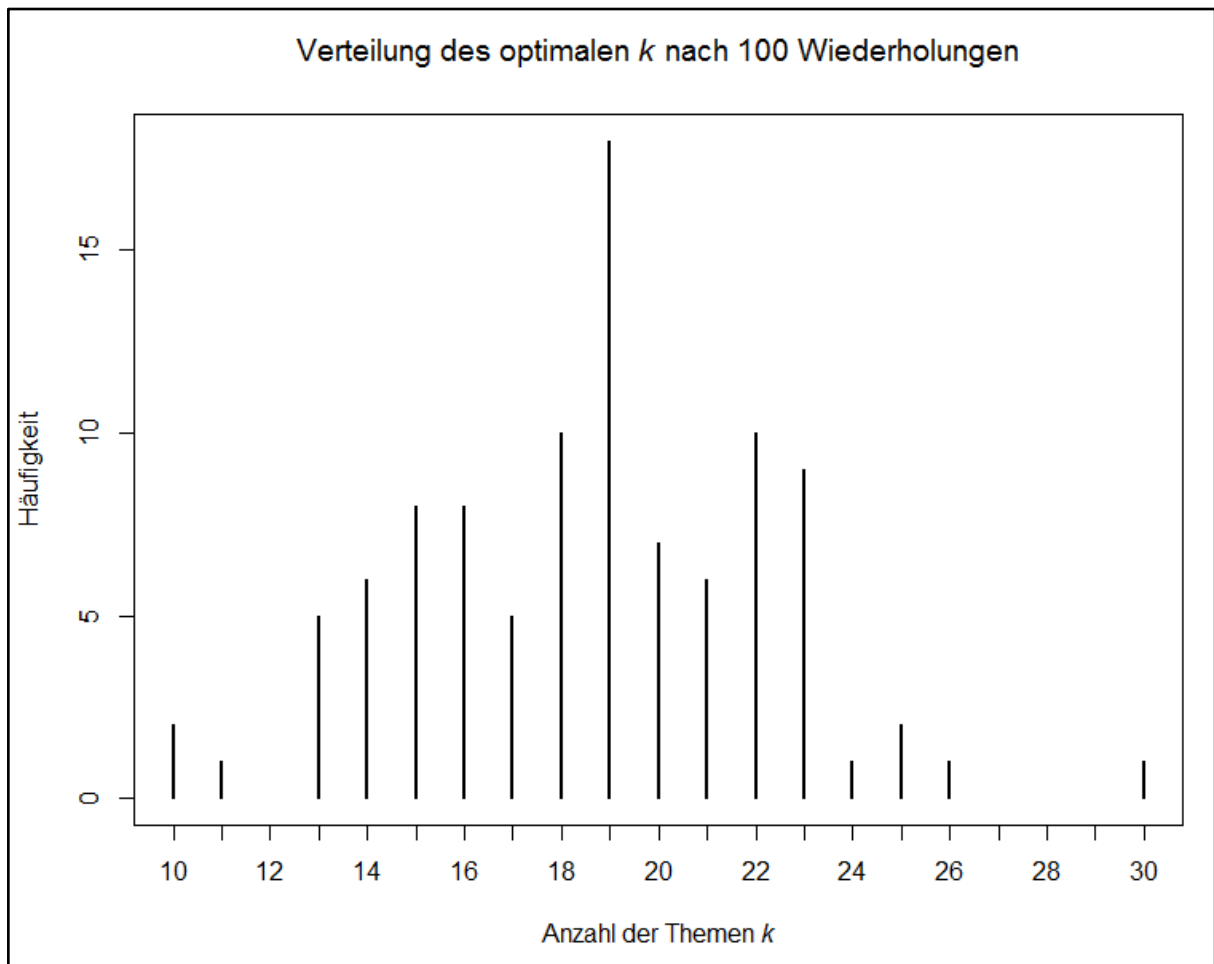


Abbildung 8. Verteilung des optimalen Wertes für k nach 100 Wiederholungen. Ausgewählt wurde, wie oft ein Wert von k die höchste Held-Out-Likelihood hatte. Insgesamt erwies sich $k = 19$ bei 18 der 100 Wiederholungen als der beste Wert.

5.3.4 Themen-Validität und Publikationsspitzen (Schwerpunkte 3 und 4)

Die Themen in Studie 1 wurden auf Grundlage eines standardisierten Vokabulars in PSYNDEX gebildet. Diese Begriffe sind im Thesaurus definiert, wodurch es möglich ist, die beim Inspizieren der Themen vermuteten Inhalte zu verifizieren. In dieser Studie wurden jedoch Freitexte (Abstracts) verwendet. Daher war es sinnvoll, zusätzliche Kriterien der externen Themvalidierung anzuwenden bzw. deren Eignung für Freitexte zu überprüfen (dritter methodischer Schwerpunkt). Konkret wurden zwei Ansätze verfolgt: (1) die Korrespondenz von thematisch relevanten Ereignissen mit Publikationsverläufen (z. B. Evans, 2014) und (2) Übereinstimmung mit einem Code-System (z. B. Jacobi et al., 2015). Ersteres wurde durch die inhaltliche Analyse der Thementrends im historischen Kontext umgesetzt. Durch den Vergleich von Publikationsspitzen (vierter methodischer Schwerpunkt) mit gesellschaftlich Ereignissen soll nachgewiesen werden, dass die psychologische Fachliteratur Themen der Gesellschaft aufgreift. Dies geschieht vor dem Hintergrund der durch den Publikationsprozess bedingten Verzögerung von ein bis zwei Jahren (Krampen et al., 2004) und mittels Sichtung der Abstracts der repräsentativsten Studien je Thema. Der zweite Validierungsansatz entspricht dem Vorgehen in Studie 1: Analog zu Griffiths und Steyvers (2004) wurde die Korrespondenz der Themen mit dem APA-Klassifikationssystem (Tuleya, 2007; ZPID, 2016) untersucht.

5.4 Ergebnisse

Die erwarteten Themen für die Jahre 1980–2017 (bzgl. der deutschen Wiedervereinigung, Traumatisierung bei Geflüchteten, Fremdenfeindlichkeit) konnten mithilfe von Topic Modeling bestätigt werden. Bei diesen Themen konnten aus geschichtswissenschaftlicher Perspektive Publikationsspitzen mit gesellschaftlich relevanten Ereignissen in Verbindung gebracht werden. Abbildung 9 zeigt dies am Beispiel des Themas zu Traumatisierung bei Geflüchteten. Darüber hinaus konnten konkrete Ereignisse auch den Publikationsspitzen der Themen zu Projekten mit Geflüchteten sowie zu Schülerinnen und Schülern mit Migrationshintergrund zugeordnet werden (siehe ausführlicher in der Originalpublikation, Anhang B).

Für 18 der 19 Themen konnten signifikante Unterschiede in der erwarteten Differenz der mittleren Dokument-Thema-Wahrscheinlichkeit zwischen empirischen bzw. nicht-empirischen Studien gefunden werden (Abbildung 10). Dabei wies das Thema

“Sprachentwicklung von Migrant*innenkindern” (Thema 8) die höchste Wahrscheinlichkeit auf, von empirischen Studien behandelt zu werden, das Thema 17 zur Identitätsentwicklung die höchste Wahrscheinlichkeit von nicht-empirischen Beiträgen. Die 18 Themen mit signifikanten Unterschieden verteilen sich zu gleichen Anteilen auf die Ausprägungen “empirisch” bzw. “nicht-empirisch”.

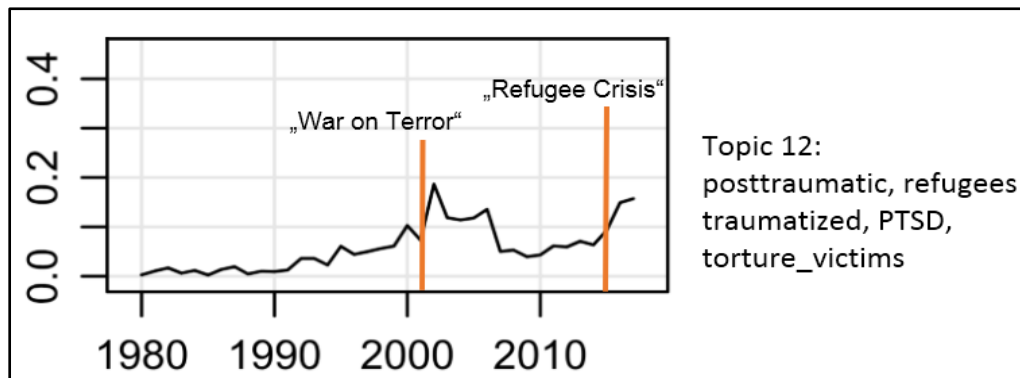


Abbildung 9. Zeitlicher Verlauf der Prävalenz des Themas zu Traumatisierung bei Geflüchteten. Markiert sind zwei relevante zeitgeschichtliche Ereignisse.

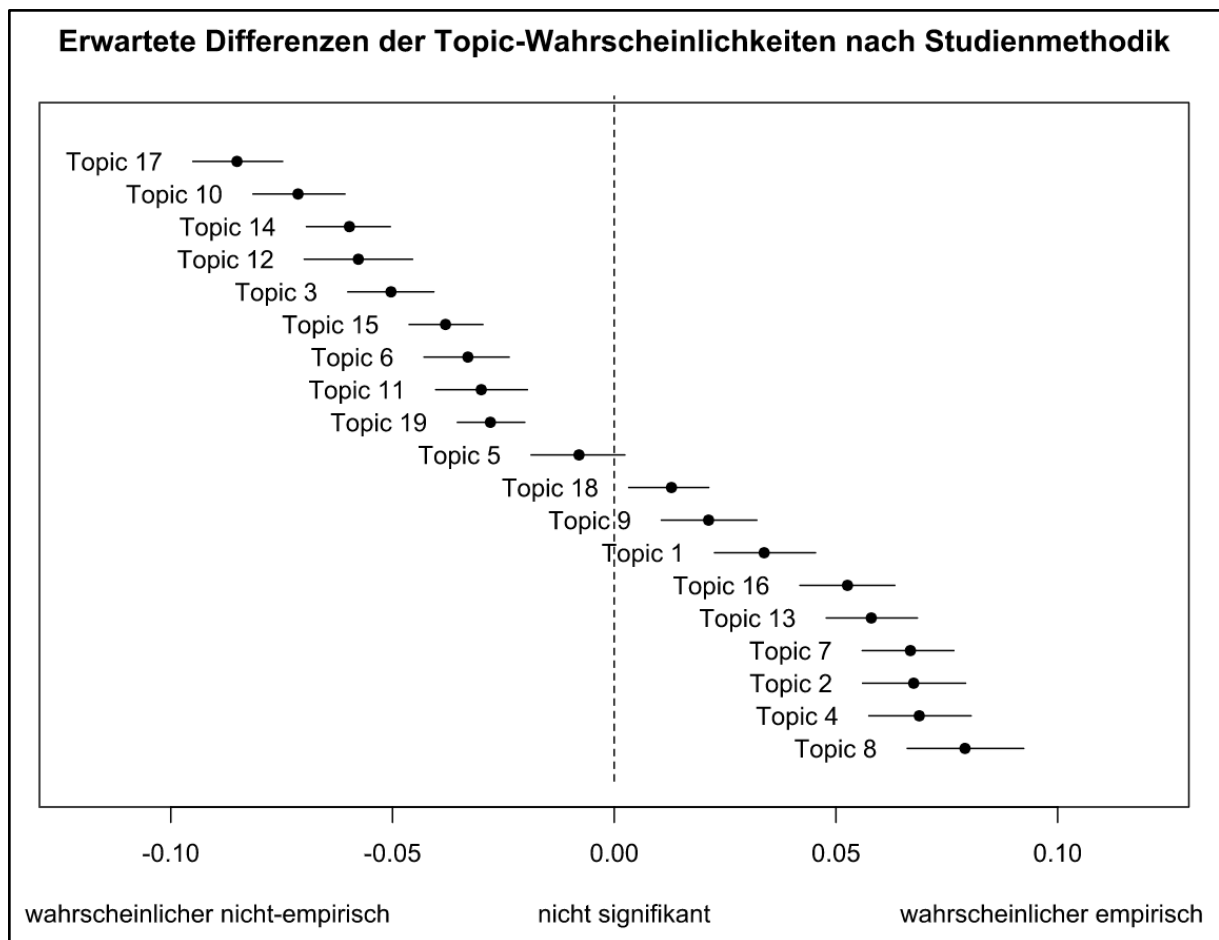


Abbildung 10. Regressionsbasierte erwartete Prävalenzdifferenzen nach Studienmethodik (einschließlich 95%-Konfidenzintervalle). Themen der linken Seite werden mit höherer Wahrscheinlichkeit von nicht-empirischen Beiträgen behandelt, Themen der rechten Seite von empirischen Studien. Die Themen werden in der Originalpublikation in Anhang B beschrieben.

Aus Abbildung 11 lässt sich zeilenweise ablesen, wie wahrscheinlich die Themen für die jeweiligen Kategorien bzw. Disziplinen der Psychologie sind. Die Schattierung der Zellen steht für die Höhe der Dokument-Thema-Wahrscheinlichkeit. Für die Beantwortung der dritten Fragestellung bzgl. der Zuordnung der psychologischen Teildisziplinen zu den Forschungsthemen wurde als Schwellenwert die doppelte durchschnittliche Prävalenz ($2 \cdot 1/k = 2 \cdot 1/19 = 0.105$) festgelegt. Damit wurden nur die deutlichsten Zuordnungen erfasst, während interdisziplinäre Themen nicht berücksichtigt wurden. Dazu kommt, dass dieses Kriterium mit geringerer Wahrscheinlichkeit von Themen erfüllt wird, die eine insgesamt niedrige Prävalenz haben⁴⁰. Dies war insofern gewünscht, als dass nur solche Themen zu den

⁴⁰ Hat ein Thema etwa eine Gesamtprävalenz von 0.01, dann könnte eine von zehn Kategorien nur dann eine Prävalenz von 0.1 aufweisen (und damit im Bereich des Schwellenwertes liegen), wenn die restlichen neun Kategorien eine Prävalenz von Null hätten. Dies ist in dieser Studie jedoch insofern auszuschließen, als dass die

psychologischen Teildisziplinen gewertet werden sollten, zu denen auch eine substanzielle Anzahl an Publikationen vorliegt. Die entsprechende Abbildung in der Originalpublikation (Anhang B, Abbildung 3) weist daher eine Schattierungsskala mit gröberer Einteilung auf (der Bereich von 0 bis 0.1 ist dort zusammengefasst). Hierbei zeigte sich, dass die meisten Themen der Klinischen Psychologie zugeordnet werden konnten (Kategorien 33** und 32**), gefolgt von der Pädagogischen Psychologie (35**) und der Entwicklungspsychologie (28**).

Zum Zwecke der externen Validierung der Themen ist die Skala in Abbildung 11 feiner unterteilt, sodass auch tendenzielle Zuordnungen sichtbar werden. Hier lässt sich feststellen, dass sich die für jedes Thema diejenigen Kategorien hervorheben, die aus den Themeninhalten zu erwarten waren (eine Liste aller Themen kann der Originalpublikation in Anhang B entnommen werden). Beispielsweise betrifft Thema 19 politische Kommunikation und Menschenrechte, entsprechend heben sich die Kategorien “27** Kommunikationssysteme” und “42** Rechtspsychologie und Kriminologie” (einschließlich der Subkategorie “4210 Zivil und Menschenrechte”) hervor. Allerdings gilt dies auch für die Kategorie “37** Sportpsychologie & Freizeitverhalten”. Diese Kategorie hat bei Thema 13 zur deutschen Wiedervereinigung die höchste Wahrscheinlichkeit und ist auch anderen Themen zugeordnet. Da dies weder aus den wahrscheinlichsten Begriffen von Thema 13 (“akkulturation, soziale, deutschen, sozialen, immigranten, identität, integration, identifikation, assimilation, republik”), noch aus den für Thema 13 häufigen und exklusiven Wörtern (“republik, demokratischen, ost, akkulturation, westdeutschland, übersiedlern, ddr, immigranten, übersiedler, identifikation”) zu erwarten war, wurden die repräsentativsten Dokumente⁴¹ einer genaueren Prüfung unterzogen. Hier zeigte sich, dass unter anderem Publikationen zu den psychischen Effekten des Sporttreibens bei Übersiedlern aus Ost- nach Westdeutschland (Fuchs & Appel, 1994) oder zur sportlichen Aktivität im Rahmen der Bewältigung von kritischen Lebensereignissen bei Übersiedlern (Kleine & Fuchs, 1994) enthalten waren. Analoges gilt für die anderen Themen mit Bezug zur Sportpsychologie.

Kategorie (29** Gesellschaftliche Systeme) explizit die Folgen von Migration umfasst und daher stets eine gewisse “Grundprävalenz” aufweist.

⁴¹ Eine Auflistung der jeweils zehn repräsentativsten Dokumente je Thema kann dem ESM zu Studie 2 zu entnommen werden: <http://dx.doi.org/10.23668/psycharchives.2498>.

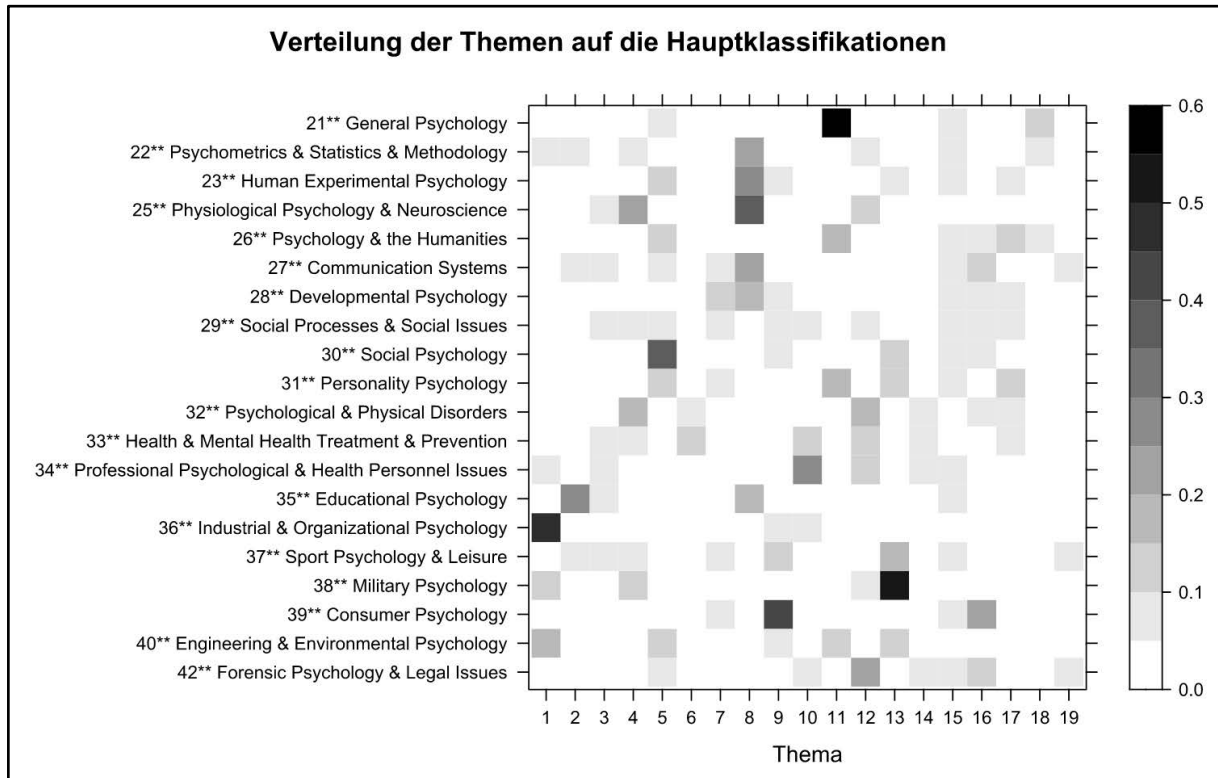


Abbildung 11. Verteilung der Themen auf die Hauptklassifikationen. Je dunkler die Zelle, desto höher ist die mittlere Dokument-Thema-Wahrscheinlichkeit. Beispielsweise hatten Dokumente, welche mit „36** Arbeits- und Organisationspsychologie“ klassifiziert waren, die höchste Wahrscheinlichkeit, Thema 1 zu behandeln. Die Farbschattierungen decken in gleichen Abständen den beobachteten Bereich von Min = 0.003 bis Max = 0.57 ab. Die Themen werden in der Originalpublikation in Anhang B beschrieben. Mit “**“ wird indiziert, dass Subkategorien eingeschlossen sind.

5.5 Diskussion

Neben der inhaltlichen Bandbreite der Themen konnte auch anhand der zeitlichen Korrespondenz mit gesellschaftlichen Ereignissen und Prozessen gezeigt werden, dass die psychologische Fachliteratur die “Große gesellschaftliche Herausforderung” (Wissenschaftsrat, 2015) “Flucht und Migration” aufgreift. Die zwei Brennpunktthemen aus Studie 1 mit Bezug zu diesem Themenbereich konnten erneut identifiziert werden. Dabei wurde das Brennpunktthema zu Migration in weitere Themen⁴² unterteilt, während das Thema “Flucht, Trauma und Folter” in nahezu gleicher Form (im Sinne identischer Begriffe und thematischer Homogenität) gefunden werden konnte. Zur Forschungsfrage 1 kann damit

⁴² “Themen” sind hier im methodischen Sinne sensu Topic Modeling gemeint. Inhaltlich könnte man hier auch von “Facetten” oder “Aspekten” sprechen.

konstatiert werden, dass die psychologische Forschung durchlässig (im Sinne inhaltlicher Kongruenz und zeitlicher Korrespondenz) für Themen der Gesellschaft ist.

Die externe Validierung anhand der zeitlichen Kongruenz von Publikationsspitzen mit Ereignissen ist naturgemäß nur bei solchen Themen möglich, denen ein Ereignis überhaupt zugeordnet werden kann. Dies war in dieser Studie bei fünf Themen möglich: Deutsche Wiedervereinigung, Fremdenfeindlichkeit, traumatisierte Geflüchtete, Projekte mit Geflüchteten sowie Schülerinnen und Schüler mit Migrationshintergrund. Die zeitlichen Verläufe dieser Themen konnten mit konkreten Ereignissen in Verbindung gebracht werden. Eine solche zeitliche Korrespondenz zur externen Validierung wurde in dieser Studie von der Koautorin und ihrem fachlichen Hintergrund als Geschichtswissenschaftlerin bestimmt. Zusätzlich wäre eine Quantifizierung im Sinne einer Abweichung des beobachteten vom erwarteten Verlauf des zeitlichen Trends denkbar. Daher wäre in zukünftigen Studien zu prüfen, ob beispielsweise Zeitreihen hierfür geeignet wären (wie z. B. bei Krampen, von Eye & Schui, 2011; vgl. Abschnitt 7.3.2).

Zu Forschungsfrage 2, ob die Psychologie dem Anspruch gerecht wird, eine evidenzbasierte Grundlage für Fragen der Gesellschaft zu schaffen, lässt sich ein gemischtes Bild feststellen: Nur bei neun der 19 Themen (47.4 %) war die Wahrscheinlichkeit empirischer Studien höher. Für eine inhaltliche Diskussion dieses Befundes sei auf die Originalpublikation in Anhang B verwiesen. Aus methodischer Sicht ist hier festzuhalten, dass anhand der Dokumentvariable "Studienmethodik" aufgezeigt werden konnte, welche Themen viel bzw. wenig empirische Evidenz aufweisen. Damit ist die Einschränkung von Studie 1 eines Umwegs über post-hoc Filterung nach empirischen Studien überwunden worden. Außerdem ist aus der Angabe von Prävalenzdifferenzen der Anteil empirischer Studien ableitbar. Für eine konkrete Bestimmung des Anteils müssten aus der Bayes'schen A-posteriori-Wahrscheinlichkeit frequentistische Aussagen abgeleitet werden. Das bedeutet, es müsste ein geeigneter Schwellenwert für die Dokument-Thema-Wahrscheinlichkeit für die Bestimmung von Anteilen definiert werden (siehe PsychArchives-ESM 1 zu Studie 3 in Anhang D; vgl. auch Abschnitt 7.5.1).

Hinsichtlich Forschungsfrage 3, der Zuordnung von psychologischen Teildisziplinen zu den Themen, zeigte sich eine Dominanz der Klinischen Psychologie, jedoch waren auch andere Disziplinen vertreten. Hierfür wurde analog zu Studie 1 das PSYINDEX-Klassifikationsschema (ZPID, 2016) verwendet, welches auch hier mit den Themen übereinstimmte. Bei einer genaueren Untersuchung zur externen Validierung ergab sich unter

anderem eine zunächst überraschend hoher Wahrscheinlichkeit für sportpsychologische Beiträge, die durch Detailsichtung von repräsentativen Dokumenten zu den Themen auch bestätigt werden konnte. Damit ermöglichte das Klassifikationssystem nicht nur eine externe Validierung der Themen, sondern wies zudem auf Facetten hin, die bei einer reinen Sichtung der wahrscheinlichsten Wörter übersehen worden werden. Dies unterstreicht einerseits die Bedeutung einer genauen Analyse der repräsentativsten Dokumente für jedes Thema (“semantische Validität” sensu Maier et al., 2018), belegt andererseits die Tauglichkeit der Klassifikationen für eine übersichtliche Betrachtung der inhaltlichen Tiefe der Themen. Letzteres ist vor allem dann von Vorteil, wenn (1) Abstracts analysiert werden und somit die Informationsdichte der Themen geringer ist als bei einem standardisierten Vokabular (bedingt durch grammatikalische Wortvariationen oder Synonyme) und wenn (2) das Modell eine hohe Anzahl von Themen umfasst, womit eine Detailprüfung von Dokumenten im Rahmen einer Automatisierung nicht möglich ist. Offen bleibt jedoch auch hier die Frage nach einer Quantifizierung der Korrespondenz zwischen Themen und Klassifikationen (vgl. Abschnitt 4.5). Damit könnten automatisch diejenigen Themen zur manuellen Prüfung herausgestellt werden, die nicht mit dem Klassifikationssystem validiert werden konnten.

Als methodische Einschränkung muss zunächst der Einschluss von ausschließlich deutschsprachigen Abstracts genannt werden, weshalb knapp 15 % der ursprünglichen Anzahl relevanter Publikationen nicht berücksichtigt werden konnten. Vor dem Hintergrund des zunehmenden Trends englischsprachiger Publikationen aus den deutschsprachigen Ländern (Schui & Krampen, 2016) ist Mehrsprachigkeit von Textquellen von zunehmender Bedeutung, weshalb verschiedene Ansätze erprobt werden sollten. Angesichts der eingeschränkten Anwendbarkeit von Topic-Modeling-Verfahren, welche Mehrsprachigkeit berücksichtigen (siehe Abschnitt 2.3) wäre zu prüfen, inwieweit eine automatische Übersetzung von Abstracts (wie etwa mit neuronalen Übersetzungsnetzen von *DeepL*⁴³) zu validen Ergebnissen führt.

Die optimale Anzahl an Themen wurde, wie auch in Studie 1 entsprechend Kosinski et al. (2016) anhand der Held-Out-Likelihood bestimmt. Die breite Streuung bei 100 Wiederholungen eines Algorithmus zur Bestimmung der optimalen Anzahl (siehe Abbildung 8) könnte indizieren, dass die Anzahl der Dokumente zu gering war (Roberts, Stewart, Tingley & Benoit, 2018, S. 55). Andererseits konnte gezeigt werden, dass hohe Held-Out-

⁴³ <https://www.deepl.com>

Likelihood mit semantisch schlechterer Interpretierbarkeit einhergehen kann (Chang et al., 2009). Die nach Abschluss der Analysen zu dieser Studie veröffentlichten Best-Practice-Empfehlungen (Banks et al., 2018; Maier et al., 2018) plädieren daher für Verwendung der semantischen Kohärenz (Mimno et al., 2011) als statistisches Maß (siehe Studie 3). Unabhängig davon ist aus der Verteilung in Abbildung 8 abzuleiten, für jeden Wert von k stets mehrere random seeds zu verwenden. An simulierten Daten könnte geprüft werden, ab welcher Anzahl verschiedener random seeds sich ein stabiles Ergebnis einstellt – in Abhängigkeit verschiedener Korpusgrößen. Da sich die Rechendauer bei größeren Korpora deutlich erhöht, ist es sinnvoll, zufällige Stichproben von Dokumenten zu verwenden. Maier, Niekler, Wiedemann und Stoltenberg (2019) empfehlen, dass die Stichproben 10–20 % der Dokumente umfassen sollten, mindestens jedoch 10000 Dokumente. Andernfalls ist das gesamte Korpus einzubeziehen.

Für eine einfachere Interpretierbarkeit der Themen ist für zukünftige Studien mit Freitexten ein Stemming *nach* der Modellinferenz sinnvoll (Schofield & Mimno, 2016). Konkret heißt das, dass erst in den “fertigen” Themen unterschiedliche Wörter mit dem gleichen Stamm (z. B. “psychisch”, “psychische”, “psychischer”) innerhalb der Themen zusammengefasst werden (vgl. Tabelle D1 in Anhang D). Für eine bessere Lesbarkeit wäre auch die Verwendung von N-Grammen in Erwägung zu ziehen (Banks et al., 2018), also die Zusammenlegung der N zusammengehörigen Wörter. So wäre für $N = 2$ (“Bigramm”) der Begriff “psychische_Störungen” ein Wort im Thema, anstatt der beiden Teile “psychische” und “Störungen”.

6. Studie 3: “Unterscheiden sich die Forschungsinteressen von Frauen und Männern in der Psychologie?”

Bittermann, A., Greiner, N. & Fischer, A. (2020). Unterscheiden sich die Forschungsinteressen von Frauen und Männern in der Psychologie? Eine Analyse von PSYINDEX-Einträgen über einen Zeitraum von 50 Jahren. *Psychologische Rundschau*, 71(2), 103–110. <https://doi.org/10.1026/0033-3042/a000482>

6.1 Hintergrund

Mitte der 1990er Jahre lösten Urs Baumanns (1995) provokante Äußerungen zu möglichen negativen Folgen eines wachsenden Frauenanteils in der Psychologie eine Debatte aus, die unter anderem geschlechtsspezifische Interessen betraf. Über verschiedene Untersuchungsjahre und -stichproben hinweg stellten sich vor allem zwei Themenfelder als von Frauen und Männern unterschiedlich stark besetzt heraus: Die Entwicklungspsychologie wurde als stärker von Frauen präferiert ermittelt (Brack, Reinhardt, Dahme & Hoffmann, 1997; Gnams et al., 2006; Hartmann, 2012), während der Bereich der Statistik und Methoden als stärker von Männern adressiert gefunden wurde (Brack et al., 1997; Hartmann, 2012; König et al., 2015).

Aus methodischer Sicht muss bemängelt werden, dass diese Befunde aus relativ kurzen Zeiträumen mit teilweise länderspezifischen und heterogenen Stichproben stammen. Auch wurden die Interessen sowohl in Umfragen (Gundlach et al., 1999; Hartmann, 2012) als auch in bibliometrischen Analysen (Gnams et al. 2006; König et al., 2015) mit eher groben Themenbereichen (wie “Allgemeine Psychologie”, “Klinische Psychologie” usw.) erfasst – entweder direkt anhand des APA-Klassifikationssystems (Tuleya, 2007) oder daran orientiert. Die gefunden geringen Effekte könnten darauf zurückzuführen sein, dass sich Unterschiede vor allem in Teilaspekten zeigen, die sich nicht auf gesamte Disziplinen der Psychologie generalisieren lassen. Beispielsweise umfasst die Klassifikation “Psychometrie, Statistik und Methoden” (wie bei König et al., 2016) mit Tests, Statistik und Forschungsmethoden drei Inhaltsbereiche, die nicht in gleichem Maße geschlechtsspezifisch präferiert sein müssen.

Angesichts dieser methodischen Einschränkungen bisheriger Studien wurden die Interessen von Frauen und Männern in der Psychologie erstmals mit einer nahezu vollständigen Erhebung von Dissertationen aus dem deutschen Sprachraum untersucht. Die lückenlose Betrachtung eines großen Zeitraumes ermöglicht es, zeitliche Trends zu analysieren. Die Untersuchung von Dissertationen hat im Gegensatz zu den häufig in Ko-Autorenschaft verfassten Artikeln in Fachzeitschriften und Buchkapiteln den Vorteil, dass deren Inhalt klar der wissenschaftlichen Arbeit einer Person zugeordnet werden kann. Da die Themen bisher aus Klassifikationskategorien abgeleitet waren, wurde hier entsprechend der Befunde von Studie 1 mit Topic Modeling ein Analyseverfahren eingesetzt, das detailliertere Einsichten ermöglicht. Wie auch in Studie 2 sollte nicht gezielt nach Themen gesucht werden, um nicht potenzielle Bereiche geschlechtsspezifischer Präferenzen zu übersehen.

Den theoretischen Ausgangspunkt von geschlechtsspezifischen Interessen stellte ein integratives Modell dar, welches versucht, das komplexe Zusammenspiel von individuellem Interesse und sozio-kulturellem Kontext während der Ausbildung und der beruflichen Entwicklung abzubilden (Su, 2020; Su, Stoll & Rounds, 2019). In dieses Modell fließt der individuelle Sozialisationsprozess mit ein, womit davon ausgegangen werden kann, dass der Ausdruck von Interesse mit sozialen Rollenerwartungen und kulturellen Werten in Zusammenhang steht und dadurch geschlechtsspezifische Ausprägungen mitbedingt werden. Dabei ist zu betonen, dass in der vorliegenden Studie dieses Modell selbst nicht getestet werden konnte⁴⁴. Vielmehr sollte es einen Hintergrund für die Fragestellungen und Diskussion dieser szientometrischen Bestandsaufnahme darstellen.

6.2 Fragestellungen

- (1) Unterscheiden sich die Forschungsinteressen von promovierenden Frauen und Männern in der Psychologie?
- (2) Welche zeitlichen Trends von Interessen bzw. Interessensunterschieden können ausgemacht werden?

⁴⁴ Das Modell beschreibt affektive Reaktionen und kognitive Bewertungen bzgl. beruflicher Tätigkeiten, also psychische Prozesse, über die Dissertationsnachweise in Fachdatenbanken keine Auskunft geben können.

6.3 Methode

Die methodischen Schwerpunkte lagen in dieser Studie auf (1) der Modellierung einer Autorenvariable, (2) der Optimierung der LDA-Hyperparameter und der Modellwahl, (3) der Themen-Reliabilität, sowie (4) dem Vergleich von Trend-Dyaden.

6.3.1 Daten

Das Korpus bildeten zunächst die kontrollierten Schlagwörter aller $N = 18050$ in PSYNDEX nachgewiesenen Dissertationen der Jahre 1968 bis 2017. Die Verwendung des kontrollierten Vokabulars wie in Studie 1 wurde einerseits mit der vergleichbaren inhaltlichen Breite begründet. Andererseits standen hier keine besonders aktuellen Themen bzw.

Brennpunktthemen im Fokus, weshalb die Begrenztheit eines standardisierten Vokabulars für die Studienfragestellung kein Problem darstellte. Weitere Gründe werden in ESM 1 in Anhang D dargelegt. Dennoch bot die vorliegende Studie Gelegenheit zu prüfen, welche Unterschiede sich ergeben, wenn bei einem Korpus aus der gesamten Psychologie anstelle des standardisierten Vokabulars Abstracts verwendet werden. Hierfür wurden, wie in Studie 2, alle deutschsprachigen Abstracts eingeschlossen (Ausschluss englischsprachiger Abstracts: 10.3 %). Die Ergebnisse sind in Anhang D zu finden.

Das Geschlecht der verfassenden Person wurde anhand des Vornamens (Naldi, Luzi, Valente & Parenti, 2004) mithilfe des *genderizeR*-Klassifikationsalgorithmus (Wais, 2006) bestimmt. Fell und König (2016) berichten in ihrer szientometrischen Anwendung dieses Algorithmus eine Korrelation von $r = .95$ zwischen wahrem Geschlecht und der automatischen Zuordnung. In der vorliegenden Studie wurde in den 1787 Fällen, in denen der Algorithmus mit einer Sicherheit von weniger als 95 % oder gar nicht zuordnen konnte, sowie bei Unisex-Vornamen (z. B. Dominique, Eike, Luca), das Geschlecht durch manuelle Recherche bestimmt. Insgesamt 79 Personen konnten nach diesem Vorgehen keinem Geschlecht zugeordnet werden, womit der finale Datensatz aus $n = 17971$ Dissertationsnachweisen bestand.

6.3.2 Topic Modeling (Schwerpunkte 1 und 2)

Die Forschungsinteressen wurden aus den Themen der Dissertationen abgeleitet. Dies entspricht dem Grundgedanken des *Author-Topic Models* (ATM) von Rosen-Zvi, Griffiths, Steyvers und Smyth (2004) und einem seitdem in der Szientometrie etablierten Vorgehen. Das ATM ermöglicht nicht nur die Verteilung der Themen innerhalb eines Dokuments zu untersuchen, sondern auch innerhalb einer Autorin bzw. Autors. Dabei wird explizit

berücksichtigt, dass ein Dokument von verschiedenen Personen verfasst sein kann. Dissertationen stellen hier einen Sonderfall dar, da jedes Dokument nur eine Autorin bzw. nur einen Autoren hat. Somit ist die Verteilung der Themen innerhalb eines Dokumentes mit der Verteilung innerhalb einer Person identisch, und die Fragestellung konnte mit dem einfacheren Topic Model, LDA, untersucht werden (erster methodischer Schwerpunkt).

Bei der Bestimmung der Themen wurden die Best-Practice-Empfehlungen von Maier et al. (2018) berücksichtigt. Es wurden Modelle mit verschiedenen LDA-Hyperparametern⁴⁵ und unterschiedlichen Anzahlen von Themen (75 bis 200) geprüft (zweiter methodischer Schwerpunkt). Das finale Modell ($k = 100$) wurde unter Einbezug von Evaluationsmetriken, anhand qualitativer Prüfung von Dokument-Thema-Zuordnungen⁴⁶ und angesichts intersubjektiver Interpretierbarkeit (mit den Koautoren der Studie) gewählt. Als statistische Maße der Modellevaluation wurde, entsprechend der Einsicht aus Studie 2, nicht mehr die Held-Out-Likelihood verwendet, sondern das Produkt aus “Semantischer Kohärenz” (Mimno et al., 2011) und “Exklusivität” (Roberts et al., 2014). Ein Thema ist dann semantisch kohärent, wenn beispielsweise die fünf Wörter mit hoher Wahrscheinlichkeit auch so zusammen in den Dokumenten auftreten (und nicht etwa die ersten beiden in einem Teil der Dokumente, und die nächsten drei in einem anderen Teil). Allerdings weisen Roberts et al. (2014) darauf hin, dass semantische Kohärenz relativ einfach erreicht werden kann, wenn es nur ein paar Themen gibt, die alle von den häufigsten Wörtern dominiert werden. Um Themen besser voneinander abgrenzen zu können schlugen sie daher vor, zu berücksichtigen, ob die wahrscheinlichsten Wörter eines Themas auch exklusiv für dieses Thema sind, also eine geringe Wahrscheinlichkeit für andere Themen haben. Daher wurden für jeden zu testeten Wert von alpha und k mittels unterschiedlicher *random seeds* zehn Modelle berechnet, aus denen das beste bzgl. des Produkts aus Kohärenz und Exklusivität gespeichert wurde.

⁴⁵ Der alpha-Hyperparameter wurden kleiner als in Studie 1 gewählt, da anzunehmen war, dass Dissertationen thematisch eingrenzter sind als wissenschaftliche Publikationen im Allgemeinen. Zum Effekt von LDA-Hyperparametern siehe Abschnitt 2.2.3)

⁴⁶ Konkret wurde für verschiedene Modellvarianten bei einer zufälligen Stichprobe von Dokumenten geprüft, ob die Anzahl der Themen im Dokument sowie ihre Rangfolge (dominantes Thema vs. untergeordnete Themen) plausibel ist.

6.3.3 Themen-Reliabilität (Schwerpunkt 3)

Dem dritten methodischen Schwerpunkt entsprechend wurden die Themen des finalen Modells hinsichtlich ihrer Reliabilität untersucht (Niekler, 2016). Das heißt, es wurde geprüft, ob sie bei wiederholten Modellinferenzen reproduzierbar sind (siehe Abschnitt 2.2.5). Entsprechend des Ansatzes von Niekler (2016) wurde die Themen-Reliabilität mittels Kosinus-Ähnlichkeit analoger Themen-Paare von mindestens .8 bei mindestens acht von zehn wiederholten Modellinferenzen bestimmt. Dieses Vorgehen verringert die Anzahl der Themen zugunsten robusterer Ergebnisse, was insbesondere angesichts der über ein Explorieren hinausgehenden Unterschiedsfragestellung dieser Studie von Bedeutung ist.

6.3.4 Prävalenzdifferenzen

Da in dieser Studie zwei Gruppen dahingehend untersucht wurden, ob die gefundenen Themen jeweils unterschiedlich wahrscheinlich sind, war es aufgrund fehlender Standards in der Literatur nötig zu definieren, wann gefundene Unterschiede als relevant erachtet werden können. Die Prävalenz beschreibt die durchschnittliche Wahrscheinlichkeit eines Themas im Korpus und wird über die mittlere Dokument-Thema-Wahrscheinlichkeit berechnet (siehe Abschnitt 2.2.1). Bei Gleichverteilung aller Themen hat jedes Thema eine Wahrscheinlichkeit von $1/k$, da sich die Wahrscheinlichkeitswerte aller Themen zu 1 addieren. So beträgt bei $k = 100$ Themen die durchschnittliche Prävalenz $1/k = 1/100 = 1\%$ (wenn also jedes Dokument jedes Thema gleich stark behandelt). Dies bedeutet also, dass die beobachteten Prävalenzen der einzelnen Themen umso kleiner ausfallen, je mehr Themen im Modell enthalten sind. Gleiches gilt entsprechend für die Prävalenzdifferenzen zweier Themen oder zweier Gruppen. Daher ist $1/k$ ein hilfreicher Ausgangspunkt um zu beurteilen, wann eine absolute Prävalenzdifferenz ($|D|$) als groß oder klein angenommen werden kann.

In Studie 2 wurde zwecks klarer Zuordnung von Forschungsthemen zu den Disziplinen der Psychologie die doppelte durchschnittliche Prävalenz als Schwellenwert verwendet (siehe Abschnitt 5.4). Für den Vergleich der zwei Gruppen Frauen und Männer wäre dieses Kriterium jedoch zu streng, da bei zwei Gruppen die Gesamtprävalenz des Themas sehr groß sein müsste, so dass eine solche Differenz überhaupt erst möglich sein könnte. Bei einem Thema mit durchschnittlicher Gesamtprävalenz $1/k$ wäre dies entsprechend dann der Fall, wenn eine Gruppe eine durchschnittliche Prävalenz von $2 \cdot 1/k$ aufweist, während die Prävalenz der anderen Gruppe nahezu Null beträgt. Angesichts der bisherigen Studien zu

geschlechtsspezifischen Forschungsinteressen war allerdings nicht zu erwarten, dass es *ausschließliche* “Frauen- bzw. Männerthemen” gibt. Weiterhin sollten hier Unterschiede auch bei Themen mit niedrigerer Prävalenz sichtbar werden. Es wurde daher folgendes Kriterium festgelegt: Eine Prävalenzerhöhung, die die Prävalenz eines anderen Themas – unter Konstanthaltung aller anderen Themenprävalenzen – mehr als halbieren würde (also die halbe durchschnittliche Prävalenz = $0.5 \cdot 1/k$), wurde als ein größerer Unterschied angesehen. Differenzen $\leq 0.5 \cdot 1/k$ wurden entsprechend als ein geringer Unterschied erachtet. Im Falle von $k = 100$ Themen bedeutet dies, dass $|D| = 0.5 \%$ der Cut-Off zwischen größeren und geringen Unterschieden ist. Dieser Cut-Off gilt auch nach Ausschluss unreliabler Themen (siehe oben), da die Prävalenzen der eingeschlossenen Themen während der Inferenz eines Modells mit $k = 100$ Themen berechnet wurden.

6.3.5 Trend-Dyaden (Schwerpunkt 4)

Wurde in den ersten beiden Studien bisher die Verläufe von Themenprävalenzen hinsichtlich Steigung bzw. Spitzen untersucht, stand hier der Verlauf zweier Gruppen innerhalb jedes Themas im Vordergrund (vierter methodischer Schwerpunkt). Um zunächst die Zeitverläufe der Themen von Schwankungen zu befreien, wurden nonlineare Trendkurven getrennt nach Geschlecht mithilfe von Multilayer Perceptrons (MLP) berechnet. Die zeitlichen Trends wurden schließlich auf drei Dimensionen untersucht:

- (1) Dimension “Prävalenzdifferenz”, das heißt, die mittlere Differenz zwischen der weiblichen und männlichen Trendkurve,
- (2) Dimension “Konvergenz” im Sinne einer über die Zeit zunehmenden Nivellierung zweier Trends, sowie
- (3) Dimension “Gesamttrend” unabhängig vom Geschlecht.

6.4 Ergebnisse

Von den 100 Themen im finalen Modell erwiesen sich 48 als reliabel und wurden in die weiteren Analysen eingeschlossen. Abbildung 12 gibt einerseits einen Überblick über die Position aller Themen auf den Dimensionen Prävalenzdifferenz und Konvergenz.

Andererseits ist markiert, welche Themen als nicht-reliabel eingestuft wurden und aus den weiteren Analysen entfernt werden mussten (grau gedruckt). Dabei ist zunächst festzustellen, dass die jeweils drei Themen mit den größten Prävalenzunterschieden (ganz links bzw. ganz

rechts in der Abbildung) zuverlässig reproduziert werden konnten (zu deren Inhalten siehe Tabelle 1 und 2). Die nicht-reliablen Themen mit größeren Unterschieden ($|D| > 0.5 \%$) waren Thema 23 (“visual perception, stimulus parameters, visual stimulation, cognitive processes, color perception”), Thema 90 (“human females, pregnancy, birth, depression (emotion), mothers”) und Thema 31 (“preschool education, social interaction, residential care institutions, child welfare, socialization”). Wie Tabelle D2 in Anhang D entnommen werden kann, verpassten die Themen 23 und 90 nur knapp das Kriterium von Übereinstimmung bei acht von zehn verschiedenen Modellinferenzen: Beide konnten siebenmal bestätigt werden. Thema 31 hingegen konnte keinmal bestätigt werden.

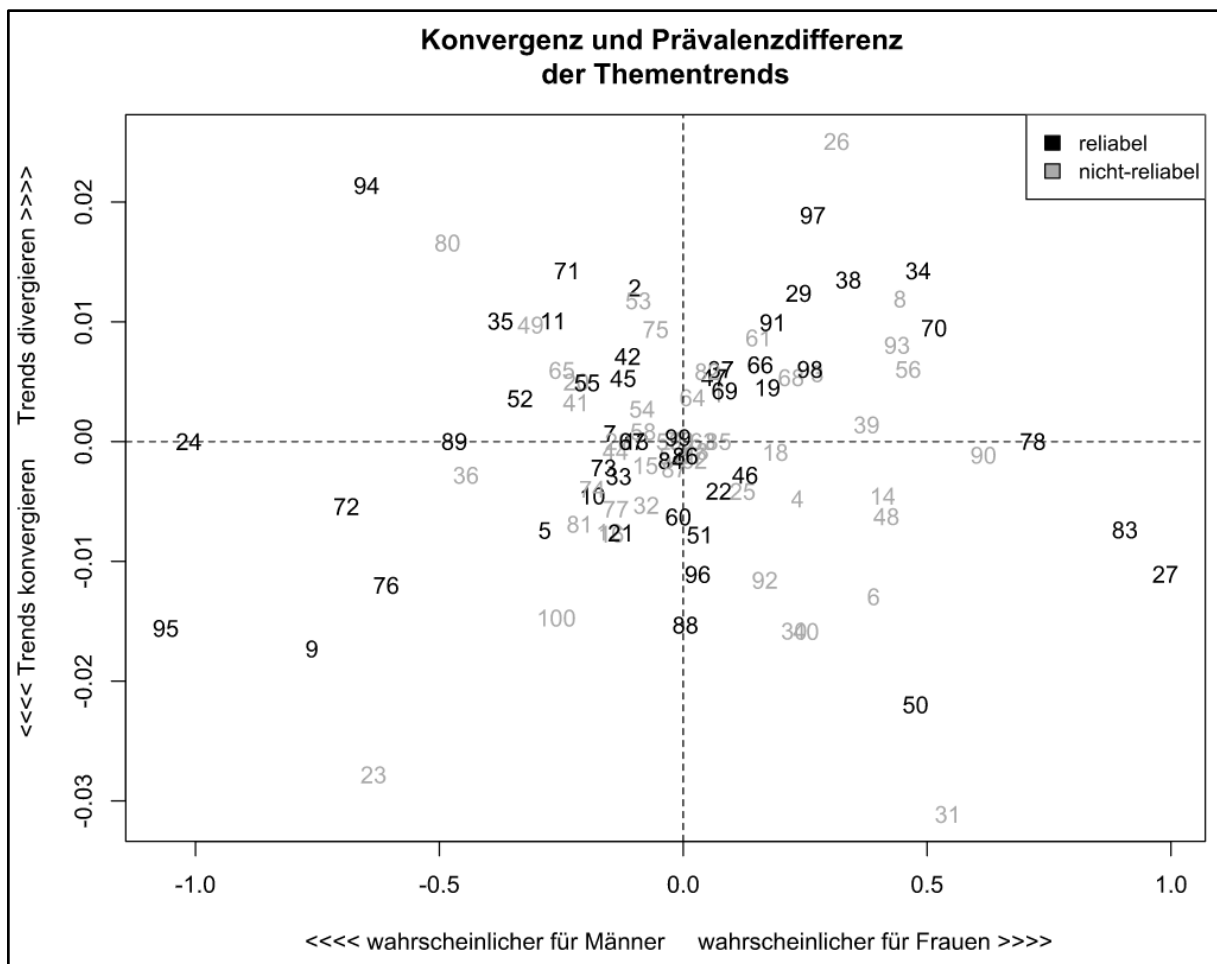


Abbildung 12. Positionen aller Themen auf den Dimensionen „Prävalenzdifferenz“ und „Konvergenz“. Die x-Achse zeigt die über alle Jahre gemittelte Prävalenzdifferenz der Trends in Prozent. Negative Werte zeigen eine höhere Wahrscheinlichkeit des Themas an, von Männern adressiert zu werden, und vice versa. Die y-Achse zeigt die Koeffizienten des linearen Regressionsmodells mit Prädiktor „Publikationsjahr“ und Kriterium „Differenz der Trends“ in Prozent. Schwarz gedruckte Themen erwiesen sich als reliabel, konnten also über verschiedene Modellinferenzen hinweg reproduziert werden.

Tabelle 11. Themen mit höchster Wahrscheinlichkeit, von Frauen behandelt zu werden.

Thema	Wahrscheinlichste Begriffe	D*(%)
27	human females, sex roles, human males, sex role attitudes, human sex differences	0.99
83	parent-child relations, family relations, childrearing practices, mothers, childhood development	0.91
78	mother-child relations, attachment behavior, infant development, mothers, mother-child communication	0.72
70	obesity, eating disorders, bulimia, human females, anorexia nervosa	0.51

Anmerkung: *D = Prävalenzdifferenz (weiblich - männlich)

Tabelle 12. Themen mit höchster Wahrscheinlichkeit, von Männern behandelt zu werden.

Thema	Wahrscheinlichste Begriffe	D*(%)
95	theories, philosophies, history of psychology, theory formulation, psychology	-1.06
24	methodology, statistical analysis, measurement, models, mathematical modeling	-1.01
9	cognitive processes, problem solving, models, theory formulation, knowledge level	-0.76
72	driving behavior, drivers, highway safety, automobiles, motor traffic accidents	-0.69
94	decision making, cognitive processes, judgment, choice behavior, uncertainty	-0.65
76	test validity, intelligence measures, test construction, measurement, testing	-0.61

Anmerkung: *D = Prävalenzdifferenz (weiblich - männlich)

Tabelle 13. Themen mit den geringsten Unterschieden zwischen Frauen und Männern.

Thema	Wahrscheinlichste Begriffe	D*(%)
86	human migration, cross-cultural differences, acculturation, immigration, cross-cultural communication	0.00
88	attribution, self-concept, emotional states, internal external locus of control, achievement	0.00
60	teacher-student interaction, teacher attitudes, teacher characteristics, teachers, teacher education	-0.01
99	occupational stress, stress management, coping behavior, working conditions, health	-0.01
84	social perception, ingroup outgroup, social identity, stereotyped attitudes, attribution	-0.03
51	academic achievement, student characteristics, college students, high school students, elementary school students	0.03
47	age differences, short-term memory, cognitive ability, aging, cognitive development	0.06

Anmerkung: *D = Prävalenzdifferenz (weiblich - männlich)

Wie Abbildung 12 weiterhin verdeutlicht, konzentrierten sich die reliablen Themen größtenteils im Bereich geringer Unterschiede (Gini-Koeffizient $G = .50$): 38 von 48 Themen (79.2 %) zeigten eine absolute Prävalenzdifferenz $|D| \leq 0.5$ %. Die reliablen Themen mit größeren Unterschieden ($|D| > 0.5$ %) werden in den Tabellen 11 bis 12 beschrieben. Tabelle 13 zeigt die Themen mit den geringsten Differenzen. Die Unterschiede unterlagen zumeist zeitlichen Schwankungen. Jedoch wiesen die Trends zweier Themen zeitlich konstant größere Unterschiede auf: Mit höherer Wahrscheinlichkeit wurde das Thema 78 "Mutter-Kind-Beziehung und Entwicklung im Kleinkindalter" von Frauen bzw. das Thema 24 zu "Statistik und Methoden" von Männern adressiert. Abbildung 13 zeigt die Prävalenzverläufe dieser Themen.

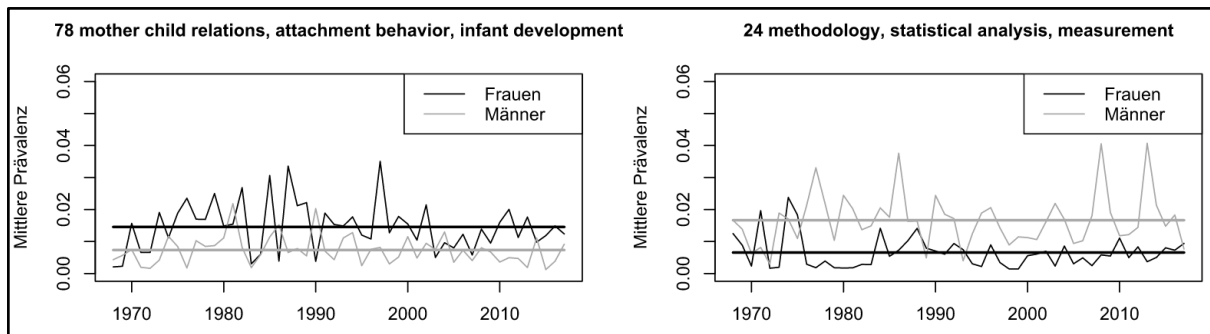


Abbildung 13. Prävalenzverlauf und Trendlinien der beiden Themen mit zeitlichen konstanten Trendunterschieden.

6.5 Diskussion

Mithilfe des Topic-Modeling-Ansatzes konnten die Interessen detaillierter erfasst werden als mit vorab definierten Kategorien, womit der Hauptbefund aus Studie 1 bestätigt werden konnte. Grundsätzlich reihen sich die Ergebnisse in die bisherige Forschung ein, vor allem, was die stärkere männliche Präferenz für Methoden (Thema 24) betrifft. Der Befund bisheriger Forschung einer stärkeren weiblichen Präferenz der Entwicklungspsychologie (Brack et al., 1997; Gnambs et al., 2006, Hartmann, 2012) muss differenziert werden. Zwar liegen zwei der betreffenden Themen im Bereich größerer Unterschiede (Themen 83 und 78 zu Entwicklung der (frühen) Kindheit), das dritte Thema zur kognitiven Entwicklung (Thema 74) war aber für beide Geschlechter gleich wahrscheinlich. Topic Modeling erwies sich auch in dieser Studie den traditionellen methodischen Ansätzen bisheriger Forschung in diesem Bereich hinsichtlich Detailgrad und Differenziertheit überlegen. Durch den Ausschluss unreliabler Themen und die optimierte Modellwahl nach aktuellen Best-Practice-Empfehlungen (Maier et al., 2018) können die Ergebnisse als besonders robust erachtet werden.

Für die Einordnung von Prävalenzdifferenzen wurde mangels Standards in der Topic-Modeling-Literatur ein methodisches Kriterium mit Bezug zur durchschnittliche Prävalenz in Abhängigkeit von der Themenanzahl im Modell vorgeschlagen. Die Berechnung des Cut-Offs mit $0.5 \cdot 1/k$ lässt sich flexibel auf andere Modelle und Stichproben anwenden. Da es sich bei den Daten dieser Studie nahezu um eine Vollerhebung handelt, wäre alternativ denkbar, Lagemaße der Werteverteilung heranzuziehen. Zumindest für Stichproben aus der Population “promovierende Psychologinnen und Psychologen aus dem deutschsprachigen Raum” könnte beispielsweise das 3. Quartil als Cut-Off herangezogen werden, um zwischen größeren und

geringen Unterschieden zu unterscheiden. Wie Abbildung 14 zeigt, stimmt dieser Wert von 0.47 % mit dem in der Studie angewandten Cut-Off von 0.5 % annähernd überein. In Zusammenschau mit dem Kriterium in Studie 2 sollte weitere Forschung an anderen Stichproben prüfen, welches Kriterium sich als praktikabler erweist, bzw. ob ein korpus- und fragestellungsunabhängiges Kriterium überhaupt möglich ist.

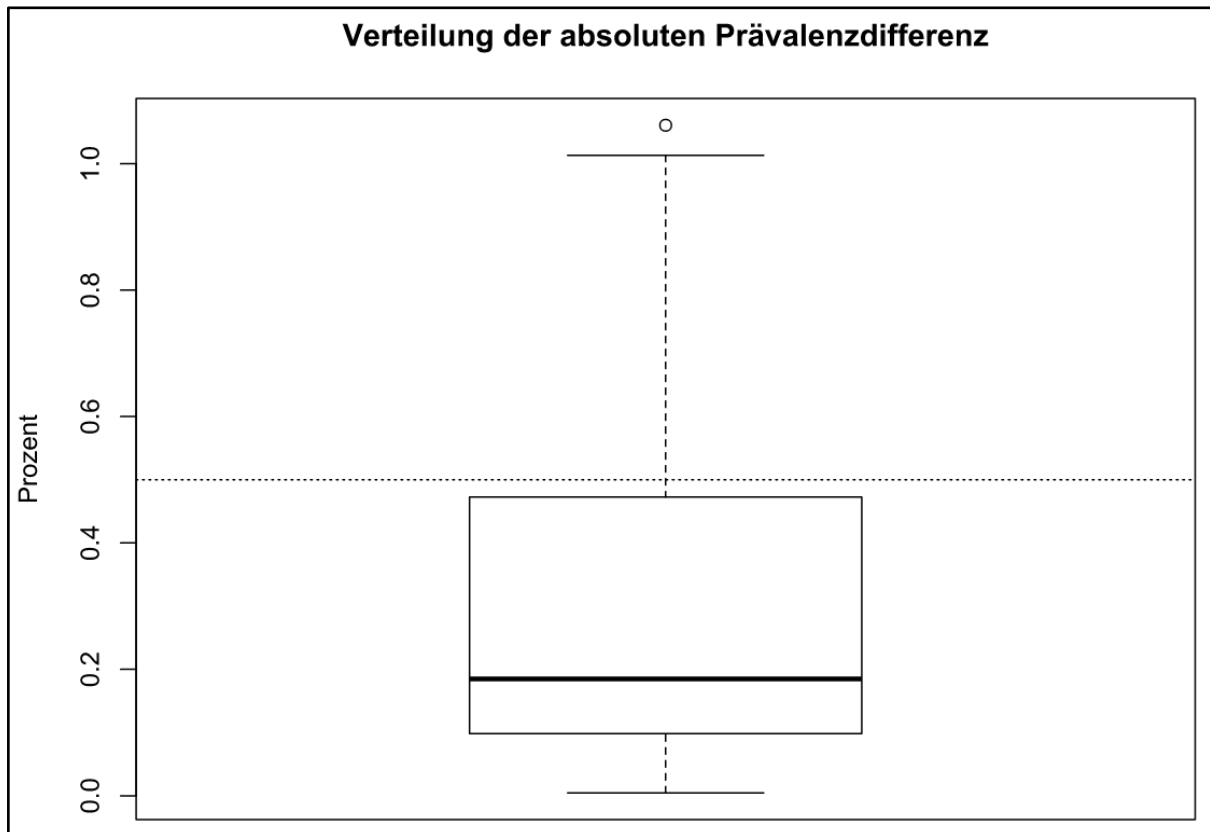


Abbildung 14. Boxplot der absoluten Prävalenzdifferenzen zwischen Männern und Frauen aller Themen. Differenzen, die im vierten Quartil liegen, können als größere Unterschiede interpretiert werden. Die gepunktete horizontale Linie zeigt den Wert für das in der dritten Studie verwendete Kriterium der Größeneinordnung $0.5 \cdot 1/k = 0.5 \%$ an (Modell mit $k = 100$ Themen).

Einschränkend muss festgehalten werden, dass potenziell weitere Einflussfaktoren auf die Wahl des Promotionsthemas jenseits des individuellen Interesse nicht berücksichtigt wurden. Neben dem Interesse können auch Gelegenheitsstrukturen die Wahl des Promotionsthemas beeinflussen, unter anderem die Entwicklung von Lehrstuhldenominationen, die Ausrichtung der Lehrstühle (so z. B. die abnehmende Zahl von Lehrstühlen zur Geschichte der Psychologie), die Bedeutung von Graduiertenschulen oder die Drittmittelstärke eines Faches und damit das Angebot an Promotionsstellen. Zu prüfen wäre hierbei, inwiefern diese Faktoren beide Geschlechter gleichermaßen oder unterschiedlich betreffen. Selbstauskünfte einer repräsentativen Stichprobe von forschungstätigen Psychologinnen und Psychologen oder eine Analyse von strukturellen Faktoren wären eine aufschlussreiche Ergänzung, um in zukünftigen Studien den Anteil weiterer Einflussfaktoren auf die Wahl des Promotionsthemas zu ermitteln.

Die Befunde dieser Studie beruhen auf den Dissertationen von Psychologinnen und Psychologen. Weiterführende Untersuchungen könnten sich damit beschäftigen, ob die gefundenen geschlechtsspezifischen Themenpräferenzen im weiteren akademischen Verlauf nach der Promotion – der in der Psychologie mit einer Abnahme des Frauenanteils verbunden ist (Antoni, 2019) – erhalten bleiben. Dazu könnte der wissenschaftliche Output von Professorinnen und Professoren (vgl. Mayer & Rathmann, 2018) untersucht werden. Für die Analyse von Publikationen in Ko-Autorenschaft wäre schließlich ein ATM (Rosen-Zvi et al., 2004) das Topic Model der Wahl.

7. Entwicklung einer nutzerfreundlichen App zur Exploration und Analyse psychologischer Forschungsthemen (“PsychTopics”)

Bittermann, A. (2019). Development of a user-friendly app for exploring and analyzing research topics in psychology. In G. Catalano, C. Daraio, M. Gregori, H. F. Moed & G. Ruocco (Hrsg.), *Proceedings of the 17th Conference of the International Society for Scientometrics and Informetrics* (S. 2634–2635). Rom: Edizioni Efesto. Verfügbar unter <https://doi.org/10.23668/psycharchives.2521>

7.1 Hintergrund

Wie in den drei zuvor beschriebenen Studien (Abschnitte 4 bis 6) gezeigt, ermöglichen Themen, die mit Topic Modeling identifiziert wurden, einen detaillierten Blick auf die Vielfalt psychologischer Forschungsthemen. Über die berichteten Ergebnisse hinaus können die ermittelte Themen auch für unterschiedlichste Anschlussfragen von Interesse sein, etwa welche Themen in einem bestimmten Zeitraum populär waren. So wurde in der Diskussion von Studie 1 die Frage aufgeworfen, welcher Zeitraum eingeschlossen werden sollte, über den ein Thema als “Hot Topic” definiert werden sollte (siehe Abschnitt 4.5). Eine Lösung hierfür kann eine Software sein, welche eine einfach zu bedienende Oberfläche für die Ergebnisse des Topic Models bereitstellt. Damit könnte individuell der zu berücksichtigende Zeitraum definiert werden. Außerdem könnten die Themen, die aus einer Fachdatenbank generiert wurden, mit den Dokumenten dieser Datenbank verknüpft werden. Im Sinne der ursprünglichen Intention von Topic Modeling als Verfahren des *Information Retrieval* (Blei et al., 2003) wäre es somit möglich, passende Literatur zu jedem Thema anzuzeigen: Einerseits für weitere Recherche, andererseits zur semantischen Validierung und Interpretation der Themen (sensu Maier et al., 2018).

Verschiedene bereits entwickelte Benutzeroberflächen bieten teilweise umfangreiche Möglichkeiten, um die Ergebnisse von Topic-Modeling-Analysen zu erkunden und Textkorpora zu erschließen (z. B. Chaney & Blei, 2012; Gardner et al., 2010; Goldstone, Galán, Lovin, Mazzaschi & Whitmore, 2014; Guille & Soriano-Morales, 2016; Kavvadias, Drosatos & Kaldoudi, 2019; Niekler, Wiedemann & Heyer, 2017). Allen solchen Tools

gemein ist eine grafische Darstellung der Themenprävalenz und ihrer zeitlichen Veränderung, sowie die Möglichkeit, zu den Themen passende Literatur bzw. Dokumente zu finden. Die Unterschiede liegen in der Visualisierung der Themen (von Wortlisten bis zu Wordclouds) und den zusätzlich präsentierten statistischen und linguistischen Metriken (von der Wortzahl bis zur Entropie der Themenverteilung im Dokument). Diese Tools richten sich in erster Linie an versierte Nutzerinnen und Nutzer, welche zahlreiche Möglichkeiten zur Korpuserkundung erhalten. Für Laien im Umgang mit Topic Models besteht jedoch die Gefahr, von der schier unendlichen Anzahl an Informationen und der methodenspezifischen Terminologie überfordert zu werden. Die Literaturrecherchen zu den Themen beschränken sich außerdem häufig auf Zitationen ohne Metadaten und Abstracts. Darüber hinaus gibt es noch keine Implementierung für psychologische Fachliteratur.

Anhand der Grundlagenforschung der drei im Rahmen dieser Dissertation durchgeführten Studien wurde daher ein digitales Produkt zur Exploration und Analyse psychologischer Forschungsthemen entwickelt. Zielgruppe sind alle an psychologischen Forschungsthemen interessierten Personen: Forschende, Lehrende, Praktikerinnen und Praktiker sowie die interessierte Öffentlichkeit. Daher lag der Schwerpunkt – neben einer nach den Erkenntnissen der drei Studien leichten Anwendung von Topic Modeling – auf Nutzerfreundlichkeit und einfacher Bedienbarkeit.

7.2 Ziele

Die Entwicklung einer ersten App-Version verfolgte als Ziele

- (1) die Nutzerfreundlichkeit der Oberfläche und
- (2) die Möglichkeit zur direkten Literaturrecherche.

Die zu entwickelnde App sollte erstens vor allem von Nutzerinnen und Nutzern ohne methodische Kenntnisse im Bereich Topic Modeling ermöglichen intuitiv bedienbar sein. Daher sollte die Auswahl an dargebotenen Informationen auf ein Minimum beschränkt werden. Außerdem sollten die Themen anhand ihrer Wörter leicht interpretierbar sein. Zweitens sollte es ermöglicht werden, themenrelevante Literatur zu finden, die auf Niveau einer Fachdatenbank präsentiert wird.

7.3 Methode⁴⁷

Im Folgenden wird neben einer Erläuterung des methodischen Vorgehens beschrieben, inwiefern die Einsichten der drei Studien im Rahmen dieser Dissertation eingeflossen sind. Zusammenfassend betrifft dies vor allem eine dynamische Analyse von Brennpunktthemen (Studie 1), eine verbesserte Modellwahl durch multiple Modellinferenzen (Studie 2), sowie den Ausschluss unreliabler Themen, der LDA-Hyperparameter-Optimierung und einer *state-of-the-art* Modellwahl (Studie 3). Die Originalpublikation ist in Anhang E zu finden.

7.3.1 Daten

Aufbauend auf Studie 1 wurde ein aktualisiertes Korpus von PSYINDEX-Einträgen aus dem Bereich der gesamten Psychologie verwendet. Unter Einschluss der Jahre 1980 bis 2017 und weiteren Kriterien aus Studie 1 umfasste der Datensatz $N = 329240$ Publikationen. Wie in den Studien 1 und 3 begründet, wurden standardisierte Schlagwörter (ZPID, 2016) als Textquelle verwendet. Zwar bleibt damit in Abschnitt 4.5 diskutierte Nachteil des potenziellen Versäumens besonders aktueller Themen bestehen, doch bieten diese Schlagwörter drei Vorteile, die im Rahmen dieser App diesem Nachteil überwiegen: (1) Diese Begriffe sind leicht interpretierbar und in einem Thesaurus definiert, (2) alle PSYINDEX-Nachweise unabhängig von der Sprache der Originalpublikation können eingeschlossen werden und (3) mit ihnen kann über das Suchportal PubPsych.eu eine direkte Literatursuche in PSYINDEX durchgeführt werden. Letzteres bedeutet konkret, dass die Wörter der Themen – mit “OR”⁴⁸ verknüpft und nach den im Modell berechneten Wort-Thema-Wahrscheinlichkeiten gewichtet – direkt eine für Nutzerinnen und Nutzer transparente Suchsyntax ermöglichen⁴⁹. Für die Gewichtung werden die Wort-Thema-Wahrscheinlichkeiten faktorisiert, mit der Wahrscheinlichkeit des letzten Begriffes als Basis. Hat beispielsweise der erste Begriff eine Wahrscheinlichkeit von 80 %, zu diesem Thema zu gehören, und der letzte Begriff eine Wahrscheinlichkeit von 60 %, so beträgt der Faktor $80 / 60 = 1.33$. Eine solche gewichtete Suchanfrage führt zur gleichen Anzahl an Treffern, jedoch mit bzgl. Relevanz optimierter Rangfolge.

⁴⁷ Die Beschreibung im Folgenden entspricht der Version in Anhang E. Die App wurde Ende November 2019 aktualisiert mit Aufnahme des Publikationsjahres 2018 und verschiedenen Optimierungen an der Oberfläche (siehe <https://bit.ly/2Ny2ipH>).

⁴⁸ Trotz der “OR”-Verknüpfung werden diejenigen Nachweise zuerst angezeigt, in denen die meisten der gesuchten Begriffe enthalten sind.

⁴⁹ So wird aus dem Thema “posttraumatic stress disorder, emotional trauma, refugees, trauma, war” die Suchanfrage “(CT=(“posttraumatic stress disorder”^{2.45} OR “emotional trauma”^{1.81} OR “refugees”^{1.21} OR “trauma”^{1.15} OR “war”)) DB=PSYINDEX”

7.3.2 Oberfläche der App

Die Oberfläche der App wurde in R Shiny (Chang, Cheng, Allaire, Xie & McPherson, 2018) programmiert. Dafür sprachen vor allem zwei Gründe: Erstens konnte der Analysecode aus den drei Studien direkt wiederverwendet werden. Zweitens ist R unter forschungsaktiven Psychologinnen und Psychologen eine weit verbreitete Programmiersprache, was den Austausch erleichtert. Zudem können Grafiken mit beliebigen R-Paketen erstellt werden.

Die Oberfläche der App sollte vor dem Hintergrund folgender Nutzungsszenarien erstellt werden:

- (1) Exploration von Brennpunktthemen (“Hot Topics”; vgl. Studie 1),
- (2) Analyse der Höhen und Tiefen in der Popularität eines Themas im Zeitverlauf (Studien 1–3) sowie
- (3) Vergleich von Publikationstrends mit gesellschaftlichen Prozessen (vgl. Studie 2).

Entsprechend des in Studie 1 diskutierten Aspektes einer individuellen Festlegung des zu analysierenden Zeitraums (vgl. Abschnitt 4.5) sollte es per Schieberegler möglich sein, einzelne Jahre bzw. Zeiträume auszuwählen. Auf Grundlage dieser Auswahl können dann populäre Themen (also mit hoher Prävalenz) und Brennpunktthemen dynamisch berechnet werden. Zur Betrachtung des Zeitverlaufs der Themenprävalenz sollten Trendgrafiken wie in den Studien 1 und 2 dargeboten werden. Zum Vergleich von Publikationstrends mit Themen der Gesellschaft wurde eine erste experimentelle Funktion implementiert, welche ab einem gewünschten Zeitpunkt den beobachteten mit einem erwarteten Verlauf vergleicht. Letzterer wird basierend auf Zeitreihen des bis zum Wunschzeitpunkt bisherigen Verlaufstrends berechnet⁵⁰. Der Trend des beobachteten Zeitraums wurde wie in Studie 3 mit Multilayer Perceptrons bestimmt.

7.3.3 Topic Modeling

Wie in Studie 3 wurde LDA mit verschiedenen Kandidaten für den alpha-Hyperparameter und die Anzahl der Themen k (unter Durchführung multipler Modellinferenzen wie in den Studien 2 und 3) angewandt. Die zu testenden Wertebereiche für k (250–550) und alpha wurden an Studie 1 angelehnt, da die Korpora vergleichbar sind (gesamte Psychologie). Das

⁵⁰ Derzeit mit den default-Parametern des R-Paketes “forecast” (Hyndman et al., 2019).

optimale Modell wurde erneut mittels manueller Prüfung der Interpretierbarkeit der Themen und hinsichtlich Stimmigkeit der Dokument-Themen-Zuordnung bestimmt.

7.4 Ergebnisse

Die App wurde unter dem Titel “PsychTopics” frei zugänglich zur Verfügung gestellt (<https://abitter.shinyapps.io/psychtopics/>). Der Code der hier berichteten Version ist bei PsychArchives.org verfügbar (<https://dx.doi.org/10.23668/psycharchives.2410>).

7.4.1 Themen

Das finale Modell umfasste 325 Themen, wobei nach Ausschluss unreliabler Themen (34.5 %) schließlich 213 Themen in die App aufgenommen wurden. Zur Darstellung der Themen wurden die fünf Wörter mit der höchsten Wahrscheinlichkeit für das jeweilige Thema verwendet. Sowohl für populäre Themen als auch für Themen mit besonders steigendem (“Hot Topics”) bzw. fallendem Trend (“Cold Topics”) sensu Griffiths und Steyvers (2004) werden die “Top 10” angezeigt.

7.4.2 Ansichten der App

Die Oberfläche der App ist responsiv, das heißt, sie passt sich der Bildschirmauflösung an. Die Grafiken und Tabellen werden entsprechend vergrößert oder verkleinert. Nutzerinnen und Nutzer können die folgenden sechs Ansichten auswählen: (1) “Populäre Themen eines bestimmten Jahres”, (2) “Populäre Themen eines Zeitraumes”, (3) “Hot Topics”, (4) “Cold Topics”, (5) “Alle Themen” sowie (6) “Erwartete Verläufe”. Über einen Schieberegler, der bei jeder Ansicht eingeblendet bleibt, kann der Nutzer den Zeitraum eingrenzen. Abbildung 15 zeigt die Oberfläche der App mit ausgewählter Ansicht (1) “Populäre Themen eines Jahres”. Das Thema, das 2017 insgesamt am stärksten in der psychologischen Fachliteratur adressiert wurde, ist “posttraumatic stress disorder, emotional trauma, refugees, trauma, war”. Die Ansicht für (2) “Themen eines Zeitraums” ist vom Aufbau identisch. Die Ansichten (3) und (4) der “Hot” bzw. “Cold Topics” zeigen Verlaufsdigramme der Themenprävalenz über den ausgewählten Zeitraum (Abbildung 16). Die Ansicht (5) für alle Themen (Abbildung 17) dient dazu, um nach gezielten Themen zu suchen oder um alle 213 Themen durchzusehen. Die Prävalenz ist hier grafisch dargestellt, um sie mit dem Maximum (dem insgesamt populärsten Thema) und der durchschnittlichen Prävalenz ($1/k$) zu vergleichen. Bei der experimentellen Ansicht (6) zum Vergleich von beobachteten und erwarteten Verläufen (Abbildung 18) kann ein Zeitpunkt festgelegt werden, ab dem der Trend vorhergesagt wird. In

den ersten fünf Ansichten kann bei jedem Thema mit einem Klick auf den “Suche in PSYINDEX”-Button eine entsprechende Suchanfrage in PubPsych.eu gestartet werden (vgl. Abschnitt 7.3.1); bei Ansicht (6) “Erwartete Verläufe” wurde dies aus Platzgründen vorläufig eingespart.

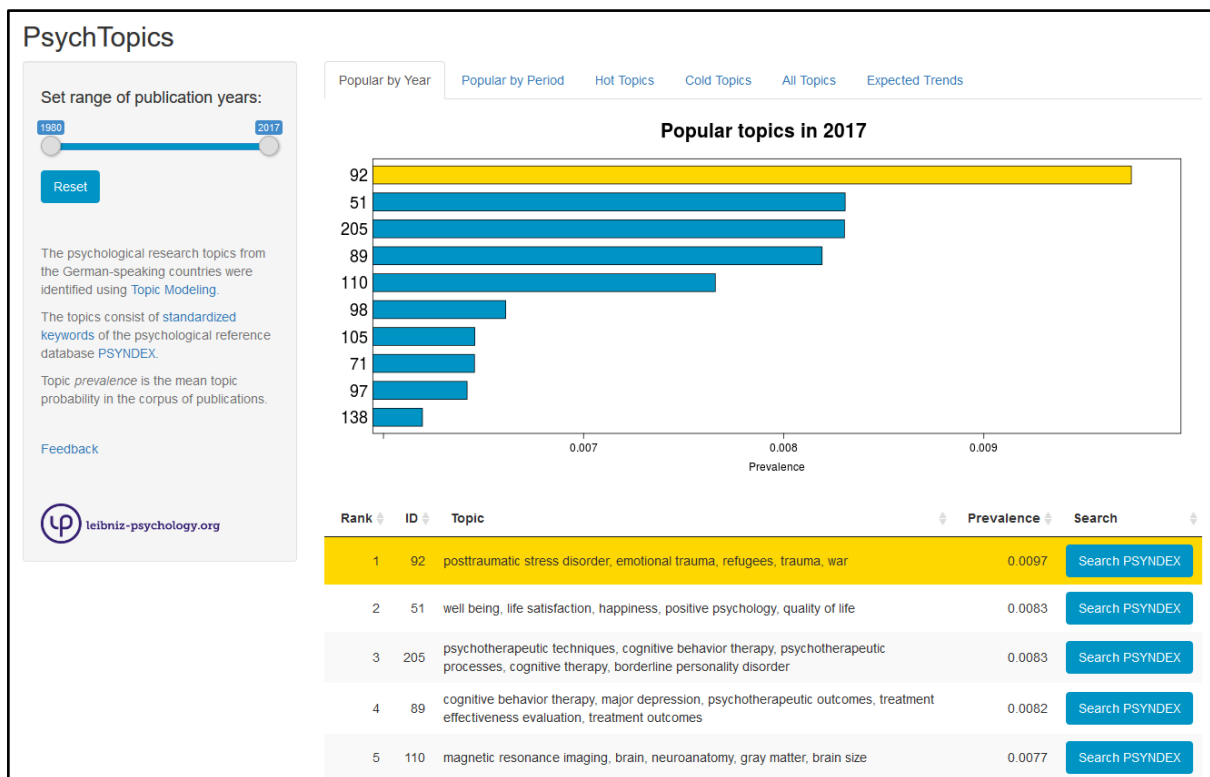


Abbildung 15. Ansicht der populären Themen für das Jahr 2017 (Themen, die in diesem Jahr insgesamt am stärksten adressiert wurden). In der Tabelle unter der Abbildung finden sich die Themen-IDs, die Wörter der Themen, die Prävalenz und ein Button, der eine entsprechende PSYINDEX-Suchanfrage im Suchportal PubPsych.eu ausführt.

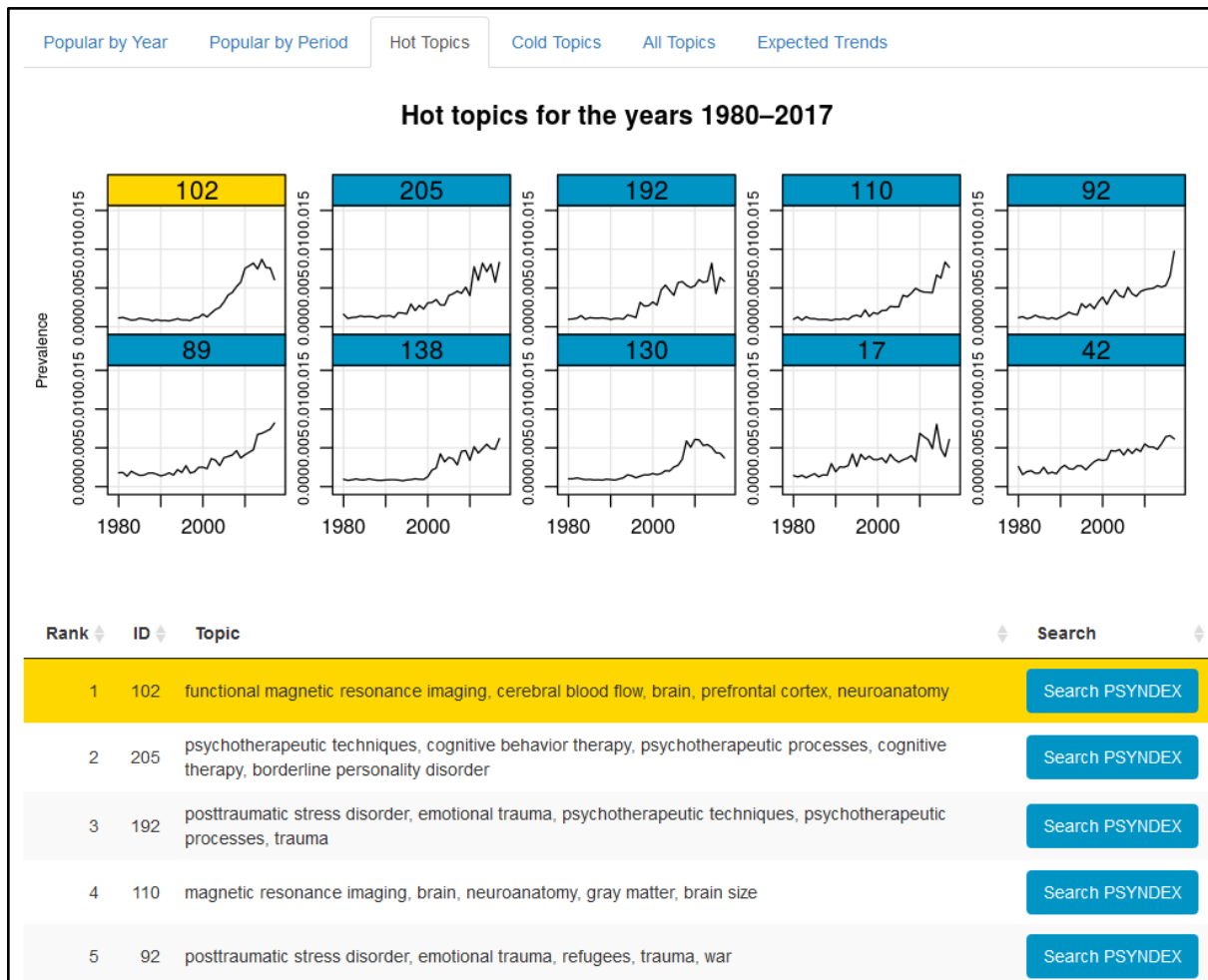


Abbildung 16. Ansicht der “Hot Topics” im Zeitraum 1980–2017, also der Themen, mit den am stärksten steigenden Trends.

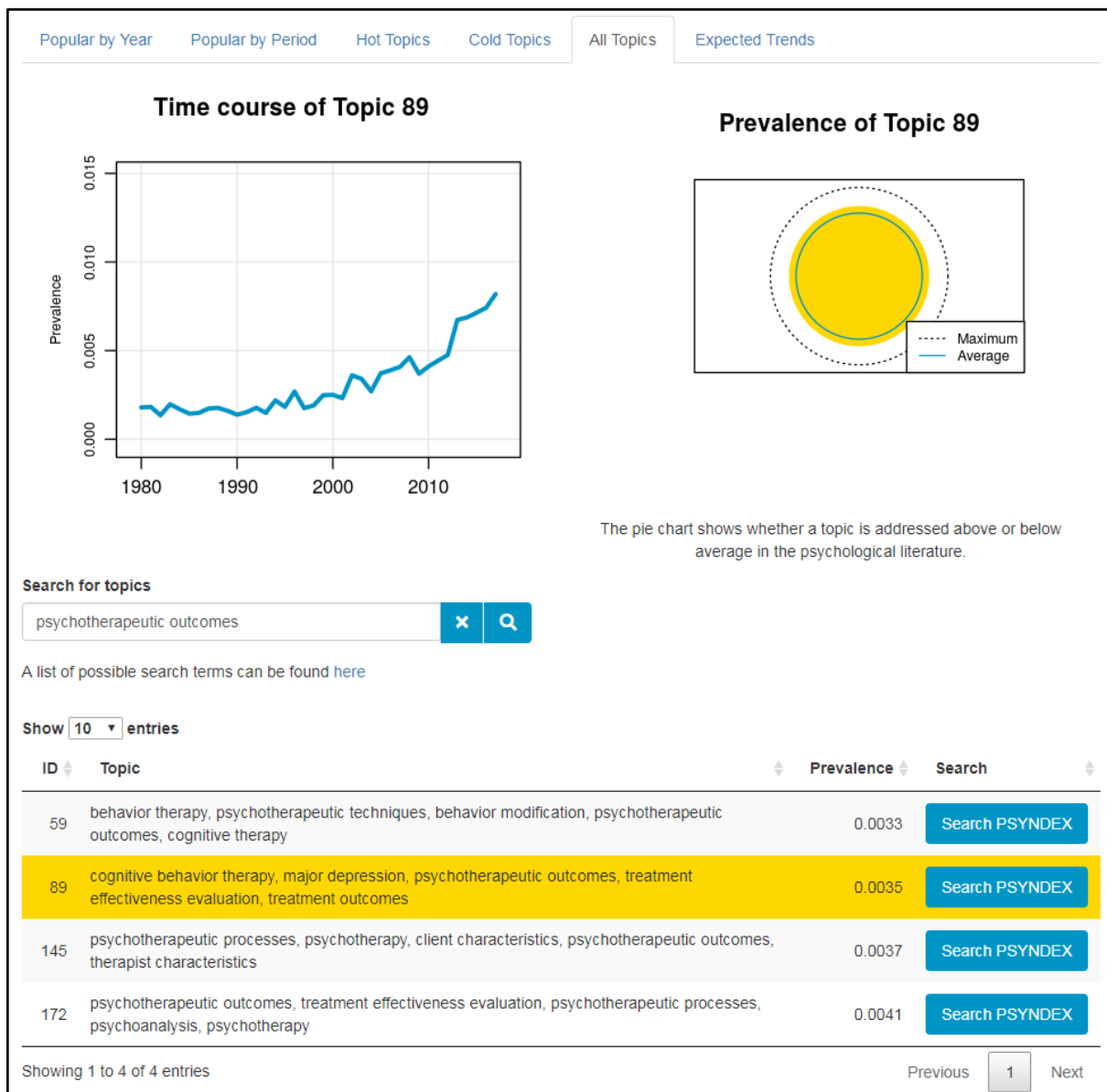


Abbildung 17. Ansicht “Alle Themen”. Hier kann anhand des standardisierten Vokabulars von psychologischen Schlagworten (ZPID, 2016) nach Themen gesucht werden (hier: “psychotherapeutic outcomes”). Angezeigt werden der zeitliche Verlauf sowie die Prävalenz des Themas im Vergleich zu Maximum und Durchschnitt (hier ein überdurchschnittlich häufig adressiertes Thema).

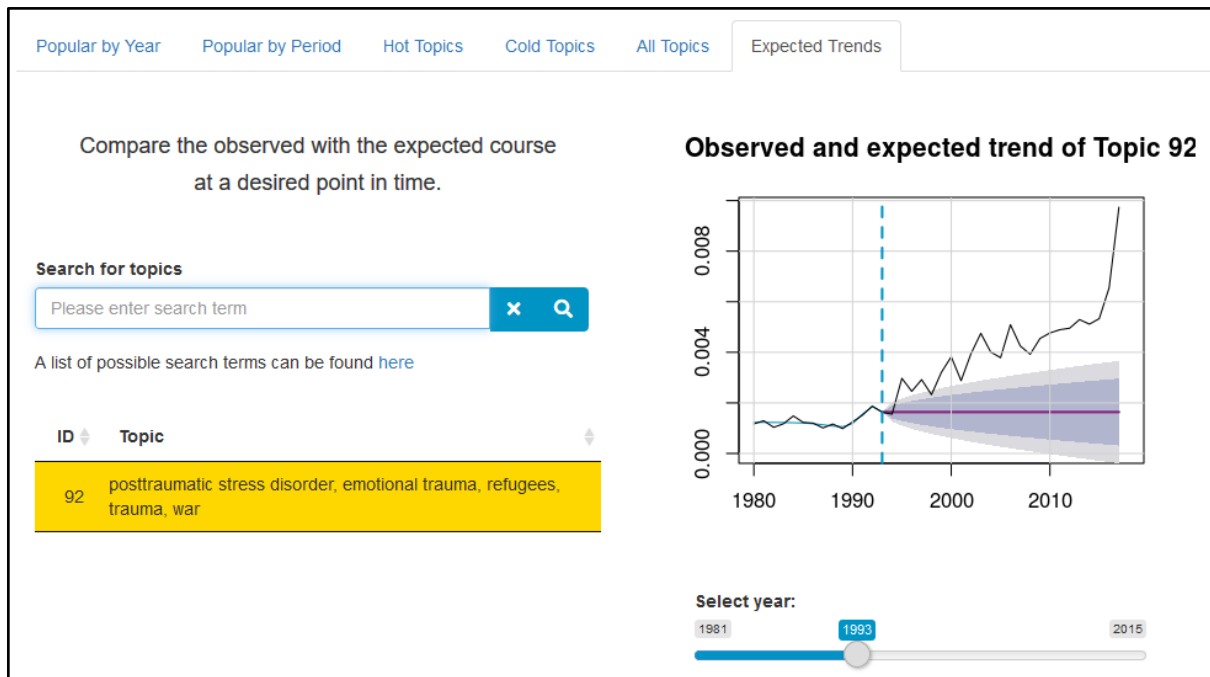


Abbildung 18. Ansicht zum Vergleich von erwarteten und beobachteten Verläufen. Das Thema zu Traumatisierung bei Geflüchteten steigt ab 1993 stärker an als vom bisherigen Verlauf (horizontale Linie; mit grauen Prognoseintervallen) zu erwarten war.

7.5 Diskussion

7.5.1 Erste Rückmeldungen und (methodische) Schlussfolgerungen

Erste Erfahrungen von Nutzerinnen und Nutzern waren sehr positiv. Vor allem die Möglichkeit, passende Literatur in PubPsych zu recherchieren wurde hervorgehoben. Für Unklarheit sorgte jedoch der Begriff der “Prävalenz” bzw. ihre Zahlenwerte. Dies ist wenig überraschend, da der probabilistische Charakter von Topic Modeling und die sinkenden Prävalenzwertebereiche mit steigender Themenanzahl eine Interpretation ohne Vorkenntnisse erschweren. Denkbar wäre, das im Rahmen von Studie 3 vorgeschlagene Kriterium zur Bestimmung von Frauen- und Männeranteilen zu adaptieren (vgl. Abschnitt “Frauenanteile” in PsychArchives-ESM 1 zu Studie 3 in Anhang D): Ein Dokument wird dann zu einem Thema zugehörig gezählt, wenn das Thema das insgesamt Dominierende im Dokument ist. Operationalisiert wird dies mit dem Kriterium: Dokument-Thema-Wahrscheinlichkeit $> 50\%$. Verteilen sich beispielsweise in einem Modell mit $k = 3$ die Themen im Dokument zu 20 % (Thema 1), 60 % (Thema 2) und 20 % (Thema 3), so wird dieses Dokument zu Thema 2 gezählt. Da hier jedoch weniger Information als in die Prävalenzen einfließt, wäre konkret zu prüfen, welche Auswirkungen eine solche Konvertierung von Wahrscheinlichkeiten zu

Häufigkeiten auf die Zeitverläufe und Trends hätte. Eine ergänzende Darstellung der Häufigkeiten zu den Prävalenzangaben wäre in jedem Fall möglich.

Die Oberfläche dieser Version ist in englischer Sprache umgesetzt; eine parallele Testversion wurde in deutscher Sprache gehalten. Eine Auswahl der Sprache durch Nutzerinnen und Nutzer könnte somit prinzipiell implementiert werden, es bliebe zu prüfen, wie nutzerfreundlich eine Umsetzung mit R Shiny wäre. Eine weitere Feinjustierung betrifft die Unterstützung kleiner Bildschirme (vor allem Smartphones): Die Abbildungen zeitlicher Trends werden nur im Breitbildformat durch Drehen des Bildschirms (“Landscape Mode”) erkennbar dargestellt. Eine Evaluierung der Benutzerfreundlichkeit im Rahmen von Nutzerstudien steht noch aus und ist für das nächste größere Update, welches die eben genannten Einschränkungen berücksichtigt, angedacht. Im Rahmen einer solchen Evaluierung könnte auch geprüft werden, inwiefern detaillierte Informationen für fortgeschrittene Interessentinnen und Interessenten angezeigt werden können, ohne an Benutzerfreundlichkeit und Übersichtlichkeit einzubüßen.

Die “Hot Topics” von Studie 1 bzgl. Neuropsychologie, Online-Therapie, Traumatisierung von Geflüchteten und Genetik werden auch in der App dargestellt. Dass sich dies trotz unterschiedlicher Werte für k und α zeigt, bestätigt die Robustheit der Ergebnisse. Nicht mehr unter den “Top 10” (doch weiterhin im Modell enthalten) sind die Themen zu visueller Aufmerksamkeit und Migration allgemein. Ihnen wurde von Gesundheitspsychologie, Psychotherapieforschung und Tests der Rang abgelaufen. Dies lässt sich einerseits darauf zurückführen, dass durch das kleinere k Themen zusammengefasst wurden, was Veränderungen ihrer Prävalenz bewirkt. Andererseits können zwischenzeitliche Aktualisierungen in PSYINDEX zu mehr Publikationen bei manchen Themen geführt haben. Die Ränge der Themen sollten daher nicht als fix erachtet werden⁵¹. Vielmehr sind die “Hot Topics” – dem probabilistischen Charakter der Methode entsprechend – als eine gemeinsame Gruppe zu interpretieren, das heißt, ohne Vergleich von konkreten Rängen. Eine Schlussfolgerung daraus ist, die Anzahl an anzuzeigenden Themen nicht auf zehn zu begrenzen, sondern vom Nutzer individuell treffen zu lassen. Eine unkompliziert auswählbare

⁵¹ Im Update vom November 2019 sind erneut kleinere Rangwechsel von Themen festzustellen, die aufgrund der noch vorliegenden Beschränkung auf zehn Hot Topics nicht vollständig sichtbar sind. Um keine falschen Schlussfolgerungen zu evozieren, wurden die Rangnummerierungen entfernt. Im nächsten Update sollen die in diesem Abschnitt diskutierten Anpassungen der Darstellung umgesetzt werden.

post-hoc-Zusammenfassung von korrelierten Themen⁵² zu Themen-Clustern (Coppola, Roberts, Stewart & Tingley, 2016; Jiang, Qiang & Lin, 2016) könnte zusätzlich helfen, einen Überblick über die Themenvielfalt zu erlangen.

7.5.2 Perspektiven zur Weiterentwicklung

Zukünftige Weiterentwicklungen dieser App sind auf mehreren Ebenen denkbar. Zunächst könnte eine engere Verknüpfung mit Suchportalen wie PubPsych.eu eine dynamischere Generierung der Themen ermöglichen. Derzeit basieren die Themen auf einem (regelmäßig erstellten) PSYNDEX-Schnappschuss, was insofern unproblematisch ist, als dass das Erscheinen neuer Themen erfahrungsgemäß erst mit einer größeren Anzahl an Dokumenten einhergeht. Für eine vollständige Automatisierung wäre es jedoch unabdingbar, eine aktuelle Datenquelle unmittelbar analysieren zu können. Damit wäre es dann auch möglich, (registrierten) Nutzerinnen und Nutzern Literaturvorschläge anhand eines Topic Models zu machen (Amami, Pasi, Stella & Faiz, 2016; vgl. Abschnitt 8.2.2).

In Studie 2 wurde die Dokumentvariable “Studienmethodik” in das Modell aufgenommen, um die Anteile empirischer Evidenz an den Themen zu untersuchen. Dies ist auch in dieser App umsetzbar und damit könnten diejenigen Themen hervorgehoben werden, welche für Forschungssynthesen interessant sein könnten oder zu welchen noch empirische Pionierarbeit geleistet werden kann. Dadurch könnte ein Beitrag zur Generierung von *Evidence Gap Maps* (Snilstveit, Vojtkova, Bhavsar & Gaarder, 2013) in der Psychologie geleistet werden.

Studie 2 untersuchte deskriptiv aus geschichtswissenschaftlicher Perspektive, inwiefern eine zeitliche Korrespondenz zwischen gesellschaftlichen Prozessen und den Publikationsverläufen psychologischer Fachliteratur vorliegt. Die in der App implementierte Funktion zum Vergleich von erwarteten und beobachteten Verläufen stellt einen ersten Schritt dar, eine solche Korrespondenz zu quantifizieren. Aktuell ist die Auswahl des Jahres, ab dem der Vergleich stattfinden soll, von den Nutzerinnen und Nutzern selbst zu treffen. Denkbar wäre, einen zweiten zeitlichen Verlauf in die Abbildung aufzunehmen: Die Prävalenz des Themas in den (Nachrichten-)Medien. Diese beiden Verläufe könnten dann (unter Berücksichtigung einer Verschiebung bedingt durch den Publikationsprozess) verglichen werden. Der Einbezug solch externer Textquellen könnte nicht nur zur Beantwortung der

⁵² wie etwa “functional magnetic resonance imaging, cerebral blood flow, brain, prefrontal cortex, neuroanatomy” und “magnetic resonance imaging, brain, neuroanatomy, gray matter, brain size”

Frage, welche gesellschaftlich bedeutsamen Themen in der Psychologie behandelt werden, sondern auch umgekehrt: “Welche Themen werden nicht von der Psychologie aufgegriffen?” Relevante Vergleichszeitpunkte würden sich aus den Spitzen des Medien-Korpus ergeben. Prinzipiell könnte dies durch Text Mining von Zeitungsartikeln (Wiedemann, 2016) mit einem anschließenden Mapping der darin gefundenen Themen auf die Themen der Psychologie umgesetzt werden. Die Übereinstimmung von Themen aus unterschiedlichen Korpora könnte etwa anhand ihrer Kullback-Leibler-Divergenz bzw. Jensen-Shannon-Divergenz (Steyvers & Griffiths, 2007) oder Kosinus-Ähnlichkeit (Niekler, 2016) bestimmt werden. Dabei würde es nötig sein, die Texte zuvor in eine “gemeinsame Sprache” zu transferieren. Dies könnte durch eine Extraktion der den Texten zugrundeliegenden zentralen Konzepte geschehen – was wiederum der PSYNDEX-Verschlagwortung mit dem APA-Thesaurus entspricht (Tuleya, 2007; ZPID, 2016). Allerdings wäre einzuwenden, dass ein Fachvokabular der Psychologie für ein Medien-Korpus nicht ausreicht. Die Verwendung von Wikipedia-Kategorien (Zarrinkalam, Kahani & Bagheri, 2019) könnte hierfür eine flexible Alternative bzw. Ergänzung sein: Texte könnten automatisch mit diesen Kategorien annotiert werden. Es wäre zu prüfen, inwieweit hierfür auf die für PSYNDEX entwickelten automatischen Indexierungsverfahren (Weiland, Gerards & Gerards, 2007) aufgebaut werden kann.

Das Einbeziehen von Korpora aus anderen Quellen als PSYNDEX wäre auch eine vielversprechende Möglichkeit, besonders aktuelle Brennpunktthemen finden zu können. Statt bisher “Hot Topics” allein in PSYNDEX zu suchen, könnten Diskussion, Beiträge und Kommentare aus den digitalen Mitteilungen und sozialen Netzwerken von Psychologinnen und Psychologen, Fachgesellschaften- und Verbänden und anderen psychologierelevanten Kanälen analysiert werden. Dies würde eine Verbindung der Vorteile der beiden in Abschnitt 2.2 skizzierten Ansätze zur Erfassung von Forschungsthemen darstellen: Ein Überblick auf Niveau von Expertinnen und Experten (erfahrungsgeladete Bestimmung) wird automatisiert erstellt (datengeleitete Bestimmung). Zu prüfen wäre, ob kognitionspsychologisch orientierte Modelle von Expertenwissen (Reimann, 1998) eine Blaupause für die Formalisierung eines automatisierten Zusammentragens aktueller Themen der Psychologie darstellen könnten. In jedem Fall würde in der App passende Fachliteratur in PSYNDEX (und ggf. anderen psychologischen Fachdatenbanken) mit den aktuellen Themen aus den sozialen Netzwerken und digitalen Mitteilungen verknüpft werden.

8. Allgemeine Diskussion

In drei empirischen Studien wurde der Einsatz von Topic Modeling zur Identifikation von psychologischen Forschungsthemen erprobt. Anhand der Anwendungsszenarien “Brennpunktthemen”, “Verhältnis zwischen Themen der Gesellschaft und Themen psychologischer Fachliteratur” und “Forschungsinteressen von Psychologinnen und Psychologen” konnten die Stärken von Topic Modeling demonstriert werden. Die Erkenntnisse aus diesen Studie flossen schließlich in die Entwicklung einer App zur Exploration und Analyse von psychologischen Forschungsthemen ein. Im Folgenden wird der wissenschaftliche Ertrag dieser Dissertation zusammengefasst bevor Anknüpfungspunkte für weitere Forschung und für forschungsbasierte Infrastruktureinrichtungen in der Psychologie aufgezeigt werden. Nach einem Resümee zu Einschränkungen von Topic Modeling erfolgt abschließend eine zusammenfassende Bewertung mit einer Topic-Modeling-Entscheidungsheuristik.

8.1 Wissenschaftlicher Ertrag vor dem Hintergrund der Forschungsziele

Die erste übergeordnete Zielstellung der Dissertation betraf als “proof-of-concept” den Vergleich von Topic Modeling mit dem in der Psychologie etablierten klassifikationsbasierten Ansatz (Studie 1). Hierbei konnte gezeigt werden, dass Topic Modeling dem Klassifikationsansatz hinsichtlich Detailgrad und inhaltlicher Vielfalt der Themen überlegen ist. Es konnten Brennpunktthemen der Psychologie gefunden werden, die mit den Klassifikationskategorien übersehen worden wären. Außerdem erlaubt Topic Modeling, anhand der errechneten Dokument-Thema-Wahrscheinlichkeiten eine gezielte Recherche besonders repräsentativer Publikationen. Bei adäquater Modellwahl (Anzahl der Themen und ggf. LDA-Hyperparameter) und der Berücksichtigung von Themen-Validität und Themen-Reliabilität wird Topic Modeling daher als Methode der Wahl zur automatisierten Identifikation psychologischer Forschungsthemen erachtet.

Als zweites Ziel sollte geprüft werden, inwieweit ein Datenbank-Klassifikationssystem der Psychologie zur externen Validierung von Themen geeignet ist (Studien 1 und 2). Dies erfolgte anhand des PSYNDEX-Klassifikationssystems (ZPID, 2016). Dabei kann festgehalten werden, dass solche Datenbank-Klassifikationen geeignet sind, um einerseits die Themeninhalte zu bestätigen und andererseits auf die inhaltliche Bandbreite der Themen

hinzuweisen (etwa wenn ein Thema in einem Teil der entsprechenden Publikationen aus einer bestimmten Perspektive beleuchtet wird; vgl. Abschnitt 5.5). Das Klassifikationssystem wird daher – bei der Analyse von Publikationen der zugehörigen Datenbank – als ergänzendes Instrument zur Validierung der Themen empfohlen. Außerdem kann diese Art der externen Validierung potenziell in automatisierte Abläufe integriert werden (siehe Abschnitt 8.3).

Das dritte Forschungsziel betraf die Erprobung von Möglichkeiten zum Umgang mit Mehrsprachigkeit von psychologischen Fachpublikationen. Die Verwendung der standardisierten Schlagwörter in PSYINDEX (ZPID, 2016) ermöglichte, Publikationen ungeachtet der Dokumentsprache in die Analysen aufzunehmen. Dieses Vokabular von mehr als 6500 vollständig in einem Thesaurus definierten Begriffen eignen sich zur Untersuchung der Forschungsthemen und -trends in Korpora, die die gesamte Psychologie betreffen. In einem solchen Fall repräsentieren sie das, was Text Mining von Freitexten (wie in Abstracts) zum Ziel hat: Die Extraktion bedeutsamer Information aus Texten (Allahyari et al., 2017). Die Verwendung dieser standardisierten Schlagwörter stößt allerdings dann an ihre Grenzen, wenn das Korpus thematisch bereits sehr eingegrenzt ist (wie in Studie 2), oder wenn besonders aktuelle bzw. gerade erst entstehende Brennpunkthemen identifiziert werden sollen. In solchen Fällen sind Freitexte zu bevorzugen – anstelle oder ergänzend zu den standardisierten Schlagwörtern. Die in den Abschnitten 4.5 und 7.5.2 diskutierten Ansätze zum Einbezug von Beiträgen und Diskussionen aus sozialen Medien können hierfür eine Lösung bieten. Zur sprachlich einheitlichen Analyse solcher Texte könnte eine automatische Annotation mit Wikipedia-Kategorien (Zarrinkalam et al., 2019) eine inhaltlich umfassendere Alternative zu den standardisierten Schlagwörtern in PSYINDEX sein.

Das vierte Ziel galt der Entwicklung einer nutzerfreundlichen App zur Exploration und Analyse psychologischer Forschungsthemen. Die App soll außerdem die Möglichkeit bieten, möglichst einfach passende Literatur zu jedem Thema zu finden. Durch die Generierung von Themen aus den standardisierten Schlagwörtern ist es möglich, die Themenbegriffen direkt für Anfragen im Suchportal PubPsych.eu zu verwenden und entsprechende Nachweise in PSYINDEX anzeigen zu lassen. Das zugrundeliegende Topic Model wurde analog zum Vorgehen in Studie 3 berechnet, was unter anderem bedeutet, nur reliable Themen einzuschließen. Bei der Auswahl geeigneter Modellkandidaten erwies sich eine Kombination aus den statistischen Maßen “Semantischer Kohärenz” (Mimno et al., 2011) und “Exklusivität” (Roberts et al., 2014) als hilfreich für die Erzeugung gut interpretierbarer

Themen, weshalb dieser Ansatz gegenüber der Verwendung der Held-Out-Likelihood (wie bei Kosinski et al., 2016) empfohlen wird.

In der vorliegenden Arbeit wurden für zwei Aspekte Kriterien vorgeschlagen, für die es noch keine Standards in der Literatur gibt: (1) Die Bestimmung von Anteilen aus den Dokument-Thema-Wahrscheinlichkeiten und (2) die Größeneinordnung von Prävalenzdifferenzen. Sollte es nötig sein, Häufigkeiten aus den Dokument-Thema-Wahrscheinlichkeiten abzuleiten und Informationsverlust in Kauf zu nehmen (etwa bei der Bestimmung von Frauen- oder Männeranteilen an einem Thema), kann ein Dokument dann als zum Thema zugehörig gezählt werden, wenn das Thema eine Dokument-Thema-Wahrscheinlichkeit von mindestens 50 % aufweist (siehe PsychArchives-ESM 1 zu Studie 3 in Anhang D). Dieses Kriterium ist strenger als alternative Vorschläge (z. B. 20 % bei Schaal et al., 2019) und zählt dadurch nur diejenigen Publikationen, die sich hauptsächlich diesem Thema widmen. Andernfalls fließen auch Publikationen ein, die das Thema nur anschnitten. Für den zweiten Aspekt, der Einordnung von Prävalenzdifferenzen muss unterschieden werden, wie viele Gruppen bzgl. der Prävalenz verglichen werden. Bei zwei Gruppen wurde vorgeschlagen, die halbe durchschnittliche Prävalenz im Korpus als Schwellenwert zur Einordnung als "größere Differenz" zu verwenden (siehe Abschnitt 6.3.4). Bei mehr als zwei Gruppen sind inhaltliche Überlegungen ausschlaggebend, etwa ob angenommen werden kann, dass manche Gruppen eine Prävalenz von nahezu Null überhaupt aufweisen können (siehe Abschnitt 5.4)

8.2 Anknüpfungspunkte für forschungsbasierte Infrastruktureinrichtungen in der Psychologie am Beispiel des ZPID

In Abschnitt 7.5.2 wurden Möglichkeiten skizziert, die im Rahmen dieser Dissertation entstandene App weiterzuentwickeln. Diese betrafen Literaturvorschläge, die Darstellung der Anteile empirischer Evidenz zu den Themen sowie den Einbezug weiterer Textquellen wie Artikel aus Zeitungsmedien oder Beiträge von Psychologinnen und Psychologen in sozialen Netzwerken. Am Beispiel des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID) sollen im Folgenden damit einhergehende und weitere Anknüpfungspunkte für forschungsbasierte Infrastruktureinrichtungen in der Psychologie aufgezeigt werden. Dabei werden konkrete Vorschläge für Produkte des ZPID unterbreitet.

8.2.1 Informations- und Recherchedienste: Brennpunktthemen

Das ZPID führt (im Januar 2020) auf der PSYNDEX-Microsite eine Rubrik zu “Psychologischen Brennpunktthemen”⁵³ auf. Dabei werden vordefinierte Suchen in PSYNDEX zu besonders nachgefragten und aktuellen Themen angeboten: “Neue Kommunikationsmedien”, “Burnout und beruflicher Stress”, sowie “Feindseligkeit und Gewalt”. Diese Themen werden manuell von Mitarbeitenden zusammengestellt, was zeitintensiv in Erstellung und Aktualisierung ist. Die geplante Weiterentwicklung der App “PsychTopics” mit Einschluss weiterer Textquellen (z. B. soziale Medien und Websites; vgl. Abschnitt 7.5.2) bietet sich hierfür als Alternative an. Durch ein solches Zusammenfließen von Informationen könnte eine zentrale Anlaufstelle für aktuelle Themen der Psychologie geschaffen werden. Zu jedem Thema kann – wie in der aktuellen Version der App – passende Literatur in PSYNDEX (und prinzipiell auch in anderen Datenbanken) mit einem Klick angezeigt werden. Durch den Einbezug von Nachrichtenmedien könnte dies auch ein Ort sein, um der Frage nach Themen der Gesellschaft und ihrem Aufgreifen durch die Psychologie (im Sinne “gesellschaftlich relevanter Brennpunktthemen”) nachzugehen (vgl. Abschnitt 7.5.2)

Bei der Untersuchung weiterer Textquellen wie Diskussionen und Kommentaren können Methoden des Text Minings verwendet werden, die über Inhalte hinausgehen und auch Bedeutungen extrahieren. So kann mit Sentimentanalysen (Pang & Lee, 2008) der affektive Gehalt in Texten untersucht werden. Als aktuelles Beispiel hierfür könnte die Rezeption des Direktstudiums zur Ausbildung in der Psychotherapie in den sozialen Medien abgebildet werden. Das Informationsangebot würde dadurch nicht nur die Präsentation von psychologierelevanten Themen und dazu passende Literatur umfassen, sondern auch darauf hinweisen, welche Themen ggf. besonders kontrovers diskutiert werden.

Für die Zusammenstellung der einzelnen Korpora (Websites, Twitter-Kanäle usw.) könnte die Community aktiv einbezogen werden. Beispielhaft könnte geprüft werden, wie die Fachgruppen ihre “Hot Topics” für den 51. DGPs-Kongress (DGPs, 2018) identifiziert haben und ob dies auf einen Text-Mining-basierten Ansatz übertragbar ist. Von Bedeutung wäre es zu ermitteln, welche Quellen von den Fachgruppen genutzt wurden bzw. woraus sich das jeweilige Expertenwissen speist. Somit könnte die in Abschnitt 7.5.2 angesprochene

⁵³ <https://psyndex.de/index.php?wahl=PSYNDEX&uwahl=Brennpunktthemen> (16.01.2020)

Verknüpfung von erfahrungsgeliteter und automatisierter Bestimmung von Themen umgesetzt werden.

Methodisch wäre zunächst zu prüfen, ob für kurze Texte in soziale Medien spezielle Varianten von Topic Models (z. B. Shi, Cheng, Xie & Xie, 2019; Yan et al., 2013) eingesetzt werden müssen bzw. eine Aggregation von Texten nötig ist (Steinskog, Therkelsen & Gambäck, 2017), oder ob LDA/STM bei einer Extraktion der zentralen Konzepte der Texte zufriedenstellend kohärente Themen erzeugt (wie mit dem standardisierten Vokabular in den Studien 1 und 3). Im Falle der Tauglichkeit von LDA/STM wäre zu prüfen, ob es zielführend ist, separate Topic Models für jedes Korpus zu rechnen und ähnliche Themen aus den verschiedenen Korpora anschließend zu mappen. Alternativ könnten alle Quellen in ein Korpus zusammengeführt und mit der Quelle als Dokumentvariable (vgl. Studie 2) verglichen werden. Auch muss das Zeitintervall für neue Modellinferenzen festgelegt werden (also der Zeitpunkt, wann neue Themen identifiziert werden), da bei sozialen Medien von einer geringeren Halbwertszeit von Themen auszugehen ist als bei Fachliteratur. Entsprechend muss erprobt werden, was ein geeignetes Zeitfenster für ein Korpus aus sozialen Medien darstellt. Beispielsweise könnten die letzten zwölf Monate analysiert werden und ein neues Topic Model einmal im Monat gerechnet werden, wodurch Trends über ein Jahr abgebildet werden und monatlich die Chance besteht, neue Themen zu identifizieren. Ältere Themen, die in aktuellen Modellen nicht mehr enthalten sind, könnten archiviert werden.

8.2.2 PsychPorta: Literaturempfehlungen

Für die vom ZPID geplante semantische Suchplattform “PsychPorta” (Bosnjak & ZPID, 2018, S. 11) bietet Topic Modeling eine Möglichkeit zur Empfehlung relevanter Fachliteratur (Amami et al., 2016). Ausgehend von den thematischen Interessen einer Nutzerin bzw. eines Nutzers kann ein individuelles Profil gebildet werden. Dazu wird aus den bisherigen Publikationen der Person ein Korpus gebildet, aus dem mittels Topic Modeling die Forschungsinteressen abgeleitet werden (vgl. Studie 3). Alternativ oder zusätzlich könnten auch fremde Publikationen einbezogen werden. Die Wort-Thema-Wahrscheinlichkeiten der einzelnen Themen bzw. Interessen der Person werden dann mit statistischen Sprachmodellen aller Publikationen verglichen, das heißt, mit Verteilungen von Wortwahrscheinlichkeiten in den Publikationen. Überschreitet die Ähnlichkeit (Kullback-Leibler-Divergenz) zwischen den Forschungsinteressen-Wortverteilungen und der Publikations-Wortverteilungen einen Schwellenwert, wird die Publikation empfohlen. Dieser Ansatz ließe sich dahingehend

erweitern, dass anhand der gebildeten Profile ähnliche Forscherinnen und Forscher vorgeschlagen werden könnten – unter Berücksichtigung der diesbezüglichen datenschutzrechtlichen Möglichkeiten. Außerdem wäre es möglich, eine “Alert-Funktion” einzurichten, mit der die Nutzerin oder der Nutzer informiert wird, sobald eine neue passende Publikation in der Datenbank verfügbar ist. Dies wäre beispielsweise für Forschende von Interesse, die sich mit Forschungssynthesen befassen: Es wird umgehend über potenzielle neue Evidenz informiert, die in die Synthese aufgenommen werden könnte.

8.2.3 PsychAuthors: Autorenprofile und -netzwerke

Die Idee, Forschungsinteressen aus den Themen der Publikationen abzuleiten, kann auch auf die Autorenplattform “PsychAuthors” (www.psychauthors.de) des ZPID übertragen werden. Hierbei handelt es sich um eine Datenbank mit Steckbriefen von forschungsaktiven Psychologinnen und Psychologen aus dem deutschsprachigen Raum. Mithilfe eines *Author-Topic Models* (ATM; Rosen-Zvi et al., 2004) könnten die Profile aller Autorinnen und Autoren der Datenbank automatisch erstellt werden. Dadurch realisierbare neue Funktionen für PsychAuthors wären beispielsweise das Aufzeigen von forschungsaktiven Expertinnen und Experten zu den jeweiligen Themen (Tu, Johri, Roth & Hockenmaier, 2010), der Anzeige der am meisten bzw. am wenigsten überraschenden Publikationen einer Person (Steyvers, Smyth, Rosen-Zvi & Griffiths, 2004) oder der Veränderung der Forschungsschwerpunkte über die Zeit (Xu et al., 2014). Gerade für die Darstellung des Lebenswerks könnten somit tiefere Einblicke in die Forschungsbiografie ermöglicht werden. Daneben wäre es möglich, themenbasierte Autorennetzwerke zu bilden (Mao, Cao, Lu & Li, 2017). Ergänzend zu Koautoren- oder Zitationsnetzwerken böten diese die Möglichkeit, themenbezogene oder -übergreifende Zusammenarbeiten aufzudecken. Damit könnte beispielsweise aufgezeigt werden, welche Arbeitsgruppen ggf. unabhängig und ohne Kenntnis voneinander am gleichen Thema arbeiten oder welche Themen besonders interdisziplinär behandelt werden. Eine Erweiterung von PsychAuthors um diese Funktionen könnte diese Plattform für neue Autorinnen und Autoren attraktiv machen.

8.3 Einschränkungen und Implikationen für weitere Forschung

Ein Ziel dieser Dissertation war zu überprüfen, inwiefern ein Klassifikationssystem psychologischer Fachliteratur (Tuleya, 2007; ZPID, 2016) zur externen Validierung geeignet ist. Auch wenn dies in zwei Studien prinzipiell gezeigt werden konnte (sensu Griffiths &

Steyvers, 2004), bleibt eine Quantifizierung der Übereinstimmung von Themeninhalten und Klassifikationskategorien ein Anliegen zukünftiger Forschung. Wie in Abschnitt 4.5 beschrieben, könnte die Übereinstimmung mittels Kolmogorov-Smirnov-Anpassungstest untersucht werden (Gleichverteilung auf alle Kategorien versus Ungleichverteilung). Durch eine auf der Kolmogorov-Smirnov-Teststatistik aufbauende Quantifizierung der externen Validität könnte spezifisch für die Psychologie ein zusätzliches Maß der Themenvalidität gebildet werden, das zusätzlich zu den Vorschlägen von Maier et al. (2018) im Kontext psychologischer Forschungsthemen verwendet werden kann. Für die entwickelte App (Abschnitt 7), für die Automatisierung von besonderer Bedeutung ist, würde das die Möglichkeit eröffnen, auch für ein Topic Model mit vielen Themen automatisch die Themen mit niedriger externer Validität herauszufiltern. Diese könnten anschließend einer manuellen Prüfung unterzogen werden, um eine Entscheidung über Verbleib oder Entfernen (Maier et al., 2018) zu treffen und die Qualität der App weiter zu erhöhen.

Für den Umgang mit Mehrsprachigkeit wissenschaftlicher Texte erwies sich für die PSYNDEX-Datenbank die Verwendung des standardisierten Vokabulars von Schlagworten (ZPID, 2016) als eine geeignete Möglichkeit. Allerdings stellt dies keine Lösung für die Analyse von Abstracts und Titeln dar (vgl. Studie 2). Für diese Textarten wäre zu prüfen, ob automatische Übersetzungen mit neuronalen Übersetzungsnetzen (wie z. B. von www.DeepL.com) zu interpretierbaren und validen Themen führen. Denkbar wäre, diese Übersetzungen mit den standardisierten Schlagwörtern zu kombinieren, um ggf. Fehlübersetzungen der neuronalen Netze zu kompensieren. Eine automatische Annotation mit Wikipedia-Kategorien (Zarrinkalam et al., 2019) könnte eine Alternative zu den standardisierten Schlagworten sein, vor allem wenn das Korpus bereits thematisch eingegrenzt ist (wie in Studie 2).

In allen Studien wurden zeitliche Trends der Themenprävalenz untersucht. Dabei wurde angenommen, dass die Themeninhalte über alle Jahre stabil bleiben, bzw. stets die gleichen Wörter für ein Thema relevant sind. Ob dies der Fall ist, könnte mit *Dynamic Topic Modeling* (DTM; Blei & Lafferty, 2006) überprüft werden. Bei dieser Topic-Modeling-Variante können sich Rangfolgen der Wörter der Themen (die Wort-Thema-Wahrscheinlichkeiten) über verschiedene Zeitabschnitte ändern, sofern sich dies aus den Publikationen ergibt. Beispielsweise könnte bzgl. Studie 2 DTM Aufschluss darüber geben, ob "Traumatisierung bei Geflüchteten" ein eigenständiges Thema ist, oder sich aus einem eher allgemeinen Thema

zu Geflüchteten heraus entwickelte. Dabei muss beachtet werden, dass DTM nur für eine geringe Anzahl an Themen im Modell geeignet ist (Wiedemann & Niekler, 2018, S. 12).

Die drei Studien konzentrierten sich auf den deutschsprachigen Raum. Da PSYINDEX strukturell kompatibel mit der APA-Fachdatenbank PsycInfo ist (Bosnjak & ZPID, 2018, S. 12), könnten die durchgeführten Analysen mit einem PsycInfo-Korpus wiederholt werden: Beide Datenbanken annotieren Metadaten entsprechend des APA-Thesaurus bzw. APA-Klassifikationssystems (Tuleya, 2007, ZPID, 2016). In der Tradition der szientometrischen Arbeiten von Krampen (z. B. Krampen et al., 2011; Krampen & Perrez, 2015) wären vergleichende Untersuchungen möglich, wie etwa eine Gegenüberstellung der “Hot Topics” aus dem angloamerikanischen Raum und den entsprechenden Brennpunktthemen aus den deutschsprachigen Ländern. Zu prüfen wäre jedoch, inwiefern sich die Qualitätsniveaus der beiden Datenbanken entsprechen, etwa ob bei PsycInfo die Metadaten maschinell oder manuell vergeben werden. Dies wäre bei einem Vergleich insofern zu berücksichtigen, als dass aktuelle Studien darauf hinweisen, dass die automatische Zuordnung von Publikationen zu Klassifikationssystemen, die mit viele Kategorien umfassen, selbst mit modernen Verfahren nicht immer zufriedenstellend gelingt (Blanck, Niekler & Kaulisch, 2019).

Topic Modeling ist in automatisierte Abläufe integrierbar. Dies ermöglicht aktuelle und umfassende Informationsangebote, wie etwa die im Rahmen dieser Dissertation entwickelte App “PsychTopics” (Abschnitt 7). Dabei kommen prinzipiell Textquellen jeder Art in Frage, also auch unstrukturierte oder kürzere Texte (z. B. Tweets). Ein adäquates Aufbereiten solcher Texte vorausgesetzt, kann so in Kombination mit *Web Scraping* bzw. *Web Mining* (Cooley, Mobasher & Srivastava, 1997) ein umfassender Überblick über die psychologische Forschungslandschaft gewonnen werden, um Antworten auf Fragen wie die folgenden zu geben: “Können zukünftige Publikationstrends aus aktuellen Diskussionen in sozialen Netzwerken abgeleitet werden?” (Zarrinkalam et al., 2019), “Schlagen sich DGPs-Empfehlungen in der Fachliteratur nieder?” (analog zu Anderson, Rowley, Stegenga, Irvin & Rosenberg, 2019), “Welche Arbeitsgruppen befassen sich (ggf. unabhängig voneinander) mit welchem Thema” (Mao et al., 2017) usw. Für solche und ähnliche Fragestellungen sind automatisierte Informationsangebote denkbar, die bei leicht verständlicher und nutzerfreundlicher Aufbereitung ein breites Publikum ansprechen können (vgl. Abschnitt 8.2.1).

8.4 Abschließende Bewertung und eine Topic-Modeling-Entscheidungsheuristik

Zusammenfassend kann Topic Modeling als geeignete Methode zur Identifikation von Forschungsthemen in der Psychologie bewertet werden. Vor allem um bisher unbekannte und neue Themen in großen Textkorpora automatisiert finden zu können, wird dieses Verfahren aktuell als die Methode der Wahl und klassifikationsbasierten Ansätzen als überlegen erachtet. Die drei Studien im Rahmen dieser Dissertation geben exemplarisch Einblick in szientometrische Anwendungsmöglichkeiten. In Kombination mit anderen Text-Mining-Methoden, welche beispielsweise den affektiven Gehalt (Pang & Lee, 2008) oder die kontextabhängige Bedeutung von Wörtern berücksichtigen (Moody, 2016), lassen sich noch anspruchsvollere Fragestellungen und umfassendere digitale Produkte denken. Auch jenseits von szientometrischen Fragestellungen sind für die Psychologie vielfältige Einsatzmöglichkeiten denkbar und bereits erprobt worden (vgl. Abschnitt 2.3).

Neben den Stärken der Methode bei einer bei adäquaten Anwendung sind auch Worte der Vorsicht angebracht. Die Interpretation von Topic-Modeling-Ergebnissen erfordert ein Umdenken im Vergleich zu frequentistischen Verfahren. Entsprechend des Bayes'schen Ansatzes werden passende Dokumente zu den Themen mit Wahrscheinlichkeiten geschätzt, keine Anzahlen angegeben. Wie bei allen unüberwachten Verfahren aus dem maschinellen Lernen sind die Ergebnisse nicht automatisch sinnvoll: "Fitting a topic model to a collection will yield patterns within the corpus whether or not they are 'naturally' there" (Blei & Lafferty, 2009, S. 17). Daher raten Grimmer und Stewart (2013, S. 270) ausdrücklich: "Validate, validate, validate" (vgl. Abschnitt 2.2.5). Doch selbst eine erfolgreiche Validierung garantiert nicht, dass die Themeninhalte in jedem Fall intersubjektiv eindeutig interpretierbar sind. Die Themen bei Topic Modeling sind latent, das heißt, sie sind nicht direkt beobachtbar, sondern werden aus einer Gruppe kookkurierender Wörter geschlossen – und zwar von Menschen in Abhängigkeit ihrer Korpuskenntnis und ihres Verständnisses der Methode. Das standardisierte Vokabular, das die Themen in der entwickelten App "PsychTopics" bildet (siehe Abschnitt 7.3.1), ist ein Versuch, durch Thesaurusdefinitionen den Interpretationsspielraum zu verringern.

Transparentes methodisches Vorgehen, etwa durch freie Offenlegung des Analysecodes, wird als wesentlich erachtet, um die Ergebnisse von Topic Modeling beurteilen zu können. Dem explorativen Charakter des Verfahrens entsprechend (Blei & Lafferty, 2009, S. 17) sollte Topic Modeling nicht unmittelbar für Evaluierungen von Forschungsleistungen missbraucht werden (Hecking & Leydesdorff, 2019), etwa indem sinkende Trends der Themenprävalenz unmittelbar als Leistungsindikatoren verstanden werden. Vielmehr sind die probabilistischen Angaben als *Indizien* zu verstehen, die eine eingehendere Prüfung nahelegen. Konkret würde eine gezielte Suche nach einem definierten Thema, welches in einem Topic Model gefunden wurde, zu Publikationshäufigkeiten und greifbareren Ergebnissen führen (Häufigkeiten statt Wahrscheinlichkeiten).

Wie Grimmer und Stewart (2013, S. 270) betonen, gibt es keine allgemein beste Methode zur automatischen Textanalyse. Selbst wenn die automatische Identifikation von Forschungsthemen Ziel der Untersuchung ist, haben das jeweilige Korpus und die jeweiligen Fragestellungen Implikationen für die Wahl eines geeigneten Verfahrens innerhalb der Vielfalt von Topic-Modeling-Varianten. In Ergänzung zu derzeitigen Best-Practice-Empfehlungen (Banks et al., 2018; Maier et al., 2018) soll daher eine Entscheidungsheuristik zur Auswahl der geeigneten Topic-Modeling-Variante Orientierung bieten. Abbildung 19 stellt grafisch dar, welche zentralen Fragen zu beantworten sind und welche Verfahren in Betracht kommen, wenn die Identifikation von Forschungsthemen das Ziel der Untersuchung ist. Die Heuristik erhebt keinen Anspruch auf Vollständigkeit, da die Zahl an Topic-Modeling-Varianten sehr groß ist und ihre jeweiligen Ausrichtungen sehr spezifisch sind. Die Auswahl erfolgt nach den etabliertesten Varianten mit verfügbaren Software-Paketen in R oder Python. Außerdem sollen die typischsten Szenarien szientometrischer Untersuchungen abgedeckt sein. Analog zu Ockhams Rasiermesser liegt der Heuristik die Idee zugrunde, bei gleicher Eignung das einfachere Verfahren zu verwenden. Daher ist markiert, in welchen Sonderfällen LDA auch für Fragestellungen mit Kovariaten möglich ist.

Kann Topic Modeling den in Abschnitt 1 beschriebenen Beitrag zum Wissenstransfer leisten? Um potenzielle Transferbereiche aufzudecken, deren Existenz noch nicht bekannt war, oder um die thematische Vielfalt bekannter Bereiche auszuloten (vgl. Studie 2), wird diese Methode als geeigneter Ansatz erachtet. Vor allem wenn die Menge auszuwertender Texte den Rahmen manueller Analysen sprengt, spielt dieses Verfahren der automatischen Textanalyse seine Stärken aus. Transparentes Vorgehen, Korpuskenntnis und Gewährsein der

Grenzen vorausgesetzt, kann Topic Modeling dabei unterstützen, wissenschaftliche Evidenz zu identifizieren, die für ein Übertragen in Gesellschaft, Politik, Kultur oder Wirtschaft in Frage kommt.

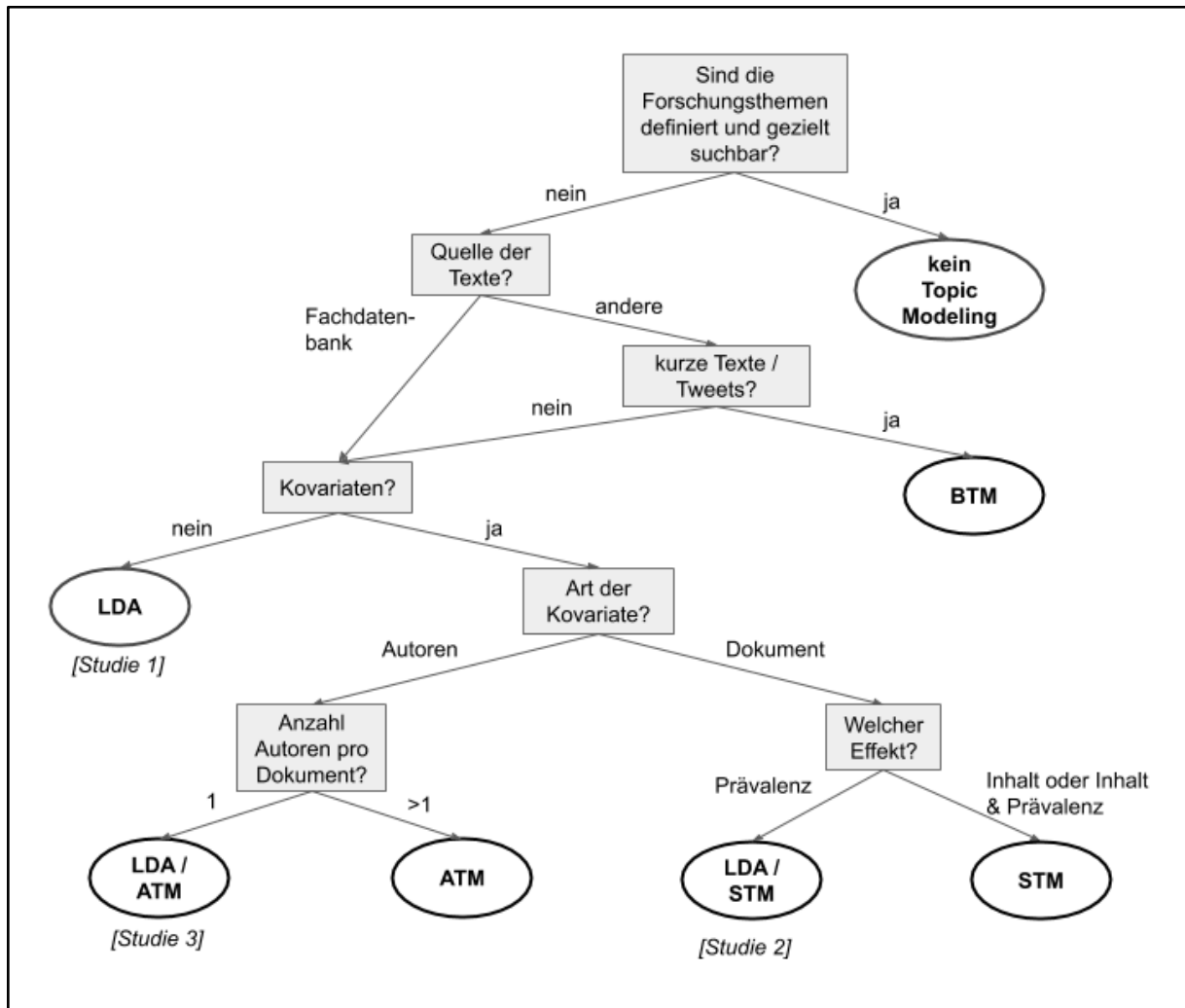


Abbildung 19. Heuristischer Entscheidungsbaum zur Wahl einer geeigneten Topic-Modeling-Variante bei der Identifikation von Forschungsthemen. Soziale Medien wie Twitter können eingeschlossen werden, falls dabei Forschende bzw. forschungsrelevante Stakeholder untersucht werden und damit Forschungsthemen ermittelt werden können. ATM = Author-Topic Model (Rosen-Zvi et al., 2004), BTM = Biterm Topic Model (Yan et al., 2013), LDA = Latent Dirichlet Allocation (Blei et al., 2003), STM = Structural Topic Model (Roberts et al., 2013).

Literaturverzeichnis

- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B. & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint*. <https://arxiv.org/abs/1707.02919v2>
- Amami, M., Pasi, G., Stella, F. & Faiz, R. (2016, June). An LDA-based approach to scientific paper recommendation. In E. Métais, F. Meziane, M. Saraee, V. Sugumaran & S. Vadera (Hrsg.) *International conference on applications of natural language to information systems* (S. 200–210). Cham: Springer. Verfügbar unter <https://boa.unimib.it/retrieve/handle/10281/137815/195091/nldb%20article.pdf>
- Anderson, D., Rowley, B., Stegenga, S., Irvin, P. S. & Rosenberg, J. M. (2019, April). *Evaluating Content-Related Validity Evidence Using Text Modeling*. Paper presented at National Council on Measurement in Education Annual Meeting, Toronto, Canada. Verfügbar unter <https://www.datalorax.com/talks/ncme19/#1>
- Antoni, C. H. (2019). Zur Lage der Psychologie. *Psychologische Rundschau*, 70(1), 4–26. <https://doi.org/10.1026/0033-3042/a000429>.
- Arik, E. (2013). A bibliometric analysis of a national Journal: The case of the Turkish Journal of Psychology. *Journal of Scientometric Research*, 2(3), 173–184. <https://doi.org/10.4103/2320-0057.135407>
- Arora, S., Ge, R., Halpern, Y., Mimno, D., Moitra, A., Sontag, D., Wu, Y. & Zhu, M. (2013). A Practical Algorithm for Topic Modeling with Provable Guarantees. In S. Dasgupta & D. McAllester (Hrsg.), *Proceedings of the 30th International Conference on Machine Learning*, (S. 280–288). Atlanta, GA: PMLR. Verfügbar unter <http://proceedings.mlr.press/v28/arora13.html>
- Atkins, D. C., Rubin, T. N., Steyvers, M., Doeden, M. A., Baucom, B. R. & Christensen, A. (2012). Topic models: A novel method for modeling couple and family text data. *Journal of Family Psychology*, 26 (5), 816–827. <https://doi.org/10.1037/a0029607>
- Awati, K. (2015, 29. September). *A gentle introduction to topic modeling using R* [Blog post]. Verfügbar unter <https://eight2late.wordpress.com/2015/09/29/a-gentle-introduction-to-topic-modeling-using-r/>
- Banks, G. C., Woznyj, H. M., Wesslen, R. S. & Ross, R. L. (2018). A review of best practice recommendations for text analysis in R (and a user-friendly app). *Journal of Business and Psychology*, 33(4), 445–459. <https://doi.org/10.1007/s10869-017-9528-3>
- Baumann, U. (1995). Bericht zur Lage der deutschsprachigen Psychologie 1994. Fakten und Perspektiven. *Psychologische Rundschau*, 46(1), 3–17.
- Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York, NY: Plenum Press.

- Bikienga, S. (2018). *Topic models: from LCA to LDA*. Verfügbar unter https://github.com/Salfo/topic-models/blob/master/introduction_to_topic_models.pdf
- Bittermann, A. (2018). *Extracting latent topics in large text corpora by taking metadata into account: Do empirical studies address psychological research topics differently than nonempirical studies do?* Paper presented at Big Data in Psychology 2018 conference, Trier, Germany. <https://dx.doi.org/10.23668/psycharchives.837>
- Blanck, S., Niekler, A. und Kaulisch, M. (2019). Augmenting a Research Information System with automatically acquired category and keyword information. In G. Catalano, C. Daraio, M. Gregori, H. F. Moed & G. Ruocco (Hrsg.), *Proceedings of the 17th Conference of the International Society for Scientometrics and Informetrics (2002–2013)*. Rom: Edizioni Efesto. Verfügbar unter http://issi-society.org/proceedings/issi_2019/ISSI%202019%20-%20Proceedings%20VOLUME%20II.pdf
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55, 77–84. <https://doi.org/10.1145/2133806.2133826>
- Blei, D. M. & Lafferty, J. D. (2006, Juni). Dynamic topic models. In W. W. Cohen & A. Moore (Hrsg.), *Proceedings of the 23rd international conference on Machine learning* (S. 113–120). Madison, WI: ACM. <https://doi.org/10.1145/1143844.1143859>
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 1, 17–35. <https://doi.org/10.1214/07-AOAS114>
- Blei, D. M., Lafferty, J. D. (2009) Topic models. In A. N. Srivastava & M. Sahami (Hrsg.), *Text mining: classification, clustering, and applications* (S. 71–94). Boca Raton, FL: CRC Press. Verfügbar unter <http://www.cs.columbia.edu/~blei/papers/BleiLafferty2009.pdf>
- Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Bosnjak, M. & ZPID (Leibniz Institute For Psychology Information). (2018). *ZPID-Jahresbericht 2018*. ZPID (Leibniz Institute for Psychology Information). <https://doi.org/10.23668/psycharchives.2592>
- Brack, K., Reinhardt, S., Dahme, B. & Hoffmann, B. (1997). Gibt es geschlechtsspezifische Präferenzen in der psychologischen Forschung? – Eine Untersuchung psychologischer Doktorarbeiten in der alten Bundesrepublik der Jahrgänge 1986–1990. *Psychologische Rundschau*, 48(2), 101–105.
- Burns, P. B., Rohrich, R. J. & Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic and reconstructive surgery*, 128(1), 305–310. <https://doi.org/10.1097/PRS.0b013e318219c171>

- Cascio, W. F. & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, 93(5), 1062–1081. <http://dx.doi.org/10.1037/0021-9010.93.5.1062>
- Chaney, A. J. B. & Blei, D. M. (2012, März). Visualizing topic models. *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (IWSCM)*. Verfügbar unter <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewFile/4645/5021>
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L. & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams & A. Culotta (Hrsg.), *Advances in neural information processing systems 22 (NIPS 2009)* (S. 288–296). La Jolla, CA: Neural Information Processing Systems. Verfügbar unter <http://users.umiacs.umd.edu/~jbg/docs/nips2009-rtl.pdf>
- Chang, W., Cheng, J., Allaire, J. J., Xie, Y. & McPherson, J. (2018). *shiny: Web Application Framework for R*. R package version 1.2.0. Verfügbar unter <https://cran.r-project.org/web/packages/shiny/shiny.pdf>
- Chen, E. (2011, 22. August). *Introduction to Latent Dirichlet Allocation* [Blog post]. Verfügbar unter <https://blog.echen.me/2011/08/22/introduction-to-latent-dirichlet-allocation/>
- Cooley, R., Mobasher, B. & Srivastava, J. (1997). Web mining: Information and pattern discovery on the World Wide Web. In P. Storms (Hrsg.), *Proceedings of the Ninth IEEE International Conference on Tools With Artificial Intelligence* (S. 558–567). Los Alamitos, CA: IEEE Computer Society. <https://doi.org/10.1109/TAI.1997.632228>
- Coppersmith, G., Dredze, M. & Harman, C. (2014). Quantifying mental health signals in Twitter. In P. Resnik, R. Resnik, & M. Mitchell (Hrsg.), *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (S. 51–60). Stroudsburg, PA: Association for Computational Linguistics. Verfügbar unter <http://www.aclweb.org/anthology/W14-3207>
- Coppola, A., Roberts, M. E., Stewart, B. & Tingley, D. (2016). *stmCorrViz: A Tool for Structural Topic Model Visualizations*. R package version 1.3. Verfügbar unter <https://cran.r-project.org/web/packages/stmCorrViz/stmCorrViz.pdf>
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6), 391–407. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6%3C391::AID-ASII%3E3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASII%3E3.0.CO;2-9)
- De Saussure, F. (2011). *Course in general linguistics*. Eds. Charles Bally & Albert Sechehaye. Trans. Wade Baskin, subsequently edited by Perry Meisel & Haun Saussy. NY: Columbia University Press. (Originalausgabe: Wade Baskin. NY: The Philosophical Society, 1959)
- DGPs – Deutsche Gesellschaft für Psychologie. (2018). *"Hot Topics" auf dem DGPs-Kongress 2018* [Website]. Abgerufen von https://www.dgpskongress.de/frontend/index.php?folder_id=133 (17.09.2019), archiviert unter <https://bit.ly/30wkJQX>

- Dudenredaktion (o. J.): "Thema" auf Duden online. Verfügbar unter <https://www.duden.de/rechtschreibung/Thema>
- Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3), 32–57. <https://doi.org/10.1080/01969727308546046>
- Egami, N., Fong, C. J., Grimmer, J., Roberts, M. E. & Stewart, B. M. (2018). How to make causal inferences using texts. *arXiv preprint*. <https://arxiv.org/abs/1802.02163v1>
- Erdfelder, E. & Bošnjak, M. (2016). "Hotspots in Psychology": A new format for special issues of the Zeitschrift für Psychologie. *Zeitschrift für Psychologie*, 224(3), 141–144. <https://doi.org/10.1027/2151-2604/a000249>
- Evans, M. S. (2014). A computational approach to qualitative analysis in large textual datasets. *PLoS ONE* 9 (2): e87908. <https://doi.org/10.1371/journal.pone.0087908>
- Feldman, R. & Dagan, I. (1995, August). Knowledge Discovery in Textual Databases (KDT). In U. Fayyad & R. Uthurusamy (Hrsg.), *Proceedings of the First International Conference on Knowledge Discovery and Data Mining* (S. 112–117). AAAI Press. Verfügbar unter <https://www.aaai.org/Papers/KDD/1995/KDD95-012.pdf>
- Fell, C. B. & König, C. J. (2016). Is there a gender difference in scientific collaboration? A scientometric examination of co-authorships among industrial–organizational psychologists. *Scientometrics*, 108(1), 113–141. <https://doi.org/10.1007/s11192-016-1967-5>
- Fuchs, R. & Appel, E. (1994). Belastungsregulation durch Sport. In R. Schwarzer & M. Jerusalem (Hrsg.), *Gesellschaftlicher Umbruch als kritisches Lebensereignis. Psychosoziale Krisenbewältigung von Übersiedlern und Ostdeutschen* (S. 227–240). Weinheim: Juventa.
- Gardner, M. J., Lutes, J., Lund, J., Hansen, J., Walker, D., Ringger, E. & Seppi, K. (2010). The topic browser: An interactive tool for browsing topic models. In *NIPS Workshop on Challenges of Data Visualization (Vol. 2)*. Canada: Whistler. Verfügbar unter <https://pdfs.semanticscholar.org/47ae/e7b0ec2aeb64ce2dcb46b00b5d1b6969a9d4.pdf>
- Gerards, A., Gerards, M., Kuhberg-Lasson, V., Singleton, K. & Trierweiler, L. (2014). *Dokumentationsmanual PSYINDEX* (aktualisierte Ausgabe). Trier: ZPID. Verfügbar unter: <https://www.psyindex.de/download/PSYINDEXmaterial/psyindex-dokumentationsmanual.pdf>
- Gnambs, T., Hanfstingl, B. & Leidenfrost, B. (2006). Ist Methodenlehre männlich und Entwicklungspsychologie weiblich? Geschlechtsspezifische Präferenzen bei Abschlussarbeiten in der Psychologie. In B. Gula, R. Alexandrowicz, S. Strauß, E. Brunner, B. Jenull-Schiefer & O. Vitouch (Hrsg.), *Perspektiven psychologischer Forschung in Österreich. Proceedings zur 7. Wissenschaftlichen Tagung der Österreichischen Gesellschaft für Psychologie* (S. 374–381). Pabst: Lengrich. Verfügbar unter <https://timo.gnambs.de/sites/default/files/gnambstimo/publications/gnambs-2006.pdf>

- Goldstone, A., Galán, S., Lovin, C. L., Mazzaschi, A. & Whitmore, L. (2014). An interactive topic model of signs. *Signs at 40*. Verfügbar unter <http://signsat40.signsjournal.org/topic-model/>
- Griffiths, T. L. & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101 (Suppl. 1), 5228–5235. <https://doi.org/10.1073/pnas.0307752101>
- Griffiths, T. L., Steyvers, M. & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114(2), 211–244. <https://doi.org/10.1037/0033-295x.114.2.211>
- Grimmer, J. (2011). An introduction to Bayesian inference via variational approximations. *Political Analysis*, 19(1), 32–47. <https://doi.org/10.1093/pan/mpq027>
- Grimmer, J. & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297. <https://doi.org/10.1093/pan/mps028>
- Guille, A. & Soriano-Morales, E. P. (2016, January). TOM: A library for topic modeling and browsing. In EGC (S. 451-456). Verfügbar unter <https://pdfs.semanticscholar.org/d96b/530c9074fc6a2bf3b08da2df5be5b7cfd89c.pdf>
- Gundlach, G., Tröster, H. & Moschner, B. (1999). Sind Psychologiestudentinnen wissenschaftsfeindlich? *Psychologie in Erziehung und Unterricht*, 46(1), 29–40.
- Harris, Z. (1954). Distributional structure. *Word*, 10(23), 146–162. <https://doi.org/10.1080/00437956.1954.11659520>
- Hartmann, B. (2012). *Sind Psychologiestudentinnen wissenschaftsfeindlich? Geschlechtsunterschiede im Bezug auf studienrelevante und berufsrelevante Faktoren* (Diplomarbeit). Verfügbar unter: http://othes.univie.ac.at/21135/1/2012-02-03_9902182.pdf
- Hecking, T. & Leydesdorff, L. (2019). Can topic models be used in research evaluations? Reproducibility, validity, and reliability when compared with semantic maps. *Research Evaluation*, 28(3), 263–272. <https://doi.org/10.1093/reseval/rvz015>
- Hilbert, M. & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025), 60–65. <https://doi.org/10.1126/science.1200970>
- Hildebrand, K. (11. Dezember 2015). Wort des Jahres: Warum "Flüchtlinge" abschätzig ist. *Süddeutsche Zeitung*. Verfügbar unter <https://www.sueddeutsche.de/kultur/deutsche-sprache-wort-des-jahres-warum-fluechtlinge-abschaetzig-ist-1.2778193>
- Hofmann, T. (1999, Juli). Probabilistic latent semantic analysis. In K. B. Laskey & H. Prade (Hrsg.), *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence* (S. 289–296). San Francisco, CA: Morgan Kaufmann Publishers Inc. Verfügbar unter <https://dl.acm.org/doi/pdf/10.5555/2073796.2073829?download=true>
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. & Yasmeen, F (2019). *forecast: Forecasting functions for*

- time series and linear models*. R package version 8.5. Verfügbar unter <http://pkg.robjhyndman.com/forecast>
- Ioannidis, J. P., Fanelli, D., Dunne, D. D. & Goodman, S. N. (2015). Meta-research: evaluation and improvement of research methods and practices. *PLoS biology*, *13*(10), e1002264. <https://dx.doi.org/10.1371/journal.pbio.1002264>
- Jacobi, C., Van Atteveldt, W. & Welbers, K. (2015). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, *1*, 89–106. <https://doi.org/10.1080/21670811.2015.1093271>
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, *78*(11), 15169–15211. <https://doi.org/10.1007/s11042-018-6894-4>
- Jiang, H., Qiang, M. & Lin, P. (2016). A topic modeling based bibliometric exploration of hydropower research. *Renewable and Sustainable Energy Reviews*, *57*, 226–237. <https://doi.org/10.1016/j.rser.2015.12.194>
- Johnston, B., Jones, A. & Kruger, C. (2019). *Applied Unsupervised Learning with Python: Discover hidden patterns and relationships in unstructured data with Python*. Birmingham: Packt Publishing Ltd.
- Jones, M. M., Manville, C. & Chataway, J. (2017). Learning from the UK's research impact assessment exercise: a case study of a retrospective impact assessment exercise and questions for the future. *The Journal of Technology Transfer*, 1–25. <https://doi.org/10.1007/s10961-017-9608-6>
- Kavvadias, S., Drosatos, G. & Kaldoudi, E. (2019). An Online Service for Topics and Trends Analysis in Medical Literature. In *World Congress on Medical Physics and Biomedical Engineering 2018* (S. 481–485). Springer, Singapore. https://doi.org/10.1007/978-981-10-9035-6_89
- King's College London and Digital Science. (2015). *The nature, scale and beneficiaries of research impact: An initial analysis of Research Excellence Framework (REF) 2014 impact case studies*. Bristol: HEFCE. Verfügbar unter <https://www.kcl.ac.uk/policy-institute/assets/ref-impact.pdf>
- Kleine, D. & Fuchs, R. (1994). Streß, sportliche Aktivität und Wohlbefinden. In D. Alfermann, & V. Scheid (Hrsg.), *Psychologische Aspekte von Sport und Bewegung in Prävention und Rehabilitation. Bericht über die Tagung der Asp vom 3. bis 5. September 1992 in Gießen* (S. 47–53). Köln: bps-Verlag.
- Kleiner, M. (Hrsg.). (2019). *Leibniz digital – Werkstattberichte*. Berlin: Leibniz-Gemeinschaft. Verfügbar unter: https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/Forschung/Open_Science/Leibniz_digital/Leibniz_digital_-_Brosch%C3%BCre_gesamt.pdf
- Klostermann, A. (2016, 13. September). *DGPs-Kongress 2016: Wissen schafft Verstehen – Psychologie im Dienste der Gesellschaft* [Blog post]. Verfügbar unter

https://www.dgps.de/index.php?id=143&tx_ttnews%5Btt_news%5D=1730&cHash=8813dd840f0d523b8279e00c3faf03b0

- König, C. J., Fell, C. B., Kellnhöfer, L. & Schui, G. (2015). Are there gender differences among researchers from industrial/organizational psychology? *Scientometrics*, 105(3), 1931–1952. <https://doi.org/10.1007/s11192-015-1646-y>
- Kosinski, M., Wang, Y., Lakkaraju, H. & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, 21(4), 493–506. <https://doi.org/10.1037/met0000105>
- Krampen, G. (2015). *Aktuelle Entwicklungen des Faches Psychologie: Versuch einer Bestandsaufnahme mit Perspektiven*. Vortrag auf der Fachtagung der Fachreferentinnen und Fachreferenten der Psychologie des Vereins Deutscher Bibliothekare (VDB), Leibniz-Zentrum ZPID, 30.09.–01.10.2015. Verfügbar unter https://www.zpid.de/pub/research/2015_krampen_VDB-Trier.pdf
- Krampen, G., Montada, L., Müller, M. & Schui, G. (2004). Erfahrungen und Empfehlungen zur Internationalisierung der psychologischen Forschung im deutschsprachigen Bereich - Befunde aus Experteninterviews: Ausführlicher Bericht. *ZPID Science Information Online*, 4 (2). Verfügbar unter: <http://www.zpid.de/pub/info/experteninterviews.pdf>
- Krampen, G. & Perrez, M. (2015). Publikationsschwerpunkte der Klinischen Psychologie und Psychotherapieforschung im deutsch-versus angloamerikanischen Bereich. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 44, 181–196. <https://doi.org/10.1026/1616-3443/a000320>.
- Krampen, G., von Eye, A. & Schui, G. (2011). Forecasting trends of development of psychology from a bibliometric perspective. *Scientometrics*, 87(3), 687–694. <https://doi.org/10.1007/s11192-011-0357-2>
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70). Verfügbar unter <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Leyendecker, B. (2012). Familie und Integration: Bilingualität als ein wichtiger Schlüssel zur Integration von Kindern in Familie und Gesellschaft. In M. Krüger-Potratz, H. H. Reich (Hrsg.), *Familien- und Jugendpolitik in der Einwanderungsgesellschaft. Akzente, Analysen, Aktionen* (S. 31–42). Göttingen: V & R Unipress.
- Lindahl, J., Stenling, A., Lindwall, M. & Colliander, C. (2015). Trends and knowledge base in sport and exercise psychology research: a bibliometric review study. *International Review of Sport and Exercise Psychology*, 8(1), 71–94. <https://doi.org/10.1080/1750984X.2015.1019540>
- Madsen, R. E., Hansen, L. K. & Winther, O. (2004). *Singular value decomposition and principal component analysis*. Verfügbar unter http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/4000/pdf/imm4000.pdf

- Maier, D., Niekler, A., Wiedemann, G. & Stoltenberg, D. (2019, 20. November). *How document sampling and vocabulary pruning affect the results of topic models*.
<https://doi.org/10.31219/osf.io/2rh6g>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri H. & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2–3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- Maller, J. B. (1934). Forty years of psychology. A statistical analysis of American and European publications, 1894-1933. *Psychological Bulletin*, 31(8), 533–559.
<http://dx.doi.org/10.1037/h0075938>
- Mao, J., Cao, Y., Lu, K. & Li, G. (2017). Topic scientific community in science: a combined perspective of scientific collaboration and topics. *Scientometrics*, 112(2), 851–875.
<https://doi.org/10.1007/s11192-017-2418-7>
- Mayer, S. J. & Rathmann, J. M. (2018). How does research productivity relate to gender? Analyzing gender differences for multiple publication dimensions. *Scientometrics*, 117(3), 1663–1693.
<https://doi.org/10.1007/s11192-018-2933-1>
- Mimno, D., Wallach, H. M., Naradowsky, J., Smith, D. A. & McCallum, A. (2009). Polylingual topic models. In P. Koehn & R. Mihalcea (Hrsg.), *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2* (S. 880–889). Stroudsburg, PA: Association for Computational Linguistics. Verfügbar unter
http://www.aclweb.org/old_anthology/D/D09/D09-1.pdf#page=918
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M. & McCallum, A. (2011, July). Optimizing semantic coherence in topic models. In R. Barzilay & M. Johnson (Hrsg.), *Proceedings of the 2011 conference on empirical methods in natural language processing* (S. 262–272). Edinburgh, Scotland, UK: Association for Computational Linguistics.
- Mohr, J. W. & Bogdanov, P. (2013). Topic models: What they are and why they matter. *Poetics*, 41(6), 545–569. <https://doi.org/10.1016/j.poetic.2013.10.001>
- Moody, C. E. (2016). Mixing dirichlet topic models and word embeddings to make lda2vec. *arXiv preprint*. <https://arxiv.org/abs/1605.02019v1>
- Mulunda, C. K., Wagacha, P. W. & Muchemi, L. (2018, November). Review of Trends in Topic Modeling Techniques, Tools, Inference Algorithms and Applications. In *Proceedings of the 5th International Conference on Soft Computing and Machine Intelligence (ISCMI 2018)* (S. 28–37). Nairobi: Institute of Electrical and Electronics Engineering, Inc. Verfügbar unter
https://www.researchgate.net/profile/Lawrence_Muchemi/publication/332822896_Review_of_Trends_in_Topic_Modeling_Techniques_Tools_Inference_Algorithms_and_Applications/links/5cd428d692851c4eab8df1f6/Review-of-Trends-in-Topic-Modeling-Techniques-Tools-Inference-Algorithms-and-Applications.pdf

- Murzintcev, N. (2019, 12. Mai). *Select number of topics for LDA model* [Blog post]. Verfügbar unter <https://cran.r-project.org/web/packages/lдатuning/vignettes/topics.html>
- Naldi, F., Luzi, D., Valente, A. & Parenti, I. V. (2004). Scientific and technological performance by gender. In H. F. Moed, W. Glänzel & U. Schmoch (Hrsg.), *Handbook of quantitative science and technology research* (S. 299–314). Dordrecht: Kluwer. <http://dx.doi.org/10.1007/1-4020-2755-9>
- Navarro-Colorado, B. (2018). On Poetic Topic Modeling: extracting themes and motifs from a corpus of Spanish poetry. *Frontiers in Digital Humanities*, 5, Article 15. <https://doi.org/10.3389/fdigh.2018.00015>
- Nguyen, L. (2015). *Topic modeling with more confidence: A theory and some algorithms*. Paper presented at the Pacific-Asia Knowledge Discovery and Data Mining, Ho Chi Minh City. Verfügbar unter: http://dept.stat.lsa.umich.edu/~xuanlong/Talks/B_admix_pakdd.pdf
- Niekler, A. (2016). *Automatisierte Verfahren für die Themenanalyse nachrichtenorientierter Textquellen*. Dissertation, Universität Leipzig. Verfügbar unter: http://asv.informatik.uni-leipzig.de/publication/file/350/Niekler_Diss.pdf
- Niekler, A., Wiedemann, G. & Heyer, G. (2017). Leipzig Corpus Miner-a text mining infrastructure for qualitative data analysis. In *Terminology and Knowledge Engineering 2014 (TKE 2014)*, Berlin. Verfügbar unter <https://hal.archives-ouvertes.fr/file/index/docid/1005878/filename/LeipzigCorpusMiner.pdf>
- Oh, J., Stewart, A. E. & Phelps, R. E. (2017). Topics in the Journal of Counseling Psychology, 1963–2015. *Journal of Counseling Psychology*, 64(6), 604–615. <http://dx.doi.org/10.1037/cou0000218>
- Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135. <http://dx.doi.org/10.1561/1500000011>
- Reimann, P. (1998). Novizen- und Expertenwissen. In F. Klix & H. Spada (Hrsg.), *Wissen* (S. 336–367). Göttingen: Hogrefe.
- Rehurek, R. & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Verfügbar unter <https://www.fi.muni.cz/usr/sojka/papers/lrec2010-rehurek-sojka.pdf>
- Resnik, P., Garron, A. & Resnik, R. (2013, October). Using topic modeling to improve prediction of neuroticism and depression. In D. Yarowsky, T. Baldwin, A. Korhonen, K. Livescu, & S. Bethard (Hrsg.), *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (S. 1348–1353). New York, NY: Association for Computational Linguistics. Verfügbar unter <http://www.aclweb.org/anthology/D13-1133>
- Roberts, M. E., Stewart, B. M. & Tingley, D. (2016). Navigating the local modes of big data. In J. Tucker, J. Nagler, M. Metzger, D. Penfold-Brown, R. Bonneau, R. M. & Alvarez (Hrsg.), *Data Analytics in Social Science, Government, and Industry*. New York, NY: Cambridge University Press. Verfügbar unter <http://scholar.harvard.edu/files/dtingley/files/multimod.pdf>

- Roberts, M. E., Stewart, B. M. & Tingley, D. (2019). stm: R Package for Structural Topic Models. *Journal of Statistical Software*, 91(2), 1–40. <http://dx.doi.org/10.18637/jss.v091.i02>
- Roberts, M. E., Stewart, B. M., Tingley, D. & Airoldi, E. M. (2013). *The structural topic model and applied social science*. Paper presented at NIPS 2013 workshop on topic models: computation, application, and evaluation. Verfügbar unter <https://scholar.princeton.edu/files/bstewart/files/stmnips2013.pdf>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Roberts, M. E., Stewart, B. M., Tingley, D. & Benoit, K. (2018). Package “stm”. *Estimation of the Structural Topic Model* [R package description]. Verfügbar unter <https://cran.r-project.org/web/packages/stm/stm.pdf>
- Rosen-Zvi, M., Griffiths, T., Steyvers, M. & Smyth, P. (2004). The author-topic model for authors and documents. In M. Chickering & J. Halpern (Hrsg.), *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (S. 487–494). Arlington, VA: AUAI Press. Verfügbar unter: <https://mimno.infosci.cornell.edu/info6150/readings/398.pdf>
- Schaal, G. S., Dumm, S. & Fleuß, D. (2019). Die vielen Wahrheiten algorithmenbasierter Interpretation: “Die Wahrheit über Postfaktizität” dekonstruiert. *Zeitschrift für Politikwissenschaft*, 29(2), 263–285. <https://doi.org/10.1007/s41358-019-00178-3>
- Schmiedel, T., Müller, O., & vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, 22(4), 941–968. <https://doi.org/10.1177/1094428118773858>
- Schöch, C. (2016, 14. November). *What’s in my topic model? Or: Clustering topics by semantic similarity* [Blog post]. Verfügbar unter <https://dragonfly.hypotheses.org/881>
- Schofield, A. & Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics*, 4, 287–300. https://doi.org/10.1162/tacl_a_00099
- Schubert, E. (2017). *Advanced Topics in Text Mining* [Vortragsfolien]. Verfügbar unter <https://dbs.ifi.uni-heidelberg.de/files/Team/eschubert/lectures/AdvancedTopicsOfTextMiningSchubert.pdf>
- Schui, G. & Krampen, G. (2016). ZPID-Monitor 2014 zur Internationalität der Psychologie aus dem deutschsprachigen Bereich: Der ausführliche Bericht. *ZPID Science Information Online*, 16(2). Verfügbar unter: https://www.zpid.de/pub/research/zpid-monitor_2014_lang.pdf
- Senat der Leibniz-Gemeinschaft (2018). *Stellungnahme zum Leibniz-Zentrum für Psychologische Information und Dokumentation, Trier (ZPID)*. Verfügbar unter <https://www.leibniz->

[gemeinschaft.de/fileadmin/user_upload/ARCHIV_downloads/Archiv/Evaluierung/Senatsstellungnahmen/ZPID - Senatsstellungnahme 27-11-2018 mit Anlagen.pdf](https://www.gemeinschaft.de/fileadmin/user_upload/ARCHIV_downloads/Archiv/Evaluierung/Senatsstellungnahmen/ZPID_-_Senatsstellungnahme_27-11-2018_mit_Anlagen.pdf)

- Sharma, D., Kumar, B., & Chand, S. (2017). A survey on journey of topic modeling techniques from SVD to deep learning. *International Journal of Modern Education and Computer Science*, 9(7), 50. <https://doi.org/10.5815/ijmeecs.2017.07.06>
- Shi, L., Cheng, G., Xie, S. R. & Xie, G. (2019). A word embedding topic model for topic detection and summary in social networks. *Measurement and Control*, 1–10. <https://doi.org/10.1177/0020294019865750>
- Snilstveit, B., Vojtkova, M., Bhavsar, A. & Gaarder, M. (2013). *Evidence gap maps—a tool for promoting evidence-informed policy and prioritizing future research*. The World Bank. Verfügbar unter <http://documents.worldbank.org/curated/en/212651468163487838/pdf/WPS6725.pdf>
- Steinskog, A., Therkelsen, J. & Gambäck, B. (2017, Mai). Twitter topic modeling by tweet aggregation. In J. Tiedemann & N. Tahmasebi (Hrsg.), *Proceedings of the 21st Nordic Conference on Computational Linguistics* (S. 77–86). Linköping University Electronic Press. Verfügbar unter <https://www.aclweb.org/anthology/W17-0210.pdf>
- Steyvers, M. & Griffiths, T. (2007). Probabilistic topic models. In T. K. Landauer, D. S. McNamara, S. Dennis & W. Kintsch (Hrsg.), *Handbook of Latent Semantic Analysis* (S. 427–448). New York, NY: Routledge. Verfügbar unter <http://173.236.226.255/tom/papers/SteyversGriffiths.pdf>
- Steyvers, M., & Griffiths, T. L. (2008). Rational analysis as a link between human memory and information retrieval. In N. Chater & M. Oaksford (Hrsg.), *The probabilistic mind: Prospects for a Bayesian cognitive science* (S. 329–350). Oxford: Oxford University Press.
- Steyvers, M., Smyth, P., Rosen-Zvi, M. & Griffiths, T. (2004, August). Probabilistic author-topic models for information discovery. In R. Kohavi, J. Gehrke, W. DuMouchel & J. Ghosh (Hrsg.), *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (S. 306–315). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/1014052.1014087>
- Su, R. (2020). The three faces of interests: An integrative review of interest research in vocational, organizational, and educational psychology. *Journal of Vocational Behavior*, 116 (B), Article 103240. <https://doi.org/10.1016/j.jvb.2018.10.016>
- Su, R., Stoll, G. & Rounds, J. (2019). The nature of interests: Toward a unifying theory of Trait-State Interest Dynamics. In C. D. Nye & J. Rounds (Hrsg.), *Vocational interests: Rethinking their role in understanding workplace behavior and practice* (S. 11–38). SIOP Organizational Frontiers Series. New York, NY: Routledge.
- Tang, J., Meng, Z., Nguyen, X., Mei, Q. & Zhang, M. (2014). Understanding the limiting factors of topic modeling via posterior contraction analysis. In E. P. Xing (Hrsg.), *31st International Conference on Machine Learning (ICML 2014)* (S. 190–198). Stroudsburg, PA: International Machine Learning Society. Verfügbar unter <http://proceedings.mlr.press/v32/tang14.pdf>

- Tu, Y., Johri, N., Roth, D. & Hockenmaier, J. (2010, August). Citation author topic model in expert search. In C.-R. Huang & D. Jurafsky (Hrsg.), *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (S. 1265–1273). Stroudsburg, PA: Association for Computational Linguistics. Verfügbar unter <https://dl.acm.org/doi/pdf/10.5555/1944566.1944711>
- Tuleya, L. G. (Ed.). (2007). *Thesaurus of psychological index terms* (11th ed.). Washington, DC: American Psychological Association.
- Vulić, I., De Smet, W., Tang, J. & Moens, M. F. (2015). Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. *Information Processing & Management*, 51, 111–147. <https://doi.org/10.1016/j.ipm.2014.08.003>
- Wais, K. (2006). Gender Prediction Methods Based on First Names with genderizeR. *The R Journal*, 8(1), 17–37. <http://doi.org/10.32614/RJ-2016-002>
- Waldherr, A., Miltner, P., Ostner, S., Stoltenberg, D., Pfetsch, B. & Wehden, L.-O. (2019). Induktive Kategorienbildung in der Inhaltsanalyse: Kombination automatischer und manueller Verfahren. *Forum Qualitative Sozialforschung*, 20(1), 1–30. <https://doi.org/10.17169/fqs-20.1.3058>
- Wampold, B. E., & White, T. B. (1985). Research themes in counseling psychology: A cluster analysis of citations in the process and outcomes section of the Journal of Counseling Psychology. *Journal of Counseling Psychology*, 32(1), 123-126. <http://dx.doi.org/10.1037/0022-0167.32.1.123>
- Webster, G. D., Jonason, P. K. & Schember, T. O. (2009). Hot topics and popular papers in evolutionary psychology: Analyses of title words and citation counts in *Evolution and Human Behavior*, 1979–2008. *Evolutionary Psychology*, 7(3), 348–362. <https://doi.org/10.1177/147470490900700301>
- Wehrheim, L. (2019). Economic history goes digital: topic modeling the Journal of Economic History. *Cliometrica*, 13(1), 83–125. <https://doi.org/10.1007/s11698-018-0171-7>
- Weichselgartner, E., & Baier, C. (2007). Sechs Jahre PsychSpider: Aus der Praxis des Betriebs einer Psychologie-Suchmaschine für freie Web-Inhalte. *Information: Wissenschaft und Praxis*, 58(3), 173–178.
- Weiland, P., Gerards, A. & Gerards, M. (2007). Einsatz automatischer Indexierungsverfahren in der Datenbank PSYINDEX. In A. Osswald, M. Stempfhuber & C. Wolff (Hrsg.), *Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft* (413–422). Konstanz: UVK. Verfügbar unter <http://fiz1.fh-potsdam.de/volltext/isi07/15040.pdf>
- Wiedemann, G. (2013). Opening up to big data: computer-assisted analysis of textual data in social sciences. *Historical Social Research*, 38(4), 332–358. <https://doi.org/10.12759/hsr.38.2013.4.332-358>
- Wiedemann, G. (2016). *Text Mining for Qualitative Data Analysis in the Social Sciences: A Study on Democratic Discourse in Germany*. Wiesbaden: Springer VS.

- Wiedemann, G. & Niekler, A. (2018). *Topic modeling for social scientists. Model variants*. Workshop at GESIS, Cologne, Germany, 17–18th September 2018.
- Wissenschaftsrat. (2015). *Zum wissenschaftspolitischen Diskurs über Große gesellschaftliche Herausforderungen. Positionspapier*. Verfügbar unter <https://www.wissenschaftsrat.de/download/archiv/4594-15.pdf>
- Wissenschaftsrat. (2016). *Wissens- und Technologietransfer als Gegenstand institutioneller Strategien. Positionspapier*. Verfügbar unter <https://www.wissenschaftsrat.de/download/archiv/5665-16.pdf>
- Wissenschaftsrat. (2018). *Perspektiven der Psychologie in Deutschland*. Verfügbar unter <https://www.wissenschaftsrat.de/download/archiv/6825-18.pdf>
- Wittgenstein, L. (2003). *Philosophische Untersuchungen*. Auf der Grundlage der Kritisch-genetischen Edition neu herausgegeben von Joachim Schulte. Frankfurt am Main: Suhrkamp. (Originalausgabe: Blackwell, Oxford, 1953).
- Xu, S., Shi, Q., Qiao, X., Zhu, L., Jung, H., Lee, S. & Choi, S. P. (2014). Author-Topic over Time (AToT): a dynamic users' interest model. In J. H. Park, H. Adeli, N. Park & I. Woungang (Hrsg.), *Mobile, ubiquitous, and intelligent computing* (S. 239–245). Berlin: Springer. https://doi.org/10.1007/978-3-642-40675-1_37
- Yan, X., Guo, J., Lan, Y. & Cheng, X. (2013, Mai). A biterm topic model for short texts. In D. Schwalbe, V. A. Fernandes Almeida, H. Glaser, R. A. Baeza-Yates & S. B. Moon (Hrsg.), *Proceedings of the 22nd international conference on World Wide Web* (S. 1445–1456). New York, NY: ACM. <https://doi.org/10.1145/2488388.2488514>
- Zarrinkalam, F., Kahani, M. & Bagheri, E. (2019). User interest prediction over future unobserved topics on social networks. *Information Retrieval Journal*, 22(1–2), 93–128. <https://doi.org/10.1007/s10791-018-9337-y>
- ZPID – Leibniz-Zentrum für Psychologische Information und Dokumentation. (2016). *PSYINDEX Terms* (10. Aufl.). Trier: ZPID. Verfügbar unter <https://www.psyindex.de/pub/info/PSYINDEXterms2016.pdf>
- ZPID – Leibniz-Zentrum für Psychologische Information und Dokumentation. (2020). *ZPID Brennpunktthemen in PSYINDEX recherchieren* [Website]. Verfügbar unter <https://psyindex.de/index.php?wahl=PSYINDEX&uwahl=Brennpunktthemen>

Anhang

Anhang A: Originalpublikation von Studie 1

How to Identify Hot Topics in Psychology Using Topic Modeling

André Bittermann¹ and Andreas Fischer²¹Leibniz Institute for Psychology Information (ZPID), Trier, Germany²Forschungsinstitut Betriebliche Bildung (f-bb), Nuremberg, Germany

Correspondence address:

André Bittermann

Leibniz Institute for Psychology Information, (ZPID) Trier, Germany

Universitätsring 15

54296 Trier

Germany

abi@leibniz-psychology.org

HOW TO IDENTIFY HOT TOPICS IN PSYCHOLOGY

3

How to Identify Hot Topics in Psychology Using Topic Modeling

Topics of particular significance in research-active fields have been referred to as “hotspots” (Erdfelder & Bošnjak, 2016). From a scientometric point of view, the occurrence of hotspots may reflect areas of current scientific discourse. On the other hand, hotspots may also derive from current needs of society, for example, consider the impact of topics such as digitalization, terrorism, or the German “refugee crisis” (beginning in 2015) on psychological research. Thus, addressing hotspots might help to deliver research results that are interesting to both the scientific community and/or the general public if the research is imparted comprehensibly (Friedman, 2008). Nevertheless, it is an open question how to identify the set of potentially hot topics in a domain of interest. In this paper, we will contrast two ways of identifying topics based on a corpus of scientific publications (manifest classifications vs. latent topics).

A comparatively simple and straightforward approach for identifying research topics is based on existing classification systems, such as the “Classification Codes”¹ outlined in the *Thesaurus of Psychological Index Terms* (Tuleya, 2007) published by the American Psychological Association (APA). Currently, this thesaurus provides 157 categories to describe the content included in the publication database, and each category may be considered a research topic. However, with regard to identifying hotspots, the apparent simplicity of this approach is burdened with multiple drawbacks: First, the approach is based on an established classification system, and thus some of the most recent (and hot) topics may not be represented in the analysis until the classification system is expanded accordingly; second, classifications may be too broad and abstract to capture the topics that are of particular significance in research-active fields (e.g., there is no classification code specific to “evaluation,” even if some researchers may consider treatment evaluation an interesting topic); a third problem arises due to the fact that some publications address more than one

Abstract

Latent topics and trends in psychological publications were examined to identify hotspots in psychology. Topic modeling was contrasted with a classification-based scientometric approach in order to demonstrate the benefits of the former. Specifically, the psychological publication output in the German-speaking countries containing German- and English-language publications from 1980 to 2016 documented in the PSYNDEX database was analyzed. Topic modeling based on latent Dirichlet allocation was applied to a corpus of 314,573 publications. Input for topic modeling was the controlled terms of the publications, that is, a standardized vocabulary of keywords in psychology. Based on these controlled terms, 500 topics were determined and trending topics were identified. Hot topics, indicated by the highest increasing trends in this data, were facets of neuropsychology, online therapy, cross-cultural aspects, traumatization, and visual attention. In conclusion, the findings indicate that topics can reveal more detailed insights into research trends than standardized classifications. Possible applications of this method, limitations, and implications for research synthesis are discussed.

Keywords: topic modeling, hotspots, scientometrics, trends, controlled terms

HOW TO IDENTIFY HOT TOPICS IN PSYCHOLOGY

4

topic. Consider a study that examines the neuropsychological correlates of emotional lability in traumatized refugees. If only one classification is assigned (e.g., “Neuropsychology & Neurology”) the information on disorders and migration-related aspects remains hidden; instead, when using additional classifications to categorize these contents, the respective proportions remain unspecified (i.e., there may be equal or varying shares of each content).

A more complex approach to identifying topics is to derive latent topics from the manifest content addressed within a corpus of publications through methods such as topic modeling (e.g., Blei, Ng, & Jordan, 2003; Griffiths & Steyvers, 2004). The basic idea behind topic modeling is that every document can address different topics that are not known a priori. Thus, the goal is to identify these latent topics based on the documents’ manifest contents by employing algorithms that “analyze the words of the original texts to discover the themes that run through them” (Blei, 2012, p. 77). Since information on the level of the full text, abstract, or keywords can be used for topic modeling, the resulting topics have the potential to address specific subjects based on the corpus and independent from predefined classifications.

In topic modeling, each document is assumed to address each topic to varying degrees (0-100%). For example, a paper might comprise an evaluation topic with a share of 10% and other topics with a share of 90%. This means that in contrast to a dichotomous classification (this publication is assigned or is not assigned to the classification) or multiple dichotomous classifications, a probabilistic approach such as topic modeling can deal with heterogeneous topics of a publication in terms of topic proportions. In this study, such a probabilistic method is applied for topic modeling, namely latent Dirichlet allocation (LDA; Blei et al., 2003).

By applying statistical methods to the change of mean topic probabilities over time, rising and declining trends can be identified (Griffiths & Steyvers, 2004). Once trending topics are identified, scientific knowledge can be gathered from publications addressing these topics by conducting systematic reviews and meta-analyses to synthesize the results from related published research on a certain subject. The current study aims to deliver the

foundation for such research synthesis techniques in the context of hotspots in psychology: a data-driven bottom-up approach for the identification of latent topics and trends in psychological research.

Topic Modeling in Psychological Research and Scientometrics

Big data and topic modeling represent a relatively new approach of psychological research methods that can be applied to various research questions (e.g., Chen & Wojcik, 2016; Kosinski, Wang, Lakkaraju, & Leskovec, 2016). For example, Griffiths, Steyvers, and Tenenbaum (2007) used topic models for predicting word association and the effects of semantic association and ambiguity on a variety of language-processing and memory tasks. Steyvers and Griffiths (2008) showed that both human memory and information retrieval faces similar computational demands by employing topic models. Topic models have also been used for modeling couple and family text data (Atkins et al., 2012), improving the prediction of neuroticism and depression (Resnik, Garron, & Resnik, 2013), investigating mental health signals in Twitter (Coppersmith, Dredze, & Harman, 2014), analyzing the linguistic data of patient-therapist interactions (Imel, Steyvers & Atkins, 2015), and exploring differences in language use on Facebook across gender, affiliation, and assertiveness (Park et al., 2016).

In the field of scientometric analysis, which is highly relevant for the present study, Griffiths and Steyvers (2004) applied LDA topic models to a corpus of abstracts published in the *Proceedings of the National Academy of Sciences of the United States of America* and identified “hot” and “cold” topics. A topic was defined as hot if it showed an increasing linear trend in popularity and cold if it showed a decreasing linear trend in popularity. This approach was adapted in several other research fields, for example, to identify the major biological concepts from a corpus of protein-related publication titles and abstracts (Zheng, McLean, & Lu, 2006), to conduct a bibliometric analysis of aquaculture literature (Natale, Fiore, & Hofherr, 2012), to analyze the field of development studies (Thelwall & Thelwall, 2016), or

the resulting topics would be “intervention, parents, disgust, love, hate,” which includes different semantic meanings. Documents differ in topic proportions, and this is represented by the probability of a document belonging to a topic (θ). As shown in Table 3, Document 1 addresses all four topics with equal shares, whereas Document 2 mostly addresses Topic 2 and so on. The resulting mean document-topic probabilities by topic show that Topic 1 has a mean probability of 25% which corresponds to the expect proportion $1/k$ (with k being the number of topics). Topic 2, with a mean probability of 37.5%, can be considered as the most popular, whereas Topics 3 and 4 are less popular than average.

LDA is an unsupervised method, but the number of topics (k) must be defined a priori by the analyst. Griffiths and Steyvers (2004) examined different values of k and compared the resulting log-likelihoods. Yet another approach would be to test various values of k and determine the optimal k intellectually, that is, by expert judgment to decide whether the topics are in balance between too broad or too specific (Thelwall & Thelwall, 2016). In this study, we follow the first approach.

Using Controlled Terms for Topic Modeling

For a reliable identification of the representative topics within a research field, the information about the documents’ content must be of high quality. For instance, if abstracts give a mere introduction rather than an objective summary, the resulting topics will reflect the theoretical background or the studies’ *raison d’être* rather than their actual content. The same applies to keywords that, for instance, contain the statistical methods that were used and do not represent the actual topic of the study (e.g., in the case of “analysis of variance,” from a keyword point of view it remains unclear whether the method was simply applied, discussed, or further developed). To avoid latent semantic heterogeneity within a topic, all keywords should be chosen according to the same rules (e.g., keywords for statistical procedures are assigned only if they themselves are the focus of the study and not their mere application is referred to). In most studies, authors provide keywords that further summarize the document’s

to explore hydropower research (Jiang, Qiang & Lin, 2016). The current study is the first to apply LDA-based topic modeling for a scientometric analysis of psychological research in the German-speaking countries.

A Brief Illustration of LDA-based Topic Modeling

In the following, a very brief and illustrative description of LDA and topic modeling is provided. Further details and more technical descriptions can be found in Blei (2012) and Blei et al. (2003). The underlying assumption of LDA is that a document represents a mixture of topics with different proportions (Blei et al., 2003). Using Bayesian probabilistic modeling, LDA aims to identify clusters of terms (i.e., topics) that tend to co-occur within documents (Park et al., 2016). Thus, topics are defined as a distribution over a fixed vocabulary (Blei, 2012). In a generative process, two kinds of probabilities are drawn from Dirichlet distributions over (1) the prior weight of a certain word in a topic (β) for the probabilities of terms occurring in a certain topic (ϕ), and (2) the prior weight of a certain topic in a document (α) for the probabilities of topics occurring in a certain document (θ) based on the terms within the document. “Prior” means that the α and β hyperparameters have to be set prior to the analysis. Lower values of α result in documents belonging to fewer topics, and lower values of β result in more separated topics.

For a simplified illustration of the main idea behind topic modeling in a scientometric context, imagine a corpus consisting of four documents and a model of four topics. For the sake of brevity, each document shall consist of 16 terms (see Table 1). In this idealized example, LDA reveals four topics by clustering co-occurring terms, of which the five most frequent terms are shown in Table 2. Note that each topic actually consists of all unique terms of the corpus, that is, of all four documents. The terms are sorted by frequency to best represent different topics. For the sake of illustration, the results presented in Table 2 can be considered ideal because these topics reflect optimal semantic differences. As a real LDA analysis for this very small sample corpus is based exclusively on term co-occurrences, one of

content. If these keywords are uncontrolled (i.e., can be freely chosen), (1) it is not guaranteed that they actually represent the main concepts, ideas, and topics of the publication; (2) they sometimes are long phrases and not terms; and (3) keywords from different authors might be different terms for the same idea (e.g., adaptation vs. adaption vs. adjustment) or, conversely, the same words for different ideas. These are problematic aspects for topic modeling using LDA, since topics are identified according to word co-occurrences (Blei et al., 2003). Topic models aim to capture semantically related topics (Wallach, Mimno, & McCallum, 2009), but they do not generate topics based on the words’ inherent semantic relations.

The PSYNDEX database is developed and hosted by the Leibniz Institute for Psychology Information (ZPID; Trier, Germany) and is a comprehensive database containing German- and English-language publications in psychology and closely related disciplines from the German-speaking countries. In early July 2017, there were more than 327,400 documents indexed in PSYNDEX (accessible at www.PubPsych.eu). The PSYNDEX editorial staff assigns controlled terms (CTs) from the aforementioned *Thesaurus of Psychological Index Terms* published by the APA (Tuleya, 2007; ZPID, 2016). In the context of topic modeling, this controlled vocabulary has several advantages: (1) The CTs correspond with the content of the publications. (2) The terms’ standardized spelling avoids synonyms or variations in expressions. (3) The corpus for topic modeling consists of only those words that are relevant to the content. Stop words that contain little topical content (e.g., “the”, “a”, “and”) have to be neither defined nor deleted. (4) All CTs are available in German and English; therefore, the whole corpus of publications can be used irrespective of the documents’ language. (5) In contrast to abstract texts, the terms do not have to be stemmed with the resulting problem of word fragments. (6) Since the corpus contains fewer words, computation time decreases and fewer memory resources are needed. In a pretest with 3,846 documents, LDA based on CTs took less than 7% of the time needed for an abstract-based LDA while revealing comparable results. Thus, in contrast to prior research using abstracts as

primary data for topic modeling analysis (e.g., Griffiths & Steyvers, 2004; Jiang et al., 2016), the current study employs CTs for topic modeling.

Objectives

The objectives of the current study are twofold:

- (1) to examine trends of latent topics, and
- (2) to contrast latent topics with manifest classifications.

LDA-based topic modeling will be applied to a corpus of psychological publications from the German-speaking countries retrieved from PSYNDEX. Increasing and decreasing linear trends as well as nonlinear trends will be identified. Furthermore, the topics will be contrasted with classifications in terms of thematic specificity.

Method

Data

Data were extracted from the PSYNDEX database on July 3, 2017. A total of 316,996 of the indexed psychological articles, book chapters, reports, and dissertations were published between 1980 and 2016. Biographies or historical sources (reprints or selected readings) were excluded, since they usually address the topic retrospectively, resulting in $N = 314,573$ publications.

Software

Analyses were conducted in RStudio version 1.0.153 (RStudio Team, 2016) based on R version 3.4.2 (R Core Team, 2017). For text mining and topic modeling, the packages *tm* 0.7-1 (Feinerer, Hornik, & Meyer, 2008) and *topicmodels* 0.2-6 (Grün & Hornik, 2011) were used. Additional operations were conducted with packages *dplyr* 0.5.0, *readr* 1.1.0, *splitstackshape* 1.4.2, *Xmisc* 0.2.1, *lattice* 0.20-35, and *met* 7.3-12.

Topic Modeling

LDA was applied using Gibbs sampling with parameters as suggested by Awati (2015), that is, 4,000 omitted Gibbs iterations at beginning, 2,000 Gibbs iterations, 500

fluctuations—and considering the duration of a typical publication cycle—an estimation over a shorter time span implies severe overfitting (Fischer, 2015) and may not represent a topic's significance well. The complete R code used in the analyses is provided in the Electronic Supplementary Material (ESM 1).

Results

Model Selection

The corpus of $N = 314,573$ documents contained 6,073 unique terms. By comparing log-likelihoods of the resulting models (as shown in Table 4), $k = 500$ was determined as the optimal number of topics. A table containing the top 15 terms of all topics can be found in the Electronic Supplementary Material (ESM 2).

Trends in Topics

Linear trends in changes of mean document-topic probabilities (θ) over time were analyzed according to Griffiths and Steyvers (2004) with an additional examination of nonlinearity. Significantly increasing linear trends could be found for 128 of the topics, and significantly decreasing linear trends could be found for 135 of the topics, both at the $p = .0001$ level. The 10 topics with highest increasing linear trends (i.e., hot topics) are listed in Table 5. Figure 1 shows their mean document-topic probabilities (θ) by publication year. The major hot topics are neuropsychology and genetics, online therapy, human migration, traumatization, and visual attention. A closer look at the terms of these topics (Table 5) reveals that these major themes can be further specified. Traumatization, for example, can be further specified with three narrower topics: traumatization of refugees during war and torture (Topic 86), therapy of emotional trauma (Topic 344), and trauma-related disorders and processes (Topic 95).

Since the focus of the current study was on hot topics, additional trend analyses are reported briefly (see ESM 2 for topic terms and more information on the following topics). Strongly decreasing linear trends (i.e., cold topics) could be found in topics referring to

omitted in-between Gibbs iterations, and five repeated random starts. Parameters of the symmetric Dirichlet priors were set according to Tang, Meng, Nguyen, Mei, and Zhang (2014), that is, $\alpha = 0.1$ (resulting in documents belonging to fewer topics) and $\beta = 0.01$ (resulting in well-separated topics). Concerning the number of topics k , we inspected the log-likelihood estimates for various values of k , which is referred to as the commonly used approach (Kosinski et al., 2016). We ran models with 100, 150, 200, 300, 400, and 500 topics comparable to Griffiths & Steyvers (2004), who tested values of 50, 100, 200, 300, 400, 500, 600, and 1,000 topics. Values of k higher than 500 were discarded, since more topics decrease understanding and verifiability by experts (De Battisti, Ferrara, & Salini, 2015). Text input for the topic models were the publications' controlled keyword terms (CTs). They were prepared for LDA by removing spaces, parentheses, hyphens, slashes, and apostrophes.

Modeling Trends

Previous research employed linear regression models for identifying increasing and decreasing trends (Griffiths & Steyvers, 2004; Paul & Girju, 2009; Ponweiser, Grün, & Hornik, 2014). Hot topics were defined by the highest linear slopes. We extended this approach by taking nonlinearity into account to identify nonlinear trends. Specifically, we applied multilayer perceptrons (MLPs) with two hidden-units to model the average topic probability (mean of document-topic probabilities over all documents for each topic) as a nonlinear function of the year of publication. The MLPs applied provide nonlinear regression functions with a minimal sum of squared residuals for each topic, and thus provide an estimate of R^2 , given an optimal nonlinear transformation of the year of publication (Fischer, 2015). Two hidden units were included to allow for nonmonotonic functions while at the same time minimizing the risk (and amount) of overfitting (Fischer, 2015). The difference between R^2_{MLP} and R^2_{linear} is applied as an indicator of the amount of nonlinearity that is not accounted for by the linear model. More specifically, nonlinearity is defined by $R^2_{MLP} > 2 \cdot R^2_{linear}$. Trends were estimated over a period of more than two years. Because of random

human-factors engineering (Topic 310), psychosomatic disorders (Topic 361), incarceration (Topic 472), social and political processes in West and East Germany (Topics 41, 393, and 186), experimental methodology (Topic 342), group psychotherapy (Topic 491), community mental health services (Topic 163), and infectious disorders (Topic 479).

The comparison of R^2_{linear} and R^2_{MLP} revealed topics with a considerable amount of nonlinearity that is not accounted for by the linear model. The largest difference between R^2_{linear} and R^2_{MLP} (i.e., nonlinear trends) could be found for topics referring to psychodiagnosis and testing (Topic 467, with peaks in 2006 and 2011), outpatient psychotherapy (Topic 334, with peak in 1998), family relations (Topic 259), prevention and health promotion (Topic 162, highest peak in 1991), Internet and information systems (Topic 481, with peak in 1999), organizational psychology (Topic 345), sexual relations with clients in psychotherapy (Topic 237, peaks in 1995 and 1998), racial and ethnic attitudes (Topic 130, peak in 1993), health behavior and dental health (Topic 44), and relations between socioeconomic background of the family and education (Topic 138).

Relationship Between Topics and Classifications

According to the second objective of the current study, we investigated whether topics can be allocated to a specific PSYNDEX subject classification. If this is the case, topics either match the classifications' content or they provide more detailed information within the classification. If this is not the case, topics cover themes that could only be matched by multiple classifications. For every document, the assigned classifications were compared to the documents' most probable topics in order to examine content similarities and differences. Similar to Griffiths and Steyvers' (2004) approach for identifying diagnostic topics, Figure 2 shows a level plot of mean document-topic probabilities (θ) by topics and main classifications for the hot topics. For creating the level plot, publications were grouped by classification (in case of multiple classifications, the document was assigned to each classification). Then, mean θ probabilities were determined by each classification. This allowed the investigation of

the extent to which a topic's semantic content (as reflected by its top terms) corresponds with the classification system. For the sake of clarity, only main classification categories are included in Figure 2 (as the complete APA classification system consists of 157 codes).

The darker the cells of the level plot, the higher the mean θ . If a topic column shows different colors, the θ values are not equally distributed over the classifications, that is, the topics' semantic content cannot be reflected by a single classification. Clearly, the topics do not match the classifications perfectly, but they do show correspondence with various classifications (in this case, only one dark cell is observed for each topic). For example, the highest mean θ for Topic 371 (referring to human migration and cross-cultural aspects) can be observed in "2900 Social Processes & Social Issues." Since it also shows a relatively high mean θ in various other classifications, this topic cannot be described by a single classification. The hot topics concerning neuropsychology (Topics 364, 249, and 323) and genetics (Topic 459) show their highest mean θ in "2500 Physiological Psychology & Neuroscience," but also in other classifications. No distinctively matching classification can be identified for Topics 86 (traumatization of refugees) and 95 (traumatization-related disorders).

Selecting Publications for Research Synthesis

The publications related to a topic can be filtered by (1) using the document-topic probabilities (θ) or (2) using the keywords that constitute the topic for literature search. We employed the first approach on the example of Hot Topic 386 (online therapy) and sorted documents by θ in decreasing order. This resulted in a list of all publications in the corpus, with the ones most likely addressing the topic ranking highest. The results were then filtered by selecting only empirical studies with values of θ higher than $1/k$ (i.e., the average document-topic probability). This means that Topic 386 occurs in these empirical studies with a probability above average. The distribution of θ values is shown in Figure 3. Inclusion criteria for subsequent research synthesis approaches can be applied to this subset of 1,083

there is no need for stop words, they offer excellent readability, and since the topics consists of CTs, they can be used directly for subsequent literature searches.

A significant disadvantage of using CTs is the time of their first thesaurus inclusion as a potential artefact. For instance, recently added CTs such as "Political Asylum" or "Asylum Seeking" (both included in 2015) cannot describe a topic during the years before their addition. Nevertheless, if one is interested in recent topics, the following approach for defining hotspots besides considering trends over time could be employed: By building a corpus for the respective recent years (e.g., 2015-2016), popular topics could be defined by examining the highest mean document-topic probabilities. This represents a cross-sectional approach using all currently available CTs.

Similar to classifications, thesaurus-based CTs have limitations regarding their semantic detail. The uncontrolled keywords of the current study are "topic modeling, hotspots, scientometrics, trends, controlled terms," with more or less corresponding CTs "Mathematical Modeling, Scientific Communication, Trends" (no matches for the quite specific keywords "hotspots" and "controlled terms"). The use of words in the abstract would overcome these shortcomings, since every word of the original text can be included. Downsides, on the other hand, are the problem of defining stop words (e.g., Schofield, Magnusson, & Mimno, 2017) and a much larger corpus vocabulary with higher computational demands that would require several days of calculation time or the use of a computer cluster.

The number of topics k was determined by computing models for values of k and inspecting the respective log-likelihoods, which is referred to as the commonly used approach (Kosinski et al., 2016). The log-likelihoods increased with higher values of k , indicating that a model with more topics could show an even better fit. However, a model with more topics is more difficult to be understood and verified by experts (De Battisti et al., 2015). Besides, inspecting the hot topics for the applied values of k in this study revealed stable themes of neuropsychology, online therapy, human migration, traumatization, and visual attention.

documents. Since the documents are ranked by θ , a list can be generated that allows for an inspection of relevant documents in the order of their topic probabilities.

Discussion

The current study applied LDA-based topic modeling for a scientometric analysis of psychological research as a data-driven bottom-up approach for the identification of latent topics and trends. In a model with 500 topics, strongly increasing linear trends were found for topics addressing neuropsychology and genetics, online therapy, human migration and cross-cultural aspects, traumatization, and visual attention. These topics were referred to as hot topics in psychology. Additionally, it was shown how the resulting topics can be used for purposes of research synthesis.

The topics' contents corresponded with respective classifications, but as expected, they could not be matched to a single subject classification. Thus, the topics provided information beyond the scope of a predefined classification system. Prior scientometric research in psychology used classifications for determining trends (e.g., Krampen & Trierweiler, 2013; Krampen, 2016). From our results, it can be concluded that this approach is feasible as long as the classifications' specificity is satisfactory. Using topic modeling, we were able to find specific topics that would have not been easy to detect by a classification-based approach, for example, lifestyle of adolescents and popular culture (Topic 49), attitude change of public opinion (Topic 72), values in individualism versus collectivism (Topic 144), or traumatization of refugees because of war and torture (Topic 86). Most topics represent a mixture of classifications.

Methodological Limitations

Topic models were based on standardized keywords (controlled terms, CTs) of the publications. This approach resulted in a much smaller number of corpus terms than would have resulted from using abstracts. CTs reflect a document's content in a condensed manner and offer several advantages in the context of topic modeling: Computation times are shorter,

In this study, basic LDA was employed using the R programming language. Newer developments such as dynamic topic modeling with a focus on changes over time (Blei & Lafferty, 2006) or correlated topic models that aim to capture correlations between the occurrence of latent topics (Blei & Lafferty, 2007) could further improve the identification of hot topics in psychology. The analysis of abstracts from different languages by employing polylingual topic models (Mimno et al., 2009) or multilingual probabilistic topic modeling (Vulić, De Smet, Tang, & Moens, 2015) could be of interest for future research as well.

Implications for Research Synthesis

The topic modeling approach presented in this paper can be applied to the identification of hotspots in psychology. Erdfelder and Bošnjak (2016) related hotspots to the presence of a significant number of primary studies within a research-active field. We expanded the scope of hotspots by including all types of publications with the exception of historical studies and biographies in order to gain a comprehensive view on the topics that are addressed. For subsequent research synthesis purposes, primary studies which address hot topics can be easily identified in PSYNDEX by filtering the documents. This results in a list with the documents that show the highest document-topic probabilities at the highest ranks. A more common approach for selecting documents would be using the keywords (CTs) that constitute the topic.

In this paper, only the top 15 terms of each hot topic were reported. Since a topic consists of a long list of terms, with various frequencies and term-to-topic probabilities, we encourage readers to take a closer look at the topics of interest. A sophisticated method for visualizing and interpreting topics is provided by "LDAvis" (Sievert & Shirley, 2014), which defines the relevance for ranking terms within topics based on weight parameters and can be employed in R with the "LDAvis" package (Sievert & Shirley, 2015).

Other Possible Applications

A researcher who wants to develop a new area of interest can learn more about the subject's structure by looking at the underlying topics. A publication database could be explored more in depth using topics, as illustrated by topic-based web browsers of Wikipedia (Chaney & Blei, 2012) or the Signs journal (Goldstone, Galán, Lovin, Mazzaschi, & Whitmore, 2014). Moreover, for a better navigation through a model with many topics, scientific documents could be divided into several topic clusters (Yau, Porter, Newman, & Suominen, 2014). Such clusters could constitute empirically derived themes as an alternative to manifest classifications.

Since a document-topic probability is computed for every publication, those papers could be recommended that show the highest probabilities for the respective topic. A more sophisticated approach was presented by Wang and Blei (2011), who developed an algorithm for recommending scientific articles to users of an online community.

Authors usually indicate their fields of interest, for example, "psychotherapy research" or "cognitive processes." Here, topic modeling can be used to identify research topics based on the authors' publications (Lu & Wolfram, 2012; Rosen-Zvi, Griffiths, Steyvers, & Smyth, 2004). This procedure results in a publication-based profile of authors which can be applied to find experts for specific topics, find authors with similar topics, or analyze authors' change of publication-based interests over time.

Conclusion

Topic modeling is a feasible method for an exploratory analysis of topics in psychological publications and for identifying hot research topics. The identification of specific topics in a large corpus of publications offers new possibilities of exploring research beyond predefined classifications. Furthermore, topics can be the starting point for subsequently applied research synthesis methods.

References

- Atkins, D. C., Rubin, T. N., Steyvers, M., Doeden, M. A., Baucom, B. R., & Christensen, A. (2012). Topic models: A novel method for modeling couple and family text data. *Journal of Family Psychology, 26*(5), 816-827. doi:10.1037/a0029607
- Awati, K. (2015, September 29). A gentle introduction to topic modeling using R [Blog post]. Retrieved from <https://eight2late.wordpress.com/2015/09/29/a-gentle-introduction-to-topic-modeling-using-r/>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84. doi:10.1145/2133806.2133826
- Blei, D. M., & Lafferty, J. D. (2006, June). Dynamic topic models. In W. Cohen & A. Moore (Eds.), *Proceedings of the 23rd International Conference on Machine Learning* (pp. 113-120). New York, NY: ACM. doi:10.1145/1143844.1143859
- Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics, 1*, 17-35. doi:10.1214/07-AOAS114
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993-1022. doi:10.1162/jmlr.2003.3.4-5.993
- Chaney, A. J. B., & Blei, D. M. (2012, March). Visualizing topic models. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (IWSM)*. Retrieved from <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewFile/4645/5021>
- Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology. *Psychological Methods, 21*(4), 458-474. doi:10.1037/met0000111
- Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. In P. Resnik, R. Resnik, & M. Mitchell (Eds.), *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical*

Acknowledgments

We thank Lisa Trierweiler and Katja Singleton for helpful comments and recommendations during the writing process, Jürgen Wiesenhütter and Veronika Kuhberg-Lasson for valuable input during early phases of this research, and Andreas Konz and Jannik Lorenz for hardware support.

Reality (pp. 51-60). Stroudsburg, PA: Association for Computational Linguistics. Retrieved from <http://www.aclweb.org/anthology/W14-3207>

De Battisti, F., Ferrara, A., & Salini, S. (2015). A decade of research in statistics: A topic model approach. *Scientometrics, 103*(2), 413-433. doi:10.1007/s11192-015-1554-1

Erdfelder, E., & Bošnjak, M. (2016). „Hotspots in Psychology“: A new format for special issues of the *Zeitschrift für Psychologie. Zeitschrift für Psychologie, 224*(3), 141-144. doi:10.1027/2151-2604/a000249

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software, 25*(5), 1-54. doi:10.18637/jss.v025.i05

Fischer, A. (2015). How to determine the unique contributions of input-variables to the nonlinear regression function of a multilayer perceptron. *Ecological Modelling, 309*, 60-63. doi:10.1016/j.ecolmodel.2015.04.015

Friedman, D. P. (2008). Public outreach: A scientific imperative. *Journal of Neuroscience, 28*(46), 11743-11745. doi:10.1523/JNEUROSCI.0005-08.2008

Goldstone, A., Galán, C., Lovin, C. L., Mazzaschi, A., & Whitmore, L. (2014). *An Interactive Topic Model of Signs*, edited by Andrew Goldstone. *Signs at 40*. Retrieved from <http://signsat40.signsjournal.org/topic-model>

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*(suppl 1), 5228-5235. doi:10.1073/pnas.0307752101

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*(2), 211-244. doi:10.1037/0033-295X.114.2.211

Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software, 40*(13), 1-30. doi:10.18637/jss.v040.i13

Imel, Z. E., Steyvers, M., & Atkins, D. C. (2015). Computational psychotherapy research: Scaling up the evaluation of patient-provider interactions. *Psychotherapy, 52*(1), 19-30. doi:10.1037/a0036841

- Jiang, H., Qiang, M., & Lin, P. (2016). A topic modeling based bibliometric exploration of hydropower research. *Renewable and Sustainable Energy Reviews*, 57, 226-237. doi:10.1016/j.rser.2015.12.194
- Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, 21(4), 493-506. doi:10.1037/met0000105
- Krampen, G. & Trierweiler, L. (2013). Research on emotions in developmental psychology contexts: Hot topics, trends, and neglected research domains. In C. Mohiyeddini, M. Eysenck & S. Bauer (Eds.), *Handbook of psychology of emotions. Recent theoretical perspectives and novel empirical findings. Volume 1* (pp. 63-79). New York: Nova Science Publishers.
- Krampen, G. (2016). Scientometric trend analyses of publications on the history of psychology: Is psychology becoming an unhistorical science? *Scientometrics*, 106(3), 1217-1238. doi:10.1007/s11192-016-1834-4
- Lu, K., & Wolfram, D. (2012). Measuring author research relatedness: A comparison of word-based, topic-based, and author cocitation approaches. *Journal of the Association for Information Science and Technology*, 63(10), 1973-1986. doi:10.1002/asi.22628
- Mimno, D., Wallach, H. M., Naradowsky, J., Smith, D. A., & McCallum, A. (2009, August). Polylingual topic models. In P. Koehn & R. Mihalcea (Eds.), *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2* (pp. 880-889). Stroudsburg, PA: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/old_anthology/D/D09/D09-1.pdf#page=918
- Natale, F., Fiore, G., & Hofherr, J. (2012). Mapping the research on aquaculture. A bibliometric analysis of aquaculture literature. *Scientometrics*, 90(3), 983-999. doi:10.1007/s11192-011-0562-z

- Schofield, A., Magnusson, M., & Mimno, D. (2017). Understanding text pre-processing for latent Dirichlet allocation. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (pp. 432-436). New York, NY: Association for Computational Linguistics. Retrieved from <http://www.cs.cornell.edu/~xanda/winlp2017.pdf>
- Sievert, C., & Shirley, K. E. (2015). LDAvis: Interactive visualization of topic models. R package version 0.3.2 [Computer Software]. Retrieved from <https://CRAN.R-project.org/package=LDAvis>
- Sievert, C., & Shirley, K. E. (2014, June). LDAvis: A method for visualizing and interpreting topics. In J. Chuang, S. Green, M. Hearst, J. Heer, & P. Koehn (Eds.), *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63-70). Stroudsburg, PA: Association for Computational Linguistics. Retrieved from <http://www.aclweb.org/anthology/W14-3110>
- Steyvers, M., & Griffiths, T. L. (2008). Rational analysis as a link between human memory and information retrieval. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for a Bayesian cognitive science* (pp. 329-350). Oxford: Oxford University Press.
- Tang, J., Meng, Z., Nguyen, X., Mei, Q., & Zhang, M. (2014, January). Understanding the limiting factors of topic modeling via posterior contraction analysis. In E. P. Xing (Ed.), *31st International Conference on Machine Learning (ICML 2014)* (pp. 190-198). Stroudsburg, PA: International Machine Learning Society. Retrieved from <http://proceedings.mlr.press/v32/tang14.pdf>
- Thelwall, M., & Thelwall, S. (2016). Development studies research 1975-2014 in academic journal articles: The end of economics? *El Profesional de la Información*, 25(1), 47-58. doi:10.3145/epi.2016.ene.06

- Park, G., Yaden, D. B., Schwartz, H. A., Kern, M. L., Eichstaedt, J. C., Kosinski, M., ... Seligman, M. E. (2016). Women are warmer but no less assertive than men: Gender and language on Facebook. *PloS One*, 11(5), e0155885. doi:10.1371/journal.pone.0155885.003
- Paul, M. J., & Girju, R. (2009, September). Topic modeling of research fields: An interdisciplinary perspective. In R. Mitkov & G. Angelova (Eds.), *Proceedings of the International Conference RANLP-2009* (pp. 337-342). Stroudsburg, PA: Association for Computational Linguistics. Retrieved from <http://www.anthology.aclweb.org/R/R09/R09-1.pdf#page=361>
- Ponweiser, M., Grün, B., & Hornik, K. (2014). Finding scientific topics revisited. In M. Carpita, E. Bentari, & E. Qannari (Eds.), *Advances in latent variables* (pp. 93-100). Cham, Switzerland: Springer International Publishing. doi:10.1007/10104_2014_11
- R Core Team (2017). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Resnik, P., Garron, A., & Resnik, R. (2013, October). Using topic modeling to improve prediction of neuroticism and depression. In D. Yarowsky, T. Baldwin, A. Korhonen, K. Livescu, & S. Bethard (Eds.), *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1348-1353). New York, NY: Association for Computational Linguistics. Retrieved from <http://www.aclweb.org/anthology/D13-1133>
- Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004, July). The author-topic model for authors and documents. In M. Chickering & J. Halpern (Eds.), *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (pp. 487-494). Arlington, VA: AUAI Press. Retrieved from <https://mimno.infosci.cornell.edu/info6150/readings/398.pdf>
- RStudio Team (2016). RStudio: Integrated development for R [Computer software]. RStudio, Inc., Boston, MA. Retrieved from <http://www.rstudio.com/>

- Tuleya, L. G. (Ed.). (2007). *Thesaurus of psychological index terms* (11th ed.). Washington, DC: American Psychological Association.
- Vulić, I., De Smet, W., Tang, J., & Moens, M. F. (2015). Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. *Information Processing & Management*, 51(1), 111-147. doi:10.1016/j.ipm.2014.08.003
- Wallach, H. M., Mimno, D. M., & McCallum, A. (2009). Rethinking LDA: Why priors matter. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22 (NIPS 2009)* (pp. 1973-1981). LA Jolla, CA: Neural Information Processing Systems. Retrieved from <http://dirichlet.net/pdf/wallach09rethinking.pdf>
- Wang, C., & Blei, D. M. (2011, August). Collaborative topic modeling for recommending scientific articles. In C. Apte, J. Ghosh, & P. Smyth (Eds.), *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 448-456). New York, NY: ACM. doi:10.1145/2020408.2020480
- Yau, C. K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics*, 100(3), 767-786. doi:10.1007/s11192-014-1321-8
- Zheng, B., McLean, D. C., & Lu, X. (2006). Identifying biological concepts from a protein-related corpus with a probabilistic topic model. *BMC Bioinformatics*, 7(1), 58. doi:10.1186/1471-2105-7-58
- ZPID - Leibniz-Zentrum für Psychologische Information und Dokumentation (Eds.). (2016). *PSYNDEX terms* (10th ed.). Trier, Germany: ZPID. Retrieved from <https://www.zpid.de/pub/info/PSYNDEXterms2016.pdf>

Footnotes

¹A list of all codes can be retrieved from

<http://www.apa.org/pubs/databases/training/class-codes.aspx>. Each publication can be linked to one or more main classifications (e.g., “Psychometrics & Statistics & Methodology,” “Human Experimental Psychology,” or “Personality Psychology”), and/or respective subcategories (e.g., a publication classified as “Psychometrics & Statistics & Methodology” may be classified more specifically as “Sensory & Motor Testing” or “Clinical Psychology Testing”).

Table 1

Example of Four Documents Consisting of 16 Terms Each

Document			
1	2	3	4
love	happiness	disgust	amazement
hate	joy	anger	surprise
fear	serenity	rage	joy
disgust	love	hate	happiness
intervention	therapy	psychoanalysis	psychotherapy
therapist	therapist	transference	counseling
client	client	client	disorder
disorder	treatment	disorder	treatment
mother	intervention	treatment	outcome
brother	disorder	intervention	exposition
sister	parents	mother	client
father	siblings	father	therapist
school	learning	parents	parents
learning	teacher	child	mother
grades	class	grades	college
class	college	achievement	university

Table 2

Five Most Common Terms of the Resulting Topics (Idealization)

Topic			
1	2	3	4
„emotions“	„therapy“	„family“	„education“
love	client	parents	class
joy	disorder	mother	college
happiness	therapist	father	grades
disgust	intervention	child	learning
amazement	treatment	brother	teacher

Note. The topic titles are descriptive terms provided by the authors and were not generated by the model.

Table 3

Illustration of Document-Topic Probabilities (θ)

Document	Topic				Sum
	1	2	3	4	
1	0.250	0.250	0.250	0.250	1
2	0.250	0.375	0.125	0.250	1
3	0.250	0.375	0.250	0.125	1
4	0.250	0.500	0.125	0.125	1
Mean	0.250	0.375	0.188	0.188	1

Note. Document 1 addresses all four topics with equal shares ($1/k$, with k = number of topics), whereas Documents 2 to 4 show different topic probabilities. By mean probabilities, Topic 2 is addressed with more than average probability and, thus, can be interpreted as the most popular.

Table 4

Log-Likelihoods (LL) of Topic Models by Different Numbers of Topics (k)

k	100	150	200	300	400	500
LL	-8234278	-7491948	-7032583	-6403997	-5978100	-5695993

Table 5

Top 15 Terms of the Ten Hottest Topics

Topic	Top 15 Terms
364	Functional Magnetic Resonance Imaging, Cerebral Blood Flow, Prefrontal Cortex, Amygdala, Neuroanatomy, Biological Neural Networks, Cingulate Cortex, Brain, Oxygenation, Insula, Rewards, Striatum, Hippocampus, Brain Connectivity, Cognitive Control
249	Functional Magnetic Resonance Imaging, Cerebral Blood Flow, Brain, Parietal Lobe, Prefrontal Cortex, Neuroanatomy, Frontal Lobe, Temporal Lobe, Oxygenation, Neuroimaging, Magnetic Resonance Imaging, Biological Neural Networks, Occipital Lobe, Visual Cortex, Spectroscopy
386	Internet, Computer Mediated Communication, Online Therapy, Online Social Networks, Internet Usage, Electronic Communication, Communications Media, Websites, Social Media, Virtual Reality, Computer Assisted Therapy, Cellular Phones, Privacy, Telemedicine, Information Technology
459	Genes, Polymorphism, Genetics, Serotonin, Genotypes, Dopamine, Alleles, Biological Markers, Phenotypes, Attention Deficit Disorder With Hyperactivity, Susceptibility (Disorders), Neurotransmission, Brain Derived Neurotrophic Factor, Neural Receptors, Tryptophan
371	Cross-Cultural Differences, Human Migration, Cross-Cultural Communication, Cultural Sensitivity, Cross-Cultural Treatment, Multiculturalism, Expatriates, Transcultural Psychiatry, International Organizations, Cross-Cultural Counseling, Globalization, Multicultural Education, Foreign Workers, Acculturation, Racial And Ethnic Differences
323	Magnetic Resonance Imaging, Brain, Neuroimaging, Neuroanatomy, Hippocampus, Gray Matter, Brain Size, Tomography, Prefrontal Cortex, White Matter, Amygdala, Cingulate Cortex, Cerebral Cortex, Temporal Lobe, Morphology
86	Posttraumatic Stress Disorder, Emotional Trauma, Refugees, Trauma, War, Victimization, Torture, Persecution, Survivors, Violence, Injuries, Asylum Seeking, Exposure Therapy, Human Migration, Transgenerational Patterns
344	Posttraumatic Stress Disorder, Emotional Trauma, Trauma, Eye Movement Desensitization Therapy, Stress Reactions, Intrusive Thoughts, Adjustment Disorders, Acute Stress Disorder, Traumatic Neurosis, Posttraumatic Growth, Complex PTSD, Exposure Therapy, Accidents, Medical Personnel, Metaphor
365	Attention, Visual Attention, Selective Attention, Visual Search, Distraction, Cues, Reaction Time, Stimulus Parameters, Eye Movements, Attentional Capture, Visual Perception, Stimulus Salience, Visual Stimulation, Attentional Bias, Divided Attention
95	Emotional Trauma, Posttraumatic Stress Disorder, Trauma, Dissociation, Dissociative Disorders, Early Experience, Dissociative Identity Disorder, Depersonalization, Borderline Personality Disorder, Neurobiology, Introjection, Dissociative Patterns, Amnesia, Psychodynamic Psychotherapy, Depersonalization/Derealization Disorder

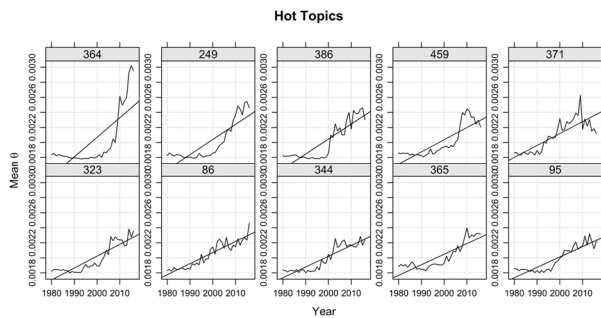


Figure 1. Mean values of document-topic probabilities θ by publication year for the 10 hottest topics with added linear regression line. The topics are described in Table 5.

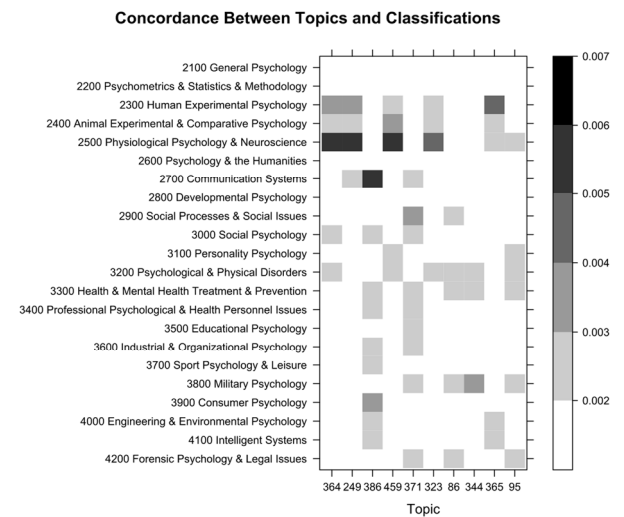


Figure 2. Level plot showing mean document-topic probabilities (θ) by topics and main classifications. Only the 10 hottest topics are displayed. Darker cells represent higher values of θ . For example, in the publications that were classified as “2500 Physiological Psychology & Neuroscience,” the highest mean θ resulted for Topics 364, 249, 459, and 323. The respective topics are described in Table 5.

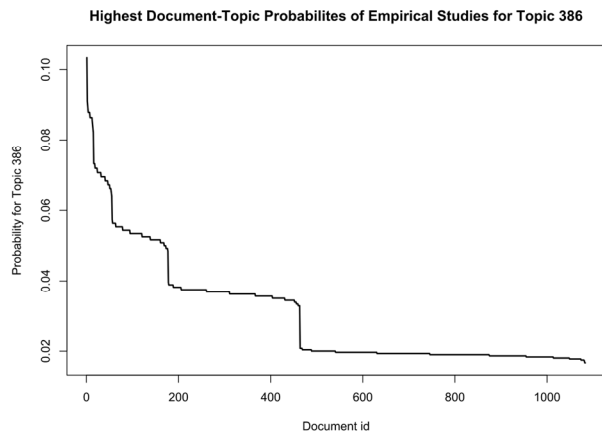


Figure 3. Document-topic probabilities (θ) of $n = 1,083$ empirical studies for Topic 386 (sorted by θ). Only documents with θ higher than average are shown.

Electronic Supplementary Material (ESM 1): R Code of the Analyses (ESM1_Code.R)

Electronic Supplementary Material (ESM 2): List of Topics for $k = 500$ (ESM2_Topics.csv)

Anhang B: Originalpublikation von Studie 2

Ist die psychologische Forschung durchlässig für aktuelle gesellschaftliche Themen?

Eine szientometrische Analyse am Beispiel Flucht und Migration mithilfe von Topic Modeling

André Bittermann¹ und Eva Maria Klos²

¹Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), Trier

²EU-Forschungsberatungs- und Koordinierungsstelle, Hochschule Trier

Zusammenfassung: Vor dem Hintergrund der Beziehung zwischen gesellschaftlichen Herausforderungen und entsprechenden Beiträgen der Wissenschaft wurde am Beispiel des Themenbereichs Flucht und Migration untersucht, inwiefern die psychologische Forschung jeweils aktuelle gesellschaftliche und politische Fragestellungen aufgreift, ob sie eine evidenzbasierte Grundlage für diese Fragestellungen schafft und welchen Teildisziplinen der Psychologie die Forschungsthemen zugeordnet werden können. Dazu wurden mit Structural Topic Modeling die Abstracts von 4.073 Publikationen aus den deutschsprachigen Ländern, dokumentiert in der psychologischen Referenzdatenbank PSYNDEX und veröffentlicht zwischen 1980 und 2017, analysiert. Es konnten 19 Themen identifiziert werden. Den stärksten zunehmenden Trend zeigten Traumatisierung von Flüchtlingen, transkulturelle Psychotherapie sowie die rehabilitative Behandlung von Patienten mit Migrationshintergrund. Im Bereich der Situation von Ausländern und Gastarbeitern, der Kriminalität von Jugendlichen sowie der sozialen Integration und Akkulturation mit Bezug zur ehemaligen DDR zeigte sich die deutlichste Abnahme der Wahrscheinlichkeit. Das Thema Sprachentwicklung von Migrantenkindern wies die höchste Wahrscheinlichkeit auf, von empirischen Studien behandelt zu werden, das Thema Identitätsentwicklung die höchste Wahrscheinlichkeit von nicht-empirischen Beiträgen. Zusammenfassend wird konstatiert, dass sich im Bereich von Flucht und Migration wesentliche gesellschaftliche und politische Entwicklungen in der psychologischen Fachliteratur widerspiegeln. Hinsichtlich empirischer Beiträge zu den Themen zeigt sich ein gemischtes Bild. Die meisten Themen haben einen klinisch-psychologischen Schwerpunkt, wobei jedoch auch andere Disziplinen vertreten sind. Methodisch kann festgehalten werden, dass der Topic-Model-Ansatz eine hilfreiche Methode für szientometrische Untersuchungen mit großen Textmengen darstellt.

Schlüsselwörter: Migration, Flucht, Szientometrie, Topic Modeling, Trends

Does Psychological Research Address Current Social Issues? A Scientometric Analysis of the Example of Refugees and Migration Using Topic Modeling

Abstract: In the context of the scientific investigation of social challenges, and drawing on the example of refugees and migration, we investigated to what extent psychological research takes up current social and political issues, whether it provides an evidence-based foundation for addressing social challenges, and to which subdisciplines of psychology the research topics can be assigned. Specifically, the abstracts of 4,073 publications from German-speaking countries, documented in the psychological reference database PSYNDEX and published between 1980 and 2017, were analyzed using structural topic modeling. A total of 19 topics were identified. The most strongly increasing trend was shown for topics concerning the traumatization of refugees, transcultural psychotherapy, and rehabilitative treatment. Topics concerning the situation of foreigners and guest workers, juvenile crime, as well as social integration and acculturation in relation to the former German Democratic Republic showed decreasing trends. The topic of language development of migrant children had the highest probability of being addressed by empirical studies; the topic of identity development was most probable for nonempirical contributions. In summary, important social and political developments in the field of refugees and migration are reflected in psychological literature. A mixed picture emerges with regard to empirical contributions on the topics. Most topics have a clinical-psychological focus, but other disciplines are also represented. From a methodological perspective, the topic modeling approach has proved to be a helpful method for scientometric investigations within large text corpora.

Keywords: migration, refugees, scientometrics, topic modeling, trends

Im Abschlussbericht zu den Perspektiven der Psychologie in Deutschland zählt der Wissenschaftsrat Migration zu

den sogenannten „Großen gesellschaftlichen Herausforderungen“, deren Bewältigung psychologische Erkennt-

nisse erfordert (Wissenschaftsrat, 2018, S. 83). Doch wird Flucht und Migration überhaupt in der psychologischen Forschung thematisiert? Und wenn ja, in welchen Bereichen liegen diesbezügliche Erkenntnisse vor? Damit psychologische Forschung zur Bewältigung gesellschaftlicher Herausforderungen beitragen kann, muss sie (1) diese Herausforderungen zunächst zum Thema machen. Grundlagenforschung und anwendungsorientierte Forschung, aber auch beispielsweise theoretische Diskussionen müssen sich damit befassen und ggf. bestehende Erkenntnisse darauf beziehen. Gerade angesichts der aktuellen „Replikationskrise“ sollten (2) Erkenntnisse dem Anspruch genügen, möglichst gesichert zu sein, um so zu einer evidenzbasierten Praxis beizutragen – nicht nur aus wissenschaftlicher, sondern auch aus ökonomischer und ethischer Sicht. Außerdem spielt es (3) eine Rolle, aus welcher fachlichen Perspektive konkrete Herausforderungen angegangen werden können. So wird sich etwa eine erfolgreiche Integration von unbegleiteten minderjährigen Geflüchteten nicht auf die Behandlung emotionaler Traumata (wo wir die Klinische Psychologie verorten können) beschränken lassen, sondern schließt vor dem Hintergrund von Sprach- und Identitätsentwicklung auch den Bereich schulischen Lernens ein (wo etwa die Entwicklungspsychologie und die Pädagogische Psychologie gefragt sind).

Von diesen Überlegungen ausgehend verfolgt die vorliegende Arbeit am Beispiel Flucht und Migration einen deskriptiven und explorativen Ansatz mit szientometrischen Methoden, um zu untersuchen, ob die psychologische Forschung durchlässig für aktuelle gesellschaftliche Themen ist. „Durchlässigkeit“ definieren wir dabei als inhaltliche Kongruenz (der Inhalt der Herausforderung spiegelt sich im Inhalt der Forschungsthemen wider) und zeitliche Korrespondenz (es können bezüglich Themenprävalenz Zusammenhänge zwischen Publikationsjahr und gesellschaftlichen Prozessen ausgemacht werden). Wenn die Psychologie durchlässig für gesellschaftliche Fragestellungen ist, dann sollten diese sich in den Forschungsthemen wiederfinden. Für den Zeitraum von 1980 bis 2017 wären im Bereich Flucht und Migration Themen zu erwarten, die in Zusammenhang stehen mit den innerbundesrepublikanischen Migrationsbewegungen nach der Wiedervereinigung, den fremdenfeindlichen Übergriffe in den 1990er Jahren, sowie den starken Migrationsbewegungen nach Deutschland 2015 und 2016, die unter anderem durch den syrischen Bürgerkrieg ausgelöst

waren. Denn all diese Ereignisse lösten in der deutschen Gesellschaft tiefgreifende Debatten aus, die die Psychologie betrafen: Psychologische Expertise war nicht nur in der gesellschaftlichen Aufnahme von Geflüchteten gefragt, sondern auch in deren Behandlung nach Gewalterfahrungen sowie in der Begleitung tiefgreifender gesellschaftlicher Umwälzungen, wie sie etwa die Wiedervereinigung bewirkt hatte. Eine thematisch breite Suche Anfang April 2018 in PSYNDEX¹, der Fachdatenbank für die Psychologie und ihrer Nachbargebiete in den deutschsprachigen Ländern, ergab für den Publikationszeitraum von 1980 bis 2017 4.933 Treffer im Bereich Flucht und Migration. Dabei bedarf es einer genaueren Betrachtung um festzustellen, welche Bezüge diese Publikationen zu dieser Thematik aufweisen.

Einen ersten Überblick zu den thematischen Schwerpunkten psychologischer Publikationen kann man durch das Klassifikationssystem der PSYNDEX-Einträge erlangen, welches den „Classification Codes“² im *Thesaurus of Psychological Index Terms* (Tuleya, 2007; ZPID, 2016) der American Psychological Association (APA) entspricht. So zeigt sich, dass von den 4.933 Fachbeiträgen 2.360 (47,84 %) in die Kategorien „Psychische und physische Störungen“ oder „Behandlung und Prävention“ bzw. deren Unterkategorien fallen³, welche den Kernbereichen der Klinischen Psychologie entsprechen. Der hierarchischen Tiefe dieses Klassifikationssystems folgend könnten nun noch genauere Untersuchungen vorgenommen werden, etwa wie viele Beiträge sich mit „Neurosen und Angststörungen“ befassen. Eine solche klassifikations- und häufigkeitsbasierte Herangehensweise stellt einen verschiedenfach eingesetzten szientometrischen Ansatz in der Psychologie dar (z. B. Preckel & Krampen, 2016). Allerdings können vorab definierte Klassifikationen für die gesamte psychologische Forschung nur einen begrenzten Detailgrad erreichen oder es müssen erst spezifische Klassifikationen von Forschungsthemen für den jeweiligen Bereich erstellt werden (Solga & Blickle, 2006). Dazu kommt, dass bei einem solchen top-down-Vorgehen nur das gefunden werden kann, wofür à priori eine Kategorie festgelegt wurde. Auch eine anschließende Häufigkeitsauszählung von standardisierten Schlagworten entsprechend des APA-Thesaurus würde nur einen begrenzten Zugewinn an Informationen erzielen, da unklar bliebe, welche Schlagwort-Kombinationen besonders häufig sind. Dabei wäre dies von entscheidender Bedeutung, da

¹ freier Zugriff über www.PubPsych.de; benutzte Suchsyntax mit global in PSYNDEX verfügbaren standardisierten Suchbegriffen: ((CT = „Refugees“ OR CT = „Human Migration“ OR CT = „Immigration“ OR CT = „Asylum Seeking“ OR CT = „Political Asylum“ OR CT = „Expatriates“ OR CT = „Foreign Workers“ OR CT = „Migrant Farm Workers“) PY < 2018 AND PY > 1979) DB = „PSYNDEX“

² Eine Liste aller Classification Codes kann unter <http://www.apa.org/pubs/databases/training/class-codes.aspx> abgerufen werden.

³ obige Suche ergänzt um „AND (SH = 32* OR SH = 33*)“

es einen Unterschied macht, ob etwa das Schlagwort „Refugees“ oft mit „Emotional Trauma“ und „War“ oder mit „Social Casework“ und „Child Welfare“ vergeben wurde. Eine manuelle Suche aller möglichen Schlagwort-Kombinationen ist angesichts eines Thesaurusumfangs von mehr als 6.000 Begriffen praktisch nicht zu bewerkstelligen. Schließlich würde eine klassische inhaltsanalytische Auswertung von Abstracts bei den knapp 5.000 Publikationen zum Thema Flucht und Migration ebenso an ihre Grenzen stoßen.

Topic Modeling (z. B. Blei, Ng & Jordan, 2003) ist eine Methode aus dem Maschinenlernen und stellt einen vielversprechenden Ansatz dar, um die eben genannten Nachteile des top-down-Filters und des hohen Aufwandes zu überwinden. Hierbei handelt es sich um eine automatisierte Inhaltsanalysetechnik, mit der die latente thematische Struktur eines gegebenen Textkorpus untersucht werden kann (Maier et al., 2018). Dies bedeutet, dass die Themen datengeleitet und ohne vorherige Annahmen, also bottom-up ermittelt werden. Ähnlich einer Clusteranalyse werden dazu diejenigen Wörter identifiziert, welche häufig gemeinsam innerhalb eines Textdokuments auftreten. Diese Cluster entsprechen den Themen im Korpus. Ausführlichere Beschreibungen zu Topic Modeling finden sich etwa bei Blei (2012) oder Maier und Kolleginnen und Kollegen (2018). Die Interpretation und Validität der gefundenen Themen hängt von der Qualität der analysierten Texte ab („Garbage-In-Garbage-Out“-Problematik). Hier stellen die in PSYNDEX nachgewiesenen Abstracts eine besonders günstige Ausgangslage dar, da von den PSYNDEX-Mitarbeitenden sichergestellt wird, dass ein Abstract auch tatsächlich den Inhalt des Beitrags widerspiegelt und den Kriterien der Vollständigkeit, Genauigkeit, Objektivität, Kürze, Verständlichkeit und Fachlichkeit entspricht (Gerards, Gerards, Kuhberg-Lasson, Singleton & Trierweiler, 2014, S. 4f.).

Fragestellungen

Am Beispiel von Flucht und Migration geht die vorliegende Studie folgenden inhaltlichen Fragen nach:

1. Ist die psychologische Forschung durchlässig für jeweils aktuelle gesellschaftliche und politische Fragestellungen?
2. Wird sie dem Anspruch gerecht, eine evidenzbasierte Grundlage für diese Fragestellungen zu schaffen?
3. Welchen psychologischen Teildisziplinen können die Forschungsthemen zugeordnet werden?

Dazu wird untersucht, welche Themen psychologische Publikationen aus den deutschsprachigen Ländern im Bereich Flucht und Migration behandeln und welche Trends zwischen 1980 und 2017 ausgemacht werden können. Um erste Hinweise zur Evidenzbasierung zu finden wird analysiert, welche Themen besonders empirisch erforscht werden und welche Themen blinde Flecken und Desiderate empirischer Forschung darstellen. Schließlich werden die Themeninhalte einem Klassifikationssystem psychologischer Forschung gegenübergestellt, um sie den Teildisziplinen zuordnen zu können.

Methode

Daten

Datengrundlage bildeten die Abstracts der in PSYNDEX nachgewiesenen Zeitschriftenartikel, Fachbücher, Buchkapitel, Berichte und Dissertationen aus den deutschsprachigen Ländern mit Stand vom 20.03.2018. Eingeschlossen wurden Beiträge, die zwischen 1980 und 2017 publiziert wurden, ein Abstract in deutscher Sprache aufwiesen, und mit mindestens einem der folgenden standardisierten Schlagwörter entsprechend des APA-Thesaurus (Tuleya, 2007; ZPID, 2016) versehen waren: „Refugees“, „Human Migration“, „Immigration“, „Asylum Seeking“, „Political Asylum“, „Expatriates“, „Foreign Workers“, „Migrant Farm Workers“⁴. Ausgeschlossen wurden Biografien und historische Quellen (wie etwa Wiederveröffentlichungen und Sammlungen bereits erschienener Beiträge), da diese die Abbildung zeitlicher Trends potenziell verzerren können. Dies resultierte in einem Korpus mit $N = 4.073$ Publikationen.

Software

Die Analysen wurden mit RStudio 1.1.414 (RStudio Team, 2016) basierend auf R 3.4.3 (R Core Team, 2017) durchgeführt. Für Text Mining und Topic Modeling wurde das Paket *stm* 1.3.3 (Roberts, Stewart & Tingley, in press) verwendet. Weitere Operationen wurden mit *dplyr* 0.7.4, *lattice* 0.20–35, *readr* 1.1.1, *splitstackshape* 1.4.2 und *Xmisc* 0.2.1 durchgeführt.

⁴ Die Suche erfolgte ausschließlich über standardisierte Suchbegriffe, da diese anhand des APA-Thesaurus vollständig definiert werden können und um potenzielle Verzerrungen durch eine subjektive Auswahl freier Suchbegriffe zu vermeiden.

Text Mining und Topic Modeling

Um Wörter auszuschließen, die zwar häufig in wissenschaftlichen Abstracts vorkommen, jedoch keinen inhaltlichen Beitrag liefern (wie etwa „Studie“, „Untersuchung“, „Ergebnis“ usw.), wurde eine stopword-Liste (einsehbar im ESM 1) erstellt. Wörter dieser Liste wurden aus dem Korpus entfernt. Aufbauend auf der Empfehlung von Banks, Wozny, Wesslen und Ross (2018) wurden weiterhin solche Wörter ausgeschlossen, die in weniger als sechs der Abstracts (im Folgenden entsprechend der Topic-Modeling-Terminologie als Dokument bezeichnet) enthalten waren, und es wurden Zahlen und Sonderzeichen entfernt. Auf das Kürzen von Wörtern auf ihren Wortstamm („stemming“) wurde einerseits zwecks besserer Lesbarkeit verzichtet, da Voruntersuchungen keinen Einfluss von stemming auf das Ergebnis zeigten. Andererseits können wichtige Differenzierungen („Flüchtlinge“ versus „Geflüchtete“) verloren gehen. Weiterhin zeigten Schofield und Mimno (2016), dass stemming negativen Einfluss auf die Topic-Stabilität haben kann. Das so aufbereitete Korpus wurde mit Structural Topic Modeling (Roberts et al., 2014) analysiert. Dabei wurden die Studienmethodik (empirisch vs. nicht-empirisch entsprechend der Metadaten in PSYNDEX) und das Publikationsjahr als Kovariaten der Topic-Prävalenz modelliert, das heißt es wurde angenommen, dass beide Merkmale einen Einfluss auf die Themenwahrscheinlichkeit haben. Dazu wurde mittels linearer Regression das Kriterium Dokument-Topic-Wahrscheinlichkeit θ auf die Prädiktoren Studienmethodik und Publikationsjahr regrediert. Die Anzahl k an Themen wurde über die „heldout likelihood“ (HL)-Methode ermittelt, was als der übliche Ansatz gilt (Kosinski, Wang, Lakkaraju & Leskovec, 2016). Basierend auf stichprobenhaften Tests für k wurden HL-Werte für 5 bis 40 Themen überprüft. Lineare zeitliche Trends der mittleren Dokument-Topic-Wahrscheinlichkeiten θ wurden entsprechend Griffiths und Steyvers (2004) über lineare Regressionen ermittelt. Der vollständige R-Code der Analysen ist im ESM 2 zu finden.

Ergebnisse

Das Vokabular im Korpus bestand aus 5.569 verschiedenen Wörtern und es konnten 19 Themen identifiziert werden. Der Anteil empirischer Studien betrug insgesamt 36,1%, der Anteil nicht-empirischer Beiträge 63,9%. Tabelle 1 zeigt für das jeweilige Thema die wahrscheinlichsten Wörter sowie die sowohl häufigen als auch exklusiven Wörter.

Abbildung 1 zeigt für jedes Thema den Verlauf der mittleren Dokument-Topic-Wahrscheinlichkeit θ über die Zeit. Signifikante ($p < .01$) Steigungen konnten für neun positive und vier negative Trends gefunden werden. Den stärksten zunehmenden linearen Trend zeigten Publikationen zu Traumatisierung von Flüchtlingen (Thema 12), transkultureller Psychotherapie (Thema 14) sowie der rehabilitativen Behandlung von Patienten mit Migrationshintergrund (Thema 18). Am stärksten nahm die Wahrscheinlichkeit für die Themen Situation von Ausländern und Gastarbeitern (Thema 9), Kriminalität von Jugendlichen (Thema 16) sowie soziale Integration und Akkulturation mit Bezug zur ehemaligen DDR (Thema 13) ab.

Abbildung 2 zeigt, welche Themen mit höherer Wahrscheinlichkeit von nicht-empirischen bzw. empirischen Beiträgen aufgegriffen wurden. Abgebildet sind für jedes Thema die durch ein Regressionsmodell erwarteten Differenzen in der mittleren Dokument-Topic-Wahrscheinlichkeit θ zwischen empirischen bzw. nicht-empirischen Studien einschließlich der 95%-Konfidenzintervalle. Bis auf Fremdenfeindlichkeit und Diskriminierung (Thema 5) wiesen alle Themen signifikante Unterschiede auf. Das Thema Sprachentwicklung von Migrantenkindern (Thema 8) wies die höchste Wahrscheinlichkeit auf, von empirischen Studien behandelt zu werden, gefolgt von den Themen psychische Störungen bei Migranten (Thema 4) und Schüler mit Migrationshintergrund (Thema 2: „migrationshintergrund, schüler, schülern“). Die höchste Wahrscheinlichkeit, von nicht-empirischen Beiträgen aufgegriffen zu werden, hatte das Thema der Identitätsentwicklung (Thema 17), gefolgt von interkultureller Beratung (Thema 10) und transkultureller Psychiatrie (Thema 14).

Analog zu Griffiths und Steyvers (2004) werden in Abbildung 3 zum Zweck der inhaltlichen Validierung die Themen dem PSYNDEX-Klassifikationsschema gegenübergestellt. In diesem Levelplot werden die mittleren Dokument-Topic-Wahrscheinlichkeiten θ der Themen getrennt nach Klassifikation der Dokumente dargestellt. Zwecks Übersichtlichkeit sind nur die zehn häufigsten Hauptklassifikationen dargestellt (das komplette Klassifikationssystem umfasst insgesamt 157 Kategorien), wobei diese zehn Hauptklassifikationen (inklusive Unterkategorien) 96,25% aller den Dokumenten zugewiesenen Klassifikationen abdecken. Je dunkler eine Zelle ist, desto höher ist das mittlere θ ; das heißt, dass ein Thema in Dokumenten mit dieser Klassifikation besonders hervorsteht. Bei annähernder Gleichverteilung der Themen innerhalb der Klassifikation sind die Zellen der entsprechenden Zeile weiß (da für 19 Themen die durchschnittliche Themen-Wahrscheinlichkeit $1/19 = 0,053$ beträgt; siehe Einteilung der Graustufen in Abbildung 3). So zeigt sich etwa, dass Dokumente, die mit „36** Arbeits- und Organisationspsychologie“ klassifiziert sind, am wahrscheinlichsten

Tabelle 1. Beschreibung der Themen anhand ihrer zehn wahrscheinlichsten Wörter und anhand derjenigen zehn Wörter, die sowohl häufig als auch exklusiv im jeweiligen Thema sind

Thema	wahrscheinlichste Wörter	häufige und exklusive Wörter
1	ausland, interviews, unternehmen, sozialen, mitarbeitern, qualitative, unterstützung, expatriates, qualitativen, beruflichen	unternehmen, expatriates, auslandseinsatz, ausland, entsendung, mobilität, mitarbeitern, arbeitsmarkt, führungskräften, qualitative
2	migrationshintergrund, schüler, schülern, schülerinnen, deutschland, schule, internationalen, schulen, kompetenzen, herkunft	schüler, schülerinnen, iglu, mathematik, mathematischen, lesekompetenz, schülern, pisa, sekundarstufe, timss
3	kinder, arbeit, jugendlichen, kindern, familien, schule, geflüchteten, eltern, erfahrungen, unterstützung	geflüchteten, geflüchtete, jugendhilfe, schritte, projekt, geflüchteter, pädagogischen, unterstützt, bindung, fachkräfte
4	migranten, störungen, deutschen, psychische, psychischen, deutschland, gesundheit, beschwerden, frauen, psychischer	beschwerden, prävalenz, lebenszufriedenheit, probanden, depressiver, störungen, belastung, psychischer, depression, depressiven
5	fremdenfeindlichkeit, diskriminierung, einstellungen, fremden, ausländern, gewalt, sozialen, rassismus, ausländerfeindlichkeit, vorurteile	fremdenfeindlichkeit, rassismus, rechtsextremismus, vorurteilen, ausländerfeindlichkeit, sozialpsychologische, vorurteile, diskriminierung, fremdenfeindlicher, fremdenfeindlichen
6	patienten, behandlung, erfahrungen, therapeuten, probleme, therapie, therapeutischen, türkischen, besonderheiten, gesammelt	gesammelt, therapeuten, ärzte, ärzten, patienten, ambulanzen, psychotherapeuten, dolmetscher, klinik, gruppentherapie
7	jugendlichen, migrationshintergrund, jugendliche, eltern, familien, jungen, deutschland, risikofaktoren, werte, faktoren	risikofaktoren, jugendalter, jugendlichen, werte, jugendliche, elterlicher, delinquenz, autonomie, wahrgenommene, diskriminierungserfahrungen
8	kinder, kindern, eltern, migrationshintergrund, deutschen, sprache, deutsch, sprachlichen, entwicklung, alter	phonologischen, erzieherinnen, deutsch, kindern, mütter, kinder, vorschulalter, sprachförderung, kindergarten, erstsprache
9	deutschland, bundesrepublik, ausländischer, situation, ausländischen, frauen, deutschen, familien, integration, probleme	ausländischer, bundesrepublik, arbeitnehmer, ausländische, ausländischen, gastarbeiter, arbeiter, demographische, ausländerinnen, arbeitsmigranten
10	migranten, beratung, interkulturellen, interkulturelle, versorgung, arbeit, psychosozialen, migrantinnen, kompetenz, interkultureller	beratung, beratungsarbeit, interkultureller, öffnung, berater, klienten, systemischen, psychosozialen, interkulturelle, interkulturellen
11	heimat, psychoanalyse, fremde, emigration, deutschland, folgen, leben, auswirkungen, psychoanalytischen, entwicklung	psychoanalyse, emigration, heimat, emigranten, jüdischen, vertreibung, tod, gespräche, entwurzelung, freud
12	flüchtlingen, flüchtlinge, traumatisierten, belastungsstörung, trauma, behandlung, arbeit, posttraumatischen, ptbs, traumatisierung	posttraumatischen, flüchtlingen, traumatisierten, ptbs, flüchtlinge, folteropfer, belastungsstörung, folter, traumatisierung, flüchtlingskinder
13	akkulturation, soziale, deutschen, sozialen, immigranten, identität, integration, identifikation, assimilation, republik	republik, demokratischen, ost, akkulturation, westdeutschland, übersiedlern, ddr, immigranten, übersiedler, identifikation
14	migranten, psychotherapie, psychiatrie, transkulturellen, migration, transkulturelle, psychiatrischen, behandlung, interkulturelle, versorgung	transkulturellen, psychiatrie, transkulturelle, leitlinien, psychotherapie, sprach, machleidt, begutachtung, psychiatrische, heise
15	integration, gesellschaft, kultur, kulturellen, kulturelle, erziehung, gesellschaftlichen, diskussion, forschung, kultureller	erziehung, anerkennung, gesellschaft, werten, gesellschaftlichen, religiosität, pädagogik, diskussion, vermittlung, religiösen
16	deutschen, mädchen, jugendlichen, gewalt, türkischen, türkischer, jungen, herkunft, deutsche, jugendliche	mädchen, polizeilichen, fernsehen, kriminalität, nichtdeutschen, griechischen, kriminalstatistik, nordrhein-westfalen, gewalt, fast
17	migration, entwicklung, migranten, familie, kultur, kulturellen, prozess, identität, überlegungen, eltern	adoleszenz, phase, fallbeispiel, patientin, angestellt, migration, migrationsprozesses, prozess, migrationsprozess, gegenübertragung
18	migrationshintergrund, patienten, behandlung, türkischem, faktoren, psychologie, rehabilitation, versorgung, interviews, patientinnen	türkischem, language, rehabilitation, kliniken, stanat, psychologie, students, schmidt, migrationshintergrund, väter
19	kommunikation, politische, deutschland, migranten, politischen, migrantinnen, letzten, möglichkeit, europa, betroffenen	partizipation, artikel, kommunikation, letzten, europäischen, akteure, europa, möglichkeit, bemühungen, politische

Thema 1 („ausland, interviews, unternehmen“) behandeln. Währenddessen greifen Dokumente mit Klassifikation „33** Behandlung und Prävention“ vor allem drei Themen auf: Thema 6 („patienten, behandlung, erfahrungen“), 10 („migranten, beratung, interkulturellen“) und 12 („flüchtlingen, flüchtlinge, traumatisierten“). Die

Klassifikation „29** Gesellschaftliche Systeme“ verteilt sich annähernd gleich auf alle Themen.

Zeitlicher Verlauf der mittleren Dokument-Topic-Wahrscheinlichkeiten

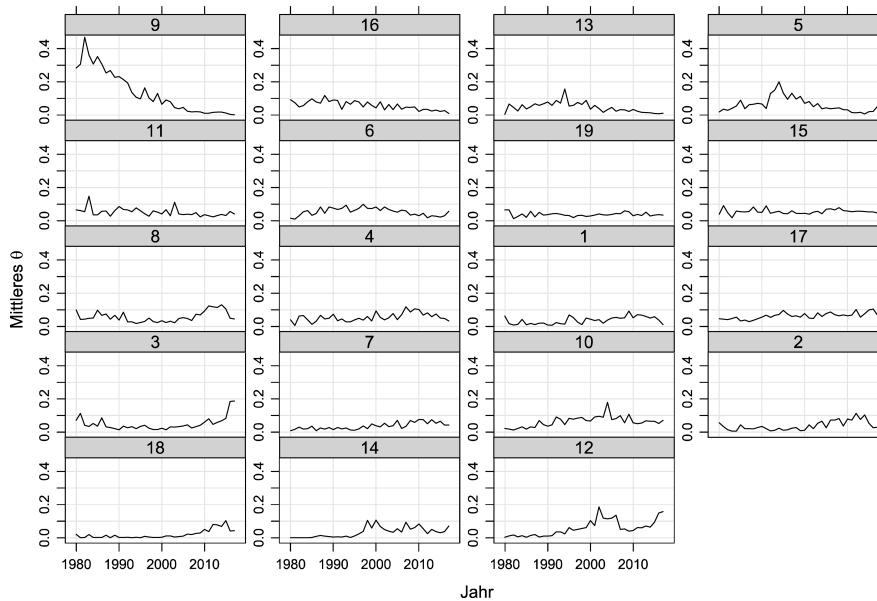


Abbildung 1. Mittlere Dokument-Topic-Wahrscheinlichkeiten (θ) der Themen, zeilenweise sortiert nach Steigung des linearen Trends von abfallend (oben links, Topic 9) bis aufsteigend (unten rechts, Topic 12). Die Themen werden in Tabelle 1 beschrieben.

Erwartete Differenzen der Topic-Wahrscheinlichkeiten nach Studienmethodik

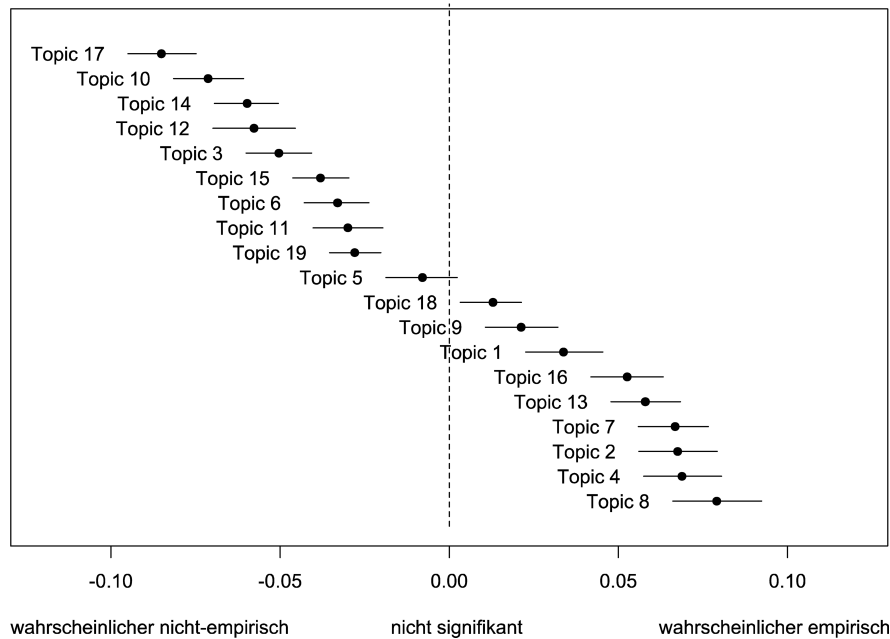


Abbildung 2. Regressionsbasierte erwartete Differenzen der Themenwahrscheinlichkeiten (θ) nach Studienmethodik (einschließlich 95%-Konfidenzintervalle). Themen der linken Seite werden mit höherer Wahrscheinlichkeit von nicht-empirischen Beiträgen behandelt, Themen der rechten Seite von empirischen Studien. Die Themen werden in Tabelle 1 beschrieben.

Diskussion

Die vorliegende Studie verwendete Structural Topic Modeling (Roberts et al., 2014), um Themen mit Bezug zu Flucht und Migration in psychologischen Fachpublikationen aus den deutschsprachigen Ländern für den Zeitraum von 1980 bis 2017 zu identifizieren. Dies diente der Beantwortung der Fragen, ob die psychologische Forschung

durchlässig für jeweils aktuelle gesellschaftliche und politische Themen ist, ob sie eine evidenzbasierte Grundlage für diese Fragestellungen schafft und welchen psychologischen Teildisziplinen die Themen zugeordnet werden können.

Verteilung der Themen auf die zehn häufigsten Hauptklassifikationen

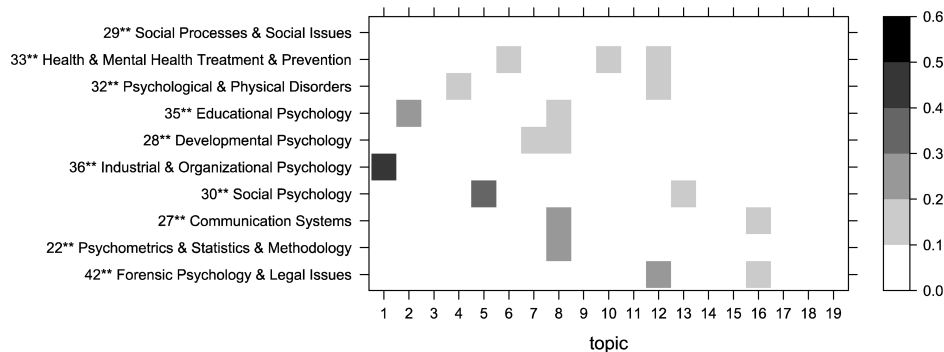


Abbildung 3: Verteilung der Themen auf die zehn häufigsten Hauptklassifikationen. Die Klassifikationen sind nach ihrer Häufigkeit sortiert (von oben nach unten). Je dunkler die Zelle, desto höher ist die mittlere Dokument-Thema-Wahrscheinlichkeit (θ). Z. B. hatten Dokumente, welche mit „3600 Arbeits- und Organisationspsychologie“ klassifiziert waren, die höchste Wahrscheinlichkeit, Thema 1 zu behandeln. Die Themen werden in Tabelle 1 beschrieben. Eine Beschreibung der Hauptklassifikationen kann unter <http://www.apa.org/pubs/databases/training/class-codes.aspx> abgerufen werden

Forschungstrends im historischen Kontext

Zur Validierung der Themeninhalte kann überprüft werden, ob ihre zeitlichen Verläufe mit gesellschaftlichen Entwicklungen und Ereignissen korrespondieren (z. B. Evans, 2014). Kausalitäten können bei einer solchen inhaltlichen Validierung mit den genutzten Methoden nicht festgestellt werden, aber ablesbare inhaltliche Schwerpunktsetzungen in der psychologischen Fachliteratur können in ihren Zeitverläufen in gesellschaftliche Kontexte eingeordnet werden. Dabei ist zu beachten, dass Publikationen von Fachbeiträgen einer Latenzzeit von ein bis zwei Jahren unterliegen (Krampen, Montada, Müller & Schui, 2004), wobei dieser Zeitraum sowohl durch Faktoren wie die Dauer der Datenerhebung (vgl. hierzu Abbildung 3 zur Unterscheidung eher empirisch vs. nicht-empirisch adressierter Themen) oder des Review- und Veröffentlichungsprozesses als auch das gesamtgesellschaftliche Interesse an der Thematik sehr unterschiedlich ausfallen kann. Somit soll die erste Fragestellung zur Durchlässigkeit psychologischer Fachbeiträge für gesellschaftliche und politische Themen aufgegriffen werden.

Die identifizierten Themen folgen in einigen Zeitabschnitten deutlich den politischen und gesellschaftlichen Debatten, die zu jenen Zeitpunkten geführt wurden (vgl. im Folgenden Abbildung 1): Die vermehrte Beschäftigung der psychologischen Fachliteratur mit der Akkulturation Anfang der 1990er Jahre (Thema 13, welches die Begriffe „Integration“ und „Assimilation“ wie auch „Identität“ und „Identifikation“ einschließt) könnte einerseits von einem Aufgreifen der Migration im Zuge des Zerfalls des Ostblocks (vgl. zur Zuwanderung aus Osteuropa: Herbert, 2001, S. 273–278), andererseits vom Ringen der zuvor jahrzehntelang getrennten Gesellschaften Ost- und Westdeutschlands mit der deutschen Wiedervereinigung zeu-

gen. So fungierten „[a]ngesichts der enormen deutsch-deutschen Integrationsprobleme [...] Ausländer für manche Politiker, Medien und Bürger als Sündenböcke“ (Pagenstecher, 2008, S. 611). Die fremdenfeindlichen, zum Teil tödlichen Angriffe und die ausufernde Gewalt gegen Migrantinnen und Migranten – nicht nur in Hoyerswerda, Rostock und Solingen – bestimmten die Tagespresse. Auch die psychologische Fachliteratur setzte in dieser Zeit inhaltliche Schwerpunkte im Themenfeld der Migration häufig rund um Fremdenfeindlichkeit, Rassismus, Rechts extremismus und Gewalt (Thema 5) und zeigte sich damit durchlässig für aktuelle gesellschaftliche Herausforderungen der Bundesrepublik.

Obwohl in beiden Themen eine enge Verbindung zu den aktuellen gesellschaftlichen Debatten festzustellen ist, wäre es falsch, von der Zahl der nach Deutschland migrierenden Menschen auf die Präsenz des Themas Migration in der Fachliteratur zu schließen: Obgleich die Zahl der Asylanträge zu Beginn der 2000er Jahre auf vergleichsweise niedrigem Stand war, ist in diesem Zeitraum eine klare Zunahme von Publikationen bezogen auf Thema 12, das unter anderem die Stichworte Flüchtlinge, Trauma und Folter umfasst, erkennbar. Warum ist dieses Thema gerade zu Beginn des neuen Jahrtausends von Interesse für die psychologische Fachliteratur? Der NATO-Einsatz im Kosovo und die Zustimmung des Bundestages zur deutschen Beteiligung unter Bundeskanzler Gerhard Schröder 1998 konfrontierten die deutsche Bevölkerung Anfang der 2000er Jahre mit den Themen Gewalt und Trauma. Die deutsche Politik debattierte zugleich über das Staatsangehörigkeitsrecht und die künftige Einwanderungspolitik. Der 11. September 2001, die folgende weitreichende Debatte um die Legitimität von Foltermaßnahmen als Verhörmethode bei Terroristen sowie der Irakkrieg ab 2003, an dem sich Deutschland nicht militärisch

beteiligte, verstärkten diese Sensibilisierung. Der Abu-Ghuraib-Folterskandal, der 2004 in den Medien publik wurde, brachte das Thema der Folter auf die Titelseiten der Zeitungen weltweit (Paul, 2008). Dass sich die psychologische Fachliteratur ab Beginn des neuen Jahrtausends also verstärkt mit den Themen Flucht, Folter, Traumatisierung und ihrer Behandlung insgesamt beschäftigte, spricht für eine Durchlässigkeit der Psychologie für aktuelle Fragestellungen. Diese Zunahme ist zum Teil aber auch auf ein methodisches Artefakt zurückzuführen, da 2002 eine Jubiläumsfestschrift eines Behandlungszentrums für Folteropfer erschien, deren Einzelbeiträge 15 % aller eingeschlossenen Beiträge in diesem Jahr ausmachten. Geflüchtete, Folter, Traumata und die Behandlung psychischer Auswirkungen von Gewalterfahrungen (Thema 12) beschäftigten die psychologische Fachliteratur erneut ab 2015 intensiv. Es kann angenommen werden, dass die Fachliteratur zu diesem Zeitpunkt auf die stark steigende Zahl von Menschen reagierte, deren Flucht von Krieg und Gewalt ausgelöst war (etwa durch den syrischen Bürgerkrieg) und die in Deutschland psychologischer Unterstützung bedurften.

Betrachtet man die Gesamtheit der Trends in ihren Verläufen von 1980 bis heute, lassen sich auch Indizien für einen Wandel in der Sprache der psychologischen Fachliteratur nachzuvollziehen, der von einer sich ändernden Sichtweise auf die Migration zeugt: Die Bezeichnung von Migranten changiert im Laufe der Zeit von „Gastarbeitern“ (Thema 9), „Ausländern“ (Themen 9 und 5) über „Flüchtlinge“ (Thema 12) hin zu „Geflüchteten“ (Thema 3). Einerseits weist diese Entwicklung auf sich wandelnde Ursachen von Migrationsbewegungen hin (vgl. zu diesen Ursachen: Herbert, 2001), andererseits auf das Bemühen der Autorinnen und Autoren, sprachliche Diskriminierung abzubauen. Eine genauere Untersuchung der zeitlichen Veränderung sprachlicher Bezeichnungen in der psychologischen Fachliteratur könnte Gegenstand zukünftiger Forschung sein.

Zusammenfassend kann festgehalten werden, dass sich die psychologische Fachliteratur sowohl den Folgen der deutsch-deutschen Wiedervereinigung als auch dem verstärkten Zuzug von Geflüchteten ab 2015 ausgiebig widmete und sich damit aktueller gesellschaftlicher Fragestellungen annahm (vgl. auch den Anstieg von Thema 3 zu Projekten mit Geflüchteten ab 2015 und von Thema 2 zu Schülerinnen und Schülern mit Migrationshintergrund ab 2000). Die starke Beschäftigung mit den Themen Flucht und Folter Anfang der 2000er Jahre, als die Zahl der Asylanträge in Deutschland vergleichsweise niedrig war,

lässt eine Vermutung zu: Die Wahrnehmung, welche Themen für die psychologische Fachwelt als relevant eingestuft wurden, veränderte sich um die Jahrtausendwende im Zuge der Digitalisierung und der Globalisierung. Vormalig räumlich weit entfernte Gewalterfahrungen kamen nicht nur durch Migranten, sondern auch durch die sich globalisierende Medienwelt unmittelbarer als zuvor nach Deutschland und mahnten die deutsche Fachwelt, Lösungswege zur Linderung der Folgen von Gewalterfahrungen zu erarbeiten.

Empirische Beiträge zu den Themen

Im Rahmen der zweiten Fragestellung sollten erste Hinweise auf den Stand der Evidenzbasierung ausgemacht werden. Für 18 der 19 Themen konnten signifikante Unterschiede in der erwarteten Differenz der mittleren Dokument-Thema-Wahrscheinlichkeit zwischen empirischen bzw. nicht-empirischen Studien gefunden werden. Bei den Themen mit hoher Wahrscheinlichkeit, von nicht-empirischen Beiträgen aufgegriffen zu werden, könnte dies ein Hinweis auf bestehende empirische Forschungslücken und -desiderate sein. Dabei gilt es aus Perspektive des Forschungsprozesses zu berücksichtigen, dass bei neuen Forschungsthemen theoretische Studien, Diskussionen oder Fallberichte einer empirischen Prüfung vorausgehen. Für Thema 17 zur Identitätsentwicklung beispielsweise sind die fünf für dieses Thema repräsentativsten Beiträge allesamt Fallberichte (eine Liste mit den Nachweisnummern der für jedes Thema zehn repräsentativsten Publikationen ist im ESM 3 zu finden). Unter den 100 typischsten Beiträgen dieses Themas finden sich 64 Fall- und Erfahrungsberichte, bei Thema 12 zur Traumatisierung von Flüchtlingen sind dies 33. Mit Blick auf die nicht-empirischen klinisch-psychologischen Themen fallen relativ hohe Anteile von Diskussionen von Versorgungsstrukturen auf. So sind dies bei Thema 10 (interkulturelle Beratung) 18% und bei Thema 14 (transkulturelle Psychiatrie) 16%. Dies ist vor dem Hintergrund der Herausforderungen in der Gesundheitsversorgung einer Gesellschaft mit rasch ansteigendem Anteil von Migrantinnen und Migranten zu sehen. Auf der Seite der Themen mit hoher Wahrscheinlichkeit, von empirischen Studien adressiert zu werden, stehen vor allem Thema 8 (Sprachentwicklung), Thema 4 (psychische Störungen) und Thema 2 (Schüler mit Migrationshintergrund). Sprachentwicklung darf im Kontext vieler Testverfahren⁵ (siehe auch Abbildung 3 und den Zusammenhang mit Klassifikation „22**

⁵ Eine Suche im Datenbanksegment PSYNDEX Tests via „(CT = “language development” seg = Test) DB = PSYNDEX“ erbrachte im April 2018 167 Treffer.

Psychometrie, Statistik und Methodik“) als gut operationalisiertes Konstrukt angesehen werden, epidemiologische und Interventionsstudien zu psychischen Störungen sind per se empirisch und ein Blick auf Tabelle 1 bei Thema 2 zeigt, dass hier mit IGLU, PISA und TIMSS drei große empirische Bildungsstudien aufgeführt sind. Solche Themen mit hoher Wahrscheinlichkeit, von empirischen Studien aufgegriffen zu werden, könnten ein Ausgangspunkt für Forschungssynthesen sein, um genauere Erkenntnisse zur empirischen Evidenz zu erhalten. Zusammenfassend kann konstatiert werden, dass sich hinsichtlich empirischer Studienanteile der Themen und entsprechender Beiträge zur Evidenzbasierung ein gemischtes Bild zeigt.

Bezüge zu psychologischen Teildisziplinen

Für die dritte Fragestellung sollte untersucht werden, welchen psychologischen Teildisziplinen die gefundenen Themen zugeordnet werden können. Da mittels Topic Modeling häufig miteinander vorkommende Wörter als Themen identifiziert werden, deren semantische Bedeutung jedoch nicht direkt offengelegt wird, werden idealerweise die Themeninhalte zur besseren Interpretation anhand statistischer oder externer Kriterien validiert (Maier et al., 2018). Bezogen auf die Frage, welche Themen die psychologische Forschung im Bereich Flucht und Migration behandelt, können Klassifikationen psychologischer Forschung als externe Kriterien der Themenvolidität herangezogen werden. Für vier Themen (Thema 4: „migranten, störungen, deutschen“, Thema 6: „patienten, behandlung, erfahrungen“, Thema 10: „migranten, beratung, interkulturellen“ und Thema 12: „flüchtlings, flüchtlinge, traumatisierten“; vgl. Abbildung 3 und Tabelle 1) kann anhand der Klassifikationen ein Bezug zur Klinischen Psychologie festgestellt werden. Da die Themen der transkulturellen Psychotherapie (Thema 14) und der rehabilitativen Behandlung von Patienten mit Migrationshintergrund (Thema 18) inhaltlich ebenfalls diesem Bereich zugeordnet werden können, macht somit unter den klassischen Anwendungsfächern die Klinische Psychologie den größten Teil der gefundenen Themen zu Flucht und Migration aus. Dies entspricht der insgesamt starken Dominanz klinisch-psychologischer Beiträge in der Psychologie (Krampen & Perrez, 2015). Der Pädagogischen Psychologie können zwei Themen zugeordnet werden (Thema 2: „migrationshintergrund, schüler, schülern“ und Thema 8: „kinder, kindern, eltern“), ebenso wie der Entwicklungspsychologie (Thema 7: „jugendlichen, migrationshintergrund, jugendlichen“ und Thema 8: „kinder, kindern, el-

tern“), der Sozialpsychologie (Thema 5: „fremdenfeindlichkeit, diskriminierung, einstellungen“ und Thema 13: „akkulturation, soziale, deutschen“) und der Forensischen Psychologie (Thema 12: „flüchtlings, flüchtlinge, traumatisierten“ und Thema 16: „deutschen, mädchen, jugendlichen“). Der Arbeits- und Organisationspsychologie entspricht Thema 1 („ausland, interviews, unternehmen“).

Methodische Diskussion

Dem explorativ-deskriptiven und szientometrischen Charakter dieser Studie entsprechend wurden die inhaltlichen Forschungsfragen und ihre Kriterien nicht deduktiv beispielsweise aus einem allgemeinen Modell zum Zusammenhang von gesellschaftlichen und wissenschaftlichen Diskursen abgeleitet. Dies wäre für zukünftige Forschung sicherlich wünschenswert, um Hypothesen zu generieren und zu prüfen. So könnten etwa das Konzept der „Durchlässigkeit“ und seine empirischen Voraussetzungen weiter präzisiert werden. Im Kontext wissenschaftsphilosophischer Überlegungen könnte diskutiert werden, was psychologische Forschung aufgrund aktuell drängender Problemlagen überhaupt leisten kann oder muss – vor allem angesichts sehr schneller Veränderungen der Gesellschaft, wie sie durch die sogenannte „Flüchtlingskrise“ 2015 ausgelöst wurden.⁶ Auf diese Weise könnte die zeitliche Korrespondenz von gesellschaftlichen Entwicklungen und Publikationsspitzen psychologischer Fachbeiträge über eine deskriptive Betrachtung hinausgehend geprüft werden, etwa durch Vergleich einer Trendprognose (anhand der Daten vor einem Ereignis) mit dem tatsächlichen Verlauf.

Die 19 gefundenen Themen decken ein breites Spektrum an Inhalten ab, welche nur teilweise mit dem APA-Klassifikationssystem korrespondieren (siehe Abbildung 3). Dass aus Sicht der Klassifikationskategorien Beiträge zu „29** Gesellschaftliche Strukturen“ am häufigsten waren, ist vor dem Hintergrund zu sehen, dass vor allem die Unterkategorie „2910 Soziale Strukturen“ explizit die Folgen von Migration erfasst (Gerards et al., 2014, S. 41). Dies erklärt, warum sich diese Klassifikation in Abbildung 3 nahezu gleich auf alle Themen verteilt. Als Themen mit deutlich zunehmendem Trend konnten die Traumatisierung von Flüchtlingen (Thema 12), transkulturelle Psychotherapie (Thema 14) sowie die rehabilitative Behandlung von Patienten mit Migrationshintergrund (Thema 18) ermittelt werden. Dies korrespondiert mit zwei der Brennpunkthemen der gesamten psychologischen Forschung bei Bittermann und Fischer (2018).

⁶ Diesen Gedanken verdanken wir einer anonymen Gutachterin/einem anonymen Gutachter.

In dieser Studie wurden wissenschaftliche Abstracts in deutscher Sprache untersucht, weshalb insgesamt knapp 15% der Beiträge zum Thema Flucht und Migration aufgrund englischer Abstract-Sprache ausgeschlossen wurden (englische Begriffe hätten eigene Themen gebildet). Außerdem weist Thema 18 darauf hin (siehe Tabelle 1), dass englische Begriffe auch in deutschsprachigen Abstracts vorkommen können. Topic-Modeling-Verfahren, welche Mehrsprachigkeit berücksichtigen (z. B. Vulić, De Smet, Tang & Moens, 2015), könnten in zukünftigen Studien eine Möglichkeit darstellen, weitere Publikationen einzubeziehen. Dies ist besonders angesichts des zunehmenden Trends englischsprachiger Publikationen aus den deutschsprachigen Ländern wichtig (Schui & Krampen, 2016).

Trends wurden analog zu Griffiths und Steyvers (2004) mittels linearer Regression unter Einbezug des gesamten Zeitraumes, also von 1980 bis 2017, ermittelt. Denkbar wäre, den Zeitraum kürzer zu fassen, um besonders aktuelle Themen zu identifizieren. Somit würden Themen, die erst kürzlich einen starken Anstieg verzeichnen (siehe Thema 3 in Abbildung 1) deutlicher hervortreten. Die Auswahl eines geeigneten Zeitfensters müsste dabei unter Berücksichtigung der Latenzzeit des Publikationsprozesses getroffen werden.

Schlussfolgerung

Im Bereich von Flucht und Migration spiegeln sich wesentliche gesellschaftliche und politische Entwicklungen in der psychologischen Fachliteratur wider. Hinsichtlich empirischer Beiträge zu den Themen zeigt sich ein gemischtes Bild. Die meisten Themen haben einen klinisch-psychologischen Schwerpunkt, wobei jedoch auch andere Disziplinen vertreten sind. Methodisch kann konstatiert werden, dass der Topic-Model-Ansatz eine geeignete Methode für szientometrische Untersuchungen mit großen Textmengen darstellt. Für zukünftige Forschung kann von Interesse sein, ob mithilfe dieser Methode auch überprüft werden kann, ob – wie vom Wissenschaftsrat (2018) gefordert – der Transfer psychologischer Erkenntnisse in die Öffentlichkeit gelingt bzw. wie dazu beigetragen werden kann.

Literatur

Banks, G. C., Woznyj, H. M., Wesslen, R. S. & Ross, R. L. (2018). A Review of Best Practice Recommendations for Text Analysis in R (and a User-Friendly App). *Journal of Business and Psychology*, 33, 445–459. <https://doi.org/10.1007/s10869-017-9528-3>
 Bittermann, A. & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, 226 (1), 3–13. <https://doi.org/10.1027/2151-2604/a000318>

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55 (4), 77–84. <https://doi.org/10.1145/2133806.2133826>
 Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5-993>
 Evans, M. S. (2014). A computational approach to qualitative analysis in large textual datasets. *PLoS ONE* 9 (2): e87908. <https://doi.org/10.1371/journal.pone.0087908>
 Gerards, A., Gerards, M., Kuhberg-Lasson, V., Singleton, K. & Trierweiler, L. (2014). Dokumentationsmanual PSYINDEX (aktualisierte Ausgabe). Trier: ZPID. Zugriff am 03.04.2018. Verfügbar unter: <https://www.psyindex.de/download/PSYINDEXmaterial/psyindex-dokumentationsmanual.pdf>
 Griffiths, T. L. & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101 (Suppl. 1), 5228–5235. <https://doi.org/10.1073/pnas.0307752101>
 Herbert, U. (2001). Geschichte der Ausländerpolitik in Deutschland. Saisonarbeiter, Zwangsarbeiter, Gastarbeiter, Flüchtlinge. München: C.H. Beck.
 Kosinski, M., Wang, Y., Lakkaraju, H. & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, 21, 493–506. <https://doi.org/10.1037/met0000105>
 Krampen, G. & Perrez, M. (2015). Publikationsschwerpunkte der Klinischen Psychologie und Psychotherapieforschung im deutsch-versus angloamerikanischen Bereich. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 44, 181–196. <https://doi.org/10.1026/1616-3443/a000320>
 Krampen, G., Montada, L., Müller, M. & Schui, G. (2004). Erfahrungen und Empfehlungen zur Internationalisierung der psychologischen Forschung im deutschsprachigen Bereich – Befunde aus Experteninterviews: Ausführlicher Bericht. *ZPID Science Information Online*, 4 (2). Zugriff am 03.04.2018. Verfügbar unter: <http://www.zpid.de/pub/info/experteninterviews.pdf>
 Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A. et al. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12 (2–3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
 Pagenstecher, C. (2008). „Das Boot ist voll“. Schreckensvisionen des vereinten Deutschland. In G. Paul (Hrsg.). *Das Jahrhundert der Bilder, Band II. 1949 bis heute* (S. 606–613). Göttingen: Vandenhoeck & Ruprecht.
 Paul, G. (2008). Der „Kapuzenmann“. Eine globale Ikone des beginnenden 21. Jahrhunderts. In G. Paul (Hrsg.). *Das Jahrhundert der Bilder, Band II. 1949 bis heute* (S. 702–709). Göttingen: Vandenhoeck & Ruprecht.
 Preckel, F. & Krampen, G. (2016). Entwicklung und Schwerpunkte in der psychologischen Hochbegabungsforschung. *Psychologische Rundschau*, 67 (1), 1–14. <https://doi.org/10.1026/0033-3042/a000289>
 R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [Computer software]. Verfügbar unter: <https://www.R-project.org/>
 Roberts, M. E., Stewart, B. M. & Tingley, D. (in press). stm: R Package for Structural Topic Models. *Journal of Statistical Software*. Zugriff am 03.04.2018. Verfügbar unter: <https://cran.r-project.org/web/packages/stm/vignettes/stmVignette.pdf>
 Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K. et al. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58 (4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
 RStudio Team. (2016). *RStudio: Integrated development for R* [Computer software]. Boston, MA: RStudio, Inc. Verfügbar unter: <http://www.rstudio.com/>

- Schofield, A. & Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics*, 4, 287–300.
- Schui, G. & Krampen, G. (2016). ZPID-Monitor 2014 zur Internationalität der Psychologie aus dem deutschsprachigen Bereich: Der ausführliche Bericht. *ZPID Science Information Online*, 16 (2). Zugriff am 03.07.2018. Verfügbar unter: https://www.zpid.de/pub/research/zpid-monitor_2014_lang.pdf
- Solga, M. & Blickle, G. (2006). In deutschsprachigen wissenschaftlichen Zeitschriften der Jahre 2004 und 2005 publizierte Forschungsbeiträge zur Arbeits- und Organisationspsychologie: Eine empirische Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie A&O*, 50 (1), 28–42. <https://doi.org/10.1026/0932-4089.50.1.28>
- Tuleya, L. G. (Ed.). (2007). *Thesaurus of psychological index terms* (11th ed.). Washington, DC: American Psychological Association.
- Vulić, I., De Smet, W., Tang, J. & Moens, M. F. (2015). Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. *Information Processing & Management*, 51, 111–147. <https://doi.org/10.1016/j.ipm.2014.08.003>
- Wissenschaftsrat (2018). *Perspektiven der Psychologie in Deutschland* (Drs. 6825-18). Zugriff am 09.04.2018. Verfügbar unter: <https://www.wissenschaftsrat.de/download/archiv/6825-18.pdf>
- ZPID – Leibniz-Zentrum für Psychologische Information und Dokumentation (Hrsg.). (2016). *PSYNDEX Terms* (10. Aufl.). Trier: ZPID. Zugriff am 03.04.2018. Verfügbar unter: <https://www.zpid.de/pub/info/PSYNDEXterms2016.pdf>

Danksagungen

Wir danken Veronika Kuhberg-Lasson für anregende Impulse und Katja Singleton für hilfreiche Hinweise beim Verfassen dieses Beitrags sowie den anonymen Gutachterinnen bzw. Gutachtern für die konstruktiven Rückmeldungen.

Open Data

Zusatzmaterialien zu diesem Artikel werden unter PsychArchives.org zur Verfügung gestellt: <http://dx.doi.org/10.23668/psycharchives.2499> und <http://dx.doi.org/10.23668/psycharchives.2498>

Dipl.-Psych. André Bittermann

Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)
Universitätsring 15
54296 Trier
abi@leibniz-psychology.org

Dr. Eva Maria Klos

Hochschule Trier
EU-Forschungsberatungs- und Koordinierungsstelle
Schneidershof
54293 Trier

Anhang C: Originalpublikation von Studie 3 (accepted version)

Unterscheiden sich die Forschungsinteressen von Frauen und Männern in der Psychologie?

Eine Analyse von PSYNDEX-Einträgen über einen Zeitraum von 50 Jahren

André Bittermann und Nina Greiner

Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), Trier

Andreas Fischer

Forschungsinstitut Betriebliche Bildung (f-bb), Nürnberg

Korrespondenzautor:

André Bittermann, Leibniz-Zentrum für Psychologische Information und Dokumentation
(ZPID), Universitätsring 15, 54296 Trier, E-Mail: abi@leibniz-psychology.org

Danksagungen

Unser besonderer Dank gilt Veronika Kuhberg-Lasson und Ute Wahner für die hilfreichen Diskussionen und Anmerkungen in allen Entstehungsphasen des Manuskripts sowie Katja Trillitzsch für die Unterstützung bei der Aufbereitung der Dissertationsdatensätze.

Zusammenfassung

Auf die Mitte der 1990er Jahre angestoßene Debatte über den steigenden Frauenanteil unter den Psychologiestudierenden folgte eine Reihe von Studien, welche unter anderem geschlechtsspezifische Interessenschwerpunkte untersuchten. Bis heute zeigt sich zu solchen Präferenzen ein insgesamt gemischtes Bild und es liegt noch keine umfassende Untersuchung über einen größeren Zeitraum vor. Daher wurde untersucht, (1) ob sich die Forschungsinteressen von promovierenden Frauen und Männern in der Psychologie unterscheiden und (2) welche zeitlichen Trends hierbei ausgemacht werden können. Analysiert wurden 17 971 in der psychologischen Referenzdatenbank PSYINDEX nachgewiesene Dissertationen aus dem deutschen Sprachraum der Jahre 1968 bis 2017. Die individuellen Forschungsinteressen wurden anhand standardisierter Schlagwörter der Dissertationen mit Latent Dirichlet Allocation (LDA) identifiziert. Zeitliche Trends wurden mit Multilayer Perceptrons (MLP) untersucht. Es konnten 48 Interessenbereiche identifiziert werden, wobei sich nur in zwei Bereichen zeitlich konstant größere Unterschiede zwischen Frauen und Männern zeigten: „Mutter-Kind-Beziehung und Entwicklung im Kleinkindalter“ wurde mit einer höheren Wahrscheinlichkeit von Frauen, „Statistik und Methoden“ stärker von Männern adressiert. Weitere Unterschiede unterlagen zeitlichen Schwankungen. Die Befunde stützen insgesamt die Annahme, dass die Ähnlichkeiten zwischen Frauen und Männern überwiegen.

Schlüsselwörter: Forschungsinteressen, Geschlechtsunterschiede, Szientometrie, Topic Modeling, Multilayer Perceptrons

Abstract

Do research interests differ between women and men in psychology? An analysis of PSYNDEX records over a period of 50 years

The debate initiated in the mid-1990s about the increasing proportion of women among psychology students in Germany was followed by a series of studies that, among other aspects, examined gender-specific interests. To date, the overall picture of such preferences is mixed, and there is no comprehensive study comprising a longer time period. Therefore, the present study examined (1) whether the research interests of female and male doctoral students in psychology differ and (2) which temporal trends can be identified for these interests. Specifically, 17,971 references from the psychological reference database PSYNDEX to dissertations from the German-speaking countries published between 1968 and 2017 were analyzed. The individual research interests were identified with latent Dirichlet allocation (LDA) using standardized keywords of the references. Temporal trends were investigated with multilayer perceptrons (MLP). A total of 48 areas of interest could be identified, where only two areas persistently showed major differences between women and men over time: "mother-child relationship and development in infancy" was addressed with a higher probability by women, "statistics and methods" was more likely to be addressed by men. Other differences were characterized by temporal fluctuations. Overall, the findings support the assumption that similarities between women and men are more predominant than the differences between them.

Keywords: research interest, sex differences, scientometrics, topic modeling, multilayer perceptrons

Unterscheiden sich die Forschungsinteressen von Frauen und Männern in der Psychologie? Eine Analyse von PSYINDEX-Einträgen über einen Zeitraum von 50 Jahren

Die provokanten Spekulationen von Urs Baumann (1995) über die Folgen eines wachsenden Frauenanteils für die akademische Psychologie lösten Mitte der 1990er Jahre eine Debatte über geschlechtsspezifische Interessen aus. Neben der Annahme einer geringeren Bereitschaft von Psychologinnen zu einer wissenschaftlichen Karriere sprach Baumann von einer zu erwartenden „Ausdünnung einzelner Fächer“ und einer „Verlagerung des Forschungsprofils“ (Baumann, 1995, S. 10), die durch eine geschlechtsabhängige Präferenz für psychologische Subdisziplinen bedingt sei. Aus heutiger Sicht kann vor dem Hintergrund einer weiterhin steigenden Promotionsquote von Frauen, die sich immer mehr dem Anteil an Masterabsolventinnen angleicht (Antoni, 2019), der Schlussfolgerung zugestimmt werden, dass von „Wissenschaftsfeindlichkeit“ bei Frauen keine Rede sein kann (Gundlach, Tröster & Moschner, 1999)¹.

Doch die Frage, ob sich die Forschungsinteressen von Psychologinnen und Psychologen tatsächlich unterscheiden, ist auch knapp 25 Jahre nach Baumanns Äußerungen noch nicht abschließend geklärt. Über verschiedene Untersuchungsjahre und -stichproben hinweg stellten sich vor allem zwei Themenfelder als von Frauen und Männern unterschiedlich stark besetzt heraus: Während Frauen eine ausgeprägtere Präferenz für Entwicklungspsychologie zu haben scheinen (Brack, Reinhardt, Dahme & Hoffmann, 1997; Gnambs, Hanfstingl & Leidenfrost, 2006, Hartmann, 2012), wird der Methoden-Bereich eher von Männern adressiert (Brack et al., 1997; Hartmann, 2012; König, Fell, Kellnoher & Schui, 2015). Weiterhin zeigten die Ergebnisse einer Umfrage unter Studierenden, dass Frauen häufiger Themen der Klinischen

¹ Baumann (1995, S. 9) führte seine Vermutung, Frauen hätten eine „vielfach nur eingeschränkte vorhandene Möglichkeit und Bereitschaft zur wissenschaftlichen Karriere“, auf Statistiken zu zunehmend sinkenden Anteilen von Frauen bei Promotionen und Habilitationen zurück. Zu den Zahlenverhältnissen von Frauen zu Männern nannte er bei Promotionen 1 zu 2.2 (heute 2.5 zu 1) und bei Habilitationen 1 zu 5.5 (heute etwa 1 zu 1).

und Pädagogischen Psychologie und seltener Themen aus der Grundlagenforschung und der ABO-Psychologie als Interessenschwerpunkt nannten (Gundlach et al., 1999). Bei einer disziplinübergreifenden Analyse von Worthäufigkeiten in Titeln, Abstracts und Schlagworten wissenschaftlicher Publikationen zeigte sich außerdem, dass in der Psychologie der Begriff „Theorie“ häufiger von Männern verwendet wurde (Thelwall, Bailey, Tobin & Bradshaw, 2019). Die Autoren kommen zu dem Schluss, dass Frauen über alle untersuchten Fächer hinweg Themen mit einer sozialen Wirkrichtung bevorzugen, während die Publikationen von Männern sich häufiger mit physischen Objekten oder abstrakten Entitäten beschäftigen. Damit gehen die Autoren über eine prominente Hypothese aus der Berufswahlforschung hinaus, die besagt, dass sich Frauen und Männer hinsichtlich einer postulierten „people-things dimension“ (Prediger, 1982) unterscheiden: Frauen weisen demnach ein stärkeres Interesse für interpersonelle Tätigkeiten auf, Männer präferieren hingegen objektbezogene Berufe (Lippa, 1998; Su, Rounds & Armstrong, 2009).

Im Hinblick auf die Entstehung solcher geschlechtsspezifischen Unterschiede gibt es eine Vielzahl an theoretischen Ansätzen². Ein aktuelles Modell, das versucht, das komplexe Zusammenspiel von individuellem Interesse und sozio-kulturellem Kontext während der Ausbildung und der beruflichen Entwicklung abzubilden, findet sich bei Su und Kollegen (Su, in Druck; Su, Stoll & Rounds, 2019): In diesem integrativen Ansatz werden drei Ebenen von Interesse unterschieden: (1) Die affektiven Reaktionen auf ein Objekt/eine Tätigkeit, (2) die kognitive Bewertung in Bezug auf Nutzen und Valenz und (3) die kognitive Bewertung in Bezug auf die Kompatibilität mit dem eigenen Selbstkonzept (Identifikation). Auf der dritten Ebene fließt der individuelle Sozialisationsprozess mit ein, in dem die Interessen einer Person durch soziale Rollenerwartungen und kulturelle Werte überformt werden, „[...] e.g., a young woman who is good at math and is fascinated by robotics may hesitate to express interest in it

² Für einen Überblick siehe z. B. Su et al. (2019).

due to its perceive[d] incompatibility with her gender identity [...].” (Su, in Druck, S. 31). Vor dem Hintergrund dieses Modells wäre zu erwarten, dass sich Frauen und Männer im Rahmen des beschriebenen Spannungsfeldes aus individuellem Interesse und gesellschaftlichen Rollenbildern auf unterschiedliche Forschungsgebiete verteilen, was sich in den Inhalten von Abschlussarbeiten und wissenschaftlichen Publikationen widerspiegeln sollte. Die vorliegende Untersuchung soll daher der Frage nachgehen, ob sich bei Psychologinnen und Psychologen Hinweise auf solche geschlechtsspezifischen Forschungsinteressen finden lassen.

Aus methodischer Sicht kann bemängelt werden, dass die Daten bisheriger Untersuchungen zu geschlechtsspezifischen Forschungsinteressen aus relativ kurzen Zeiträumen mit teilweise länderspezifischen und heterogenen Stichproben stammen. Eine umfassende Untersuchung mit einer Vollerhebung von psychologischen Dissertationen aus dem deutschsprachigen Raum ermöglicht eine genauere Betrachtung zeitlicher Entwicklungen. Denkbar wäre, dass sich parallel zu einem gesamtgesellschaftlichen Trend hin zu einer Annäherung der Geschlechterrollen (Lang & Risman, 2006) und der steigenden Promotionsquote von Psychologinnen (Antoni, 2019) Angleichungen in den Forschungsinteressen von Frauen und Männern erkennen lassen. Tatsächlich konnten König und Kollegen (2015) in den Publikationsthemen von Nachwuchspsychologinnen und -psychologen keine signifikanten Geschlechterunterschiede nachweisen, was als Hinweis auf einen Konvergenztrend interpretiert werden kann. In bisherigen Studien aus dem deutschsprachigen Raum wurden darüber hinaus die Interessen mit eher groben Kategorien (wie „Allgemeine Psychologie“, „Klinische Psychologie“ usw.) erfasst. Die Verwendung von vorab definierten Klassifikationen zur Erfassung von Forschungsthemen und -interessen wurde jedoch vor allem hinsichtlich ihres Detailgrades kritisiert (Bittermann & Fischer, 2018; Ding, 2011). Demnach wäre es möglich, dass sich Unterschiede vor allem in Teilaspekten zeigen, die sich nicht auf gesamte Disziplinen der Psychologie generalisieren lassen und daher in Kategorien

zusammengefasst zu eher geringen oder auch ausbleibenden Effekten führen können. Deshalb erscheint es sinnvoll, die Forschungsinteressen datengeleitet aus den Inhalten von wissenschaftlichen Arbeiten abzuleiten, statt a priori definierte, weit gefasste Klassifikationen zu verwenden. Die Wahl von Dissertationen als Datengrundlage bietet zudem den Vorteil, dass deren Inhalt klar der wissenschaftlichen Arbeit einer Person zugeordnet werden kann.

Fragestellungen

In dieser Studie wird daher untersucht,

(1) ob sich die Forschungsinteressen von promovierenden Frauen und Männern in der Psychologie unterscheiden und

(2) welche zeitlichen Trends von Interessen/-unterschieden ausgemacht werden können.

Dazu werden die Inhalte von Dissertationen aus einem Zeitraum von 50 Jahren extrahiert und hinsichtlich des Geschlechts der Autorin bzw. des Autors verglichen. Angesichts des deskriptiven Charakters dieser Studie sollen die Ergebnisse der Analyse zwar vor dem oben dargelegten theoretischen Hintergrund diskutiert werden, eine Prüfung der angeführten Modelle ist jedoch auf Grund dieser Zielsetzung nicht möglich. Darüber hinaus soll mit einer vollständigen Bestandsaufnahme beurteilt werden können, ob bisherige Ergebnisse an Ort und Zeit gebunden waren oder als repräsentativ angesehen werden können. Außerdem adressieren wir damit die auch heute noch aktuelle Frage, ob sich der gesamtgesellschaftliche Trend hin zu einer Annäherung der Geschlechterrollen (Lang & Risman, 2006) in Verschiebungen der Forschungsinteressen von Psychologinnen und Psychologen widerspiegelt.

Methode

Das methodische Vorgehen wird im Folgenden skizziert, ausführliche Beschreibungen einschließlich der verwendeten Software können in [PsychArchives-ESM 1](#)³ sowie [PsychArchives-ESM 2](#) eingesehen werden.

Daten

Die Datengrundlage bildeten $N = 18\,050$ in der psychologischen Referenzdatenbank PSYINDEX⁴ nachgewiesene Dissertationen aus dem deutschen Sprachraum und den Jahren 1968 bis 2017. Das Geschlecht der verfassenden Person wurde anhand des Vornamens bestimmt⁵. Nach Ausschluss nicht zuordenbarer Fälle bestand der finale Datensatz aus $n = 17\,971$ Dissertationsnachweisen⁶.

Forschungsinteressen

Das Ableiten von Forschungsinteressen aus den Themen wissenschaftlicher Arbeiten entspricht einem inzwischen etablierten Ansatz in der Szientometrie, vor allem seit dem Author-Topic Model von Rosen-Zvi, Griffiths, Steyvers und Smyth (2004). Wir gehen dementsprechend davon aus, dass sich das individuelle Interesse einer Person in der Wahl eines Forschungsgegenstandes niederschlägt, auch wenn neben dem Interesse vermutlich andere Aspekte wie strukturelle Faktoren oder pragmatische und karrieregeleitete Überlegungen hier ebenso eine Rolle spielen. Unser Fokus liegt vor allem auf horizontalen Segregationsmustern, also der Verteilung von Frauen und Männern auf verschiedene Themengebiete und weniger auf vertikalen Segregationsprozessen, die sich auf eine

³ Die elektronischen Supplementmaterialien (PsychArchives-ESM) können auf PsychArchives.org unter <https://dx.doi.org/10.23668/psycharchives.2678> und <https://dx.doi.org/10.23668/psycharchives.2677> abgerufen werden.

⁴ produziert vom Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID); freier Zugriff über www.PubPsych.de

⁵ Wir sind uns der Tatsache bewusst, dass Vornamen nicht zwingend Rückschlüsse auf das Geschlecht einer Person zulassen. In Anbetracht des Mangels von Selbstauskünften aller Personen stellte dies jedoch für diese Studie die beste Möglichkeit dar.

⁶ Da es sich bei unseren Daten nahezu um eine Vollerhebung handelt, verzichten wir im Folgenden auf Inferenzstatistiken.

heterogene Positionierung der Geschlechter im Rahmen einer hierarchischen Arbeitsmarkt- und Wissenschaftsstruktur beziehen.

Der Inhalt der Dissertationen wurde wie bei Bittermann und Fischer (2018) anhand ihrer standardisierten Schlagworte⁷ in PSYNDEX (ZPID, 2016) erfasst. Um die den Dissertationen zugrundeliegenden Themen zu bestimmen, wurde Topic Modeling basierend auf Latent Dirichlet Allocation (LDA; Blei, Ng & Jordan, 2003) unter Einbezug der Best-Practice-Empfehlungen von Maier et al. (2018) angewandt. Resultat des Topic Modelings sind Cluster häufig miteinander vorkommender Wörter (= Themen) sowie Dokument-Thema-Wahrscheinlichkeiten, anhand deren die Prävalenz eines Themas bestimmt wird. Die Prävalenz beschreibt die durchschnittliche Wahrscheinlichkeit eines Themas im Korpus. Ihre Höhe hängt von der Anzahl der Themen im Modell ab⁸. Prävalenzdifferenzen größer als die halbe durchschnittliche Prävalenz fassen wir als größere Unterschiede auf.

Zeitliche Trends

Um die Zeitverläufe der Themen von Schwankungen zu befreien, wurden nonlineare Trendkurven getrennt nach Geschlecht mithilfe von Multilayer Perceptrons (MLP) berechnet. Wie bereits bei Bittermann und Fischer (2018) wurde für jedes MLP eine Struktur mit zwei Hidden Units berechnet, um die durchschnittliche Themenwahrscheinlichkeit als nichtlineare Funktion des Erscheinungsjahres zu modellieren. Die zeitlichen Trends wurden schließlich auf drei Dimensionen untersucht:

(1) Dimension „Prävalenzdifferenz“, das heißt, die mittlere Differenz zwischen der weiblichen und männlichen Trendkurve,

⁷ Diese kontrollierten Schlagworte werden vom wissenschaftlichen Personal der Informations- und Recherchedienste am ZPID vergeben und bilden die zentralen Inhalte der Publikationen mit einem standardisierten Vokabular ab (ZPID, 2016). Für weitere Gründe, in dieser Arbeit diese standardisierten Schlagwörter und keine Abstracts zu verwenden, siehe [PsychArchives-ESM 1](#).

⁸ So beträgt bei vier Themen die durchschnittliche Prävalenz 25 % (wenn jedes Dokument jedes Thema gleich stark behandelt). Bei einem Modell mit 100 Themen beträgt die durchschnittliche Prävalenz nur noch 1 %.

(2) Dimension „Konvergenz“ im Sinne einer über die Zeit zunehmenden Nivellierung zweier Trends, sowie

(3) Dimension „Gesamttrend“ unabhängig vom Geschlecht.

Ergebnisse

Frauen- und Männeranteile im Datensatz

Von den 17 971 in die Analysen eingeschlossenen Dissertationen wurden 9 608 (53.5 %) von Frauen und 8 363 (46.5 %) von Männern verfasst. Der Frauenanteil stieg dabei kontinuierlich⁹ von 27.6 % im Jahr 1968 auf 64.2 % im Jahr 2017 und näherte sich damit zunehmend dem relativ konstanten Anteil weiblicher Erstsemester in Deutschland von ca. 79 % an (Destatis, 2019; vgl. [PsychArchives-ESM 3](#)).

((Abbildung 1 ungefähr hier einfügen))

Interessenunterschiede und Thementrends

Es konnten 48 Themen identifiziert werden. Abbildung 1 gibt einen Überblick über die Position der Themen auf den Dimensionen Prävalenzdifferenz, Konvergenz sowie Gesamttrend. Der Großteil der Themen konzentrierte sich im Bereich geringer Unterschiede (Gini-Koeffizient $G = .50$): 38 von 48 Themen (79.2 %) zeigten eine absolute Prävalenzdifferenz $|D| \leq 0.5$ %.

Im Bereich größerer Unterschiede ($|D| > 0.5$ %) lagen mit einer höheren Prävalenz bei Frauen die Themen 27 („human females, sex roles, human males, sex role attitudes, human sex differences“; $|D| = 0.99$ %), 83 („parent-child relations, family relations, childrearing practices, mothers, childhood development“; $|D| = 0.91$ %), 78 („mother-child relations, attachment behavior, infant development, mothers, mother-child communication“; $|D| = 0.72$ %) und 70 („obesity, eating disorders, bulimia, human females, anorexia nervosa“; $|D| = 0.51$ %). Nur bei Thema 78 war der Trend über alle Jahre konstant (horizontale Linie in Abbildung

⁹ $b_1 = 0.828$; $R^2_{adj} = .931$

1). Mit einer höheren Prävalenz bei Männern zeigten sich größere Differenzen bei den Themen 95 („theories, philosophies, history of psychology, theory formulation, psychology“; $|D| = 1.06\%$), 24 („methodology, statistical analysis, measurement, models, mathematical modeling“; $|D| = 1.01\%$), 9 („cognitive processes, problem solving, models, theory formulation, knowledge level“; $|D| = 0.76\%$), 72 („driving behavior, drivers, highway safety, automobiles, motor traffic accidents“; $|D| = 0.69\%$), 94 („decision making, cognitive processes, judgment, choice behavior, uncertainty“; $|D| = 0.65\%$) und 76 („test validity, intelligence measures, test construction, measurement, testing“; $|D| = 0.61\%$). Wie auf Seiten der Frauen konnte nur für eines dieser Themen ein konstanter Trend identifiziert werden (Thema 24). Eine vollständige Tabelle mit allen Themen findet sich in [PsychArchives-ESM 4](#).

Hinsichtlich zeitlicher Veränderungen konnte ein Zusammenhang zwischen dem Gesamttrend und der Konvergenz der Trends von Frauen und Männern festgestellt werden ($r = .73$): Allgemein steigende Trends neigten zur Divergenz, fallende Trends zur Konvergenz. Im Sinne eines Bodeneffektes ist eine Konvergenz bei sinkenden Trends auf die generell schiefe Verteilung von Themenprävalenzen zurückzuführen: Ein Thema wird nur von einem Teil aller Dokumente adressiert, beim Rest hat das Thema entsprechend eine Prävalenz von Null. Abbildung 1 zeigt aber auch Ausnahmen. Die Abbildungen zu den zeitlichen Verläufen aller Themen finden sich in [PsychArchives-ESM 5](#).

Diskussion

In der vorliegenden Studie wurde untersucht, ob sich die Forschungsinteressen von promovierenden Psychologinnen und Psychologen unterscheiden bzw. welche zeitlichen Muster festgestellt werden können. Zunächst ist festzuhalten, dass keine ausschließlichen „Frauen-“ bzw. „Männerthemen“ gefunden wurden. Wie Abbildung 1 zeigt, konzentrieren sich die Themen im Bereich geringer Unterschiede. Dieses Ergebnis stützt die Schlussfolgerung von Hyde (2005), dass die Ähnlichkeiten zwischen Frauen und Männern

überwiegen. Dennoch kann konstatiert werden, dass trotz einer großen Überschneidung der Forschungsinteressen von Frauen und Männern gewisse Differenzen verbleiben. Ein solcher Unterschied zeigte sich in unseren Daten im Trend zeitlich stabil und über verschiedene Modellvarianten hinweg beim Thema „Mutter-Kind-Beziehung und Entwicklung im Kleinkindalter“ (Thema 78), welches stärker von Frauen adressiert wurde. Frauen beschäftigten sich in ihren Promotionen auch eher als Männer mit Geschlechterrollen und -unterschieden (Thema 27), Kindererziehung und Entwicklung in der Kindheit (Thema 83), und Essstörungen (Thema 70), wobei für diese Themengebiete über den zeitlichen Verlauf hinweg Schwankungen vorlagen. Insgesamt decken sich diese Ergebnisse mit den Befunden, nach denen sich das Interesse von Frauen häufig auf den Menschen fokussiert bzw. eine prosoziale Dimension aufweist (Brack et al., 1997; Gundlach et al., 1999; Hartmann, 2012; Su et al., 2009; Thelwall et al., 2019).

„Theorien in der Psychologie, Geschichte der Psychologie und philosophische Aspekte“ (Thema 95) erwies sich als das am stärksten von Männern besetzte Gebiet, was mit der Annahme in Verbindung gebracht werden kann, dass Männer die Auseinandersetzung mit abstrakteren Themen suchen (Thelwall et al., 2019). Allerdings muss betont werden, dass die Prävalenz dieses Themas in den letzten zehn Jahren für beide Geschlechter stark abnahm (s. [PsychArchives-ESM 5](#)), so dass aktuell kein bedeutsamer Unterschied mehr feststellbar ist. Gleiches gilt für Thema 9, das kognitive Prozesse und Problemlösen beinhaltet.

Darüber hinaus fand sich in den Promotionen von Männern in Einklang mit vorhergehenden Untersuchungen (Brack et al., 1997; Hartmann, 2012; König et al., 2015) ein zeitlich stabiler Trend im Hinblick auf das Feld Methoden und Statistik (Thema 24). Dass sich dies Jahrzehnte nach der Diskussion um Baumann (1995) noch immer zeigt, ist bemerkenswert und gibt Anlass für zukünftige Untersuchungen. So stellt sich die Frage, ob etwa eine bei Frauen stärker ausgeprägte „Statistikangst“ (Hartmann, 2012) hier eine Rolle spielt oder ob – aus berufswahltheoretischer Sicht – Männer innerhalb eines von Frauen dominierten Faches wie

der Psychologie zu einer geschlechtsstereotypen Spezialisierung neigen, die einer immer noch eher männlichen Zuschreibung dieser Felder auf gesellschaftlicher Ebene entspricht: Eine männliche Präferenz ergab sich außerdem für „Fahrverhalten und Fahrzeuge“ (Thema 72) und tendenziell auch für „Sport“ (Thema 52) – zwei ebenfalls geschlechtsstereotype Themen (vgl. Abbildung 1 und [PsychArchives-ESM 4](#)). In Bezug auf Psychologinnen kann zudem vermutet werden, dass Geschlechterstereotype in zweifacher Weise dazu beitragen, dass sie mit einer geringeren Wahrscheinlichkeit im Bereich Methoden und Statistik promovieren. Zum einen führt die Konfrontation mit Stereotypen zu unterschiedlichen Ausprägungen des Fähigkeitsselbstkonzeptes von Männern und Frauen. Frauen entscheiden sich daher im Sinne einer Selbstselektion möglicherweise weniger häufig für dieses Gebiet. Zum anderen wäre es auch denkbar, dass Geschlechterstereotype auf Seiten von Professorinnen und Professoren dazu beitragen, dass Frauen seltener Promotionsstellen an Statistik-/Methoden-Lehrstühlen erhalten und somit strukturelle Faktoren ebenfalls eine Rolle spielen. Natürlich kann eine analoge Argumentation für Männer bei Themenbereichen der Entwicklungspsychologie vorgenommen werden. Zum Einfluss von Stereotypen im akademischen Bereich finden sich weitere Ausführungen in einem Themenheft zu Geschlechterverhältnissen in der Wissenschaft (IHF, 2015) sowie bei Zimmermann und Matthies (2010). Für die zukünftige Forschung wäre es wünschenswert, den Einfluss von bestehenden Geschlechterstereotypen in der Wissenschaft und auf allgemein gesellschaftlicher Ebene weiter zu untersuchen und aufzuweichen – ein Projekt, zu dem insbesondere Psychologinnen bereits einen Beitrag geleistet haben (vgl. Thema 27).

Mithilfe des Topic-Modeling-Ansatzes konnten die Interessen detaillierter erfasst werden als mit vorab definierten Kategorien. Grundsätzlich reihen sich die Ergebnisse in die bisherige Forschung ein, vor allem, was die stärkere männliche Präferenz für Methoden (Thema 24) betrifft. Die durch frühere Untersuchungen gestützte Annahme, dass Frauen ein größeres Interesse an Entwicklungspsychologie aufweisen, muss nach unseren Ergebnissen

differenziert werden. Zwar liegen zwei der betreffenden Themen im Bereich größerer Unterschiede (Themen 83 und 78 zu Entwicklung der (frühen) Kindheit), das dritte Thema zur kognitiven Entwicklung (Thema 47) war aber für beide Geschlechter gleich wahrscheinlich.

Analog zum Grundgedanken von Rosen-Zvi et al. (2004) haben wir die Themen der Dissertationen als Indikator für das Forschungsinteresse der Autorin bzw. des Autors herangezogen. Ob das Promotionsthema nun tatsächlich von höchstem persönlichen Interesse ist, oder mit der Wahl eines Themas eventuelle Karrierevorteile oder pragmatische Gründe (wie beispielsweise eine finanzielle Absicherung durch eine damit einhergehende Stelle am Lehrstuhl) verbunden sind¹⁰, kann auf Grundlage unserer Daten allein nicht beantwortet werden. Selbstauskünfte einer repräsentativen Stichprobe von forschungstätigen Psychologinnen und Psychologen oder eine Analyse von strukturellen Faktoren¹¹ wären eine aufschlussreiche Ergänzung, um in zukünftigen Studien den Anteil weiterer Einflussfaktoren auf die Wahl des Promotionsthemas zu ermitteln. Überdies sollten weiterführende Untersuchungen sich damit beschäftigen, ob die gefundenen geschlechtsspezifischen Themenpräferenzen im weiteren akademischen Verlauf nach der Promotion – der in der Psychologie mit einer Abnahme des Frauenanteils verbunden ist (Antoni, 2019) – erhalten bleiben.

Fazit

Zusammenfassend lässt sich feststellen, dass nur einzelne Bereiche mit größeren Interessenunterschieden ermittelt wurden, die bis auf zwei Ausnahmen zudem zeitlichen

¹⁰ Eine weiterführende Diskussion zum Verhältnis von Interessen und ausbildungs- bzw. berufsbezogenen Entscheidungen findet sich bei Su (in Druck).

¹¹ Wie eine anonyme Gutachterin bzw. ein anonymes Gutachter anmerkt, können neben dem Interesse auch Gelegenheitsstrukturen die Wahl des Promotionsthemas beeinflussen, unter anderem die Entwicklung von Lehrstuhldenominationen, die Ausrichtung der Lehrstühle (so z. B. die abnehmende Zahl von Lehrstühlen zur Geschichte der Psychologie), die Bedeutung von Graduiertenschulen oder die Drittmittelstärke eines Faches und damit das Angebot an Promotionsstellen. Zu prüfen wäre hierbei, inwiefern diese Faktoren beide Geschlechter gleichermaßen oder unterschiedlich betreffen.

Schwankungen unterlagen. Der neue methodische Ansatz und der umfassende Datensatz führen zu differenzierteren Ergebnissen als die bisherige Forschung. Die Befunde stützen die Annahme, dass die Ähnlichkeiten zwischen Männern und Frauen überwiegen. Wir schließen uns somit der Schlussfolgerung von Hyde an, dass es keine empirische Grundlage für „[...] overinflated claims of gender differences [...]“ (Hyde, 2005, S. 586) gibt. Keinesfalls sollten also aus den vorliegenden Befunden Rückschlüsse auf genuin weibliche bzw. männliche Interessen, losgelöst von Sozialisationsprozessen, Rollenerwartungen und anderen Einflussfaktoren, gezogen werden.

Elektronische Supplemente (ESM)

Zusatzmaterialien zu diesem Artikel werden unter PsychArchives.org zur Verfügung gestellt:

<http://dx.doi.org/10.23668/psycharchives.2678> und

<http://dx.doi.org/10.23668/psycharchives.2677>.

[PsychArchives-ESM 1](#) (Forschungsinteressen_ESM 1.pdf):

Detaillierte Beschreibung des methodischen Vorgehens.

[PsychArchives-ESM 2](#) (LDA.R, preprocessing.R, reliability.R, topic_coherence.R):

R-Code der Analysen.

[PsychArchives-ESM 3](#) (Forschungsinteressen_ESM 3.pdf):

Abbildung zu Frauenanteilen an den Promotionen.

[PsychArchives-ESM 4](#) (Forschungsinteressen_ESM 4.pdf):

Auflistung aller Themen.

[PsychArchives-ESM 5](#) (Forschungsinteressen_ESM 5.pdf):

Zeitliche Verläufe aller Themen.

Literaturverzeichnis

Antoni, C. H. (2019). Zur Lage der Psychologie. *Psychologische Rundschau*, 70, 4–26.

<https://doi.org/10.1026/0033-3042/a000429>

Baumann, U. (1995). Bericht zur Lage der deutschsprachigen Psychologie 1994. Fakten und Perspektiven. *Psychologische Rundschau*, 46 (1), 3–17

Bittermann, A. & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, 226 (1), 3–13. <https://doi.org/10.1027/2151-2604/a000318>

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>

Brack, K., Reinhardt, S., Dahme, B. & Hoffmann, B. (1997). Gibt es geschlechtsspezifische Präferenzen in der psychologischen Forschung? – Eine Untersuchung psychologischer Doktorarbeiten in der alten Bundesrepublik der Jahrgänge 1986–1990. *Psychologische Rundschau*, 48 (2), 101–105.

Destatis. (2019). Statistisches Bundesamt. *Studienanfänger: Deutschland, Semester, Nationalität, Geschlecht, Studienfach*. Abgerufen von <https://www-genesis.destatis.de>

Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics*, 5 (1), 187–203.

<https://doi.org/10.1016/j.joi.2010.10.008>

Gnambs, T., Hanfstingl, B. & Leidenfrost, B. (2006). Ist Methodenlehre männlich und Entwicklungspsychologie weiblich? Geschlechtsspezifische Präferenzen bei Abschlussarbeiten in der Psychologie. In B. Gula, R. Alexandrowicz, S. Strauß, E. Brunner, B. Jenull-Schiefer & O. Vitouch (Hrsg.), *Perspektiven psychologischer Forschung in Österreich. Proceedings zur 7. Wissenschaftlichen Tagung der Österreichischen Gesellschaft für Psychologie* (S. 374–381). Pabst: Lengerich.

Gundlach, G., Tröster, H. & Moschner, B. (1999). Sind Psychologiestudentinnen wissenschaftsfeindlich? *Psychologie in Erziehung und Unterricht*, 46 (1), 29–40.

Hartmann, B. (2012). *Sind Psychologiestudentinnen wissenschaftsfeindlich?*

Geschlechtsunterschiede im Bezug auf studienrelevante und berufsrelevante Faktoren

(Diplomarbeit). Verfügbar unter: [http://othes.univie.ac.at/21135/1/2012-02-](http://othes.univie.ac.at/21135/1/2012-02-03_9902182.pdf)

[03_9902182.pdf](http://othes.univie.ac.at/21135/1/2012-02-03_9902182.pdf)

Hyde, J. S. (2005). The gender similarities hypothesis. *American psychologist*, 60 (6), 581–

592. <http://dx.doi.org/10.1037/0003-066X.60.6.581>

IHF. (2015). *Beiträge zur Hochschulforschung*, 37 (3). Verfügbar unter

http://www.bzh.bayern.de/uploads/media/3-2015-gesamt_b.pdf

König, C. J., Fell, C. B., Kellnhofer, L. & Schui, G. (2015). Are there gender differences among researchers from industrial/organizational psychology? *Scientometrics*, 105 (3),

1931–1952. <https://doi.org/10.1007/s11192-015-1646-y>

Lang, M. M., & Risman, B. J. (2006). Blending into equality: Family diversity and gender convergence. In M. Davies, M. Evans & J. Lorber (Hrsg.), *Handbook of gender and women's studies* (S. 287–303). Thousand Oaks, CA: Sage.

Lippa, R. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people–things dimension. *Journal of Personality and Social Psychology*, 74 (4), 996–1009. <http://dx.doi.org/10.1037/0022-3514.74.4.996>

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri H. & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12 (2–3), 93–118.

<https://doi.org/10.1080/19312458.2018.1430754>

Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *Journal of Vocational Behavior*, 21 (3), 259–287.

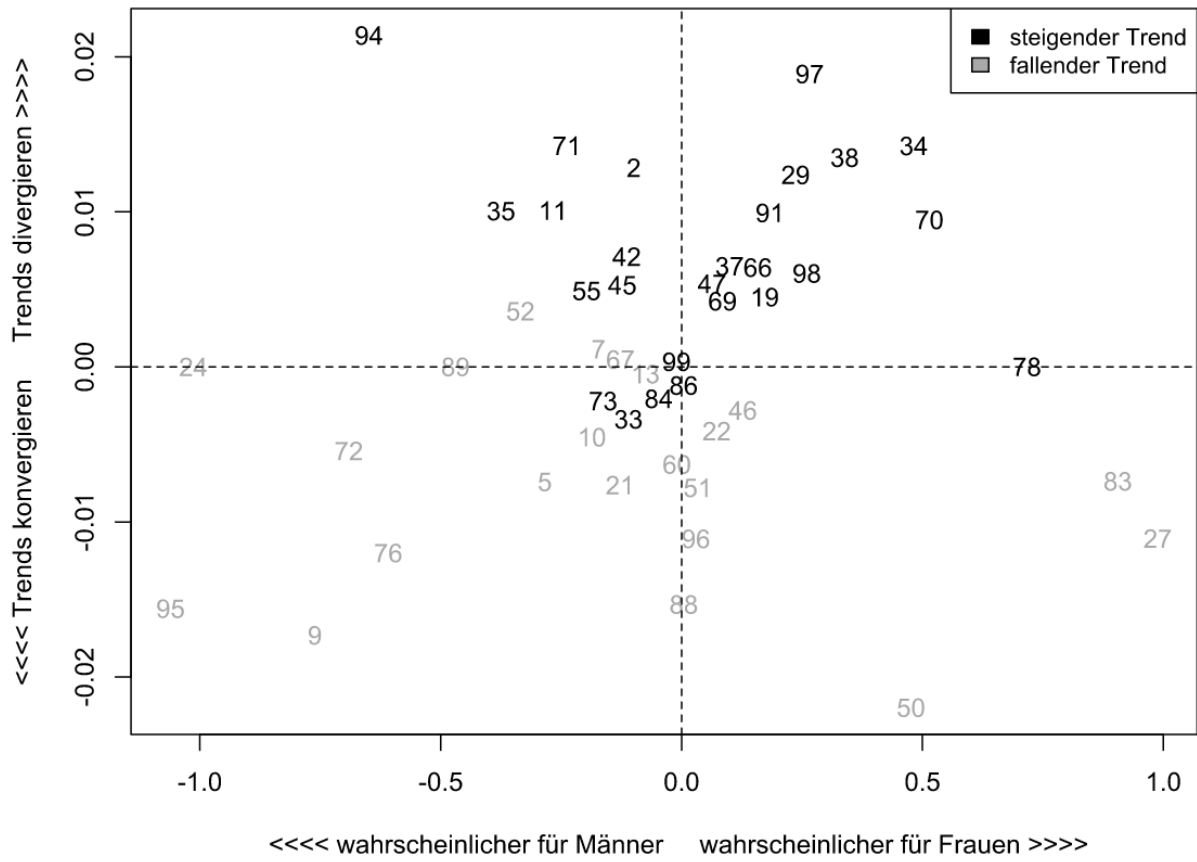
[http://dx.doi.org/10.1016/0001-8791\(82\)90036-7](http://dx.doi.org/10.1016/0001-8791(82)90036-7)

- Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. In M. Chickering & J. Halpern (Hrsg.), *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* (S. 487–494). Arlington, VA: AUAI Press. Verfügbar unter:
<https://mimno.infosci.cornell.edu/info6150/readings/398.pdf>
- Su, R. (in Druck). The three faces of interests: An integrative review of interest research in vocational, organizational, and educational psychology. *Journal of Vocational Behavior*. <https://doi.org/10.1016/j.jvb.2018.10.016>
- Su, R., Rounds, J. & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135 (6), 859–884.
<http://dx.doi.org/10.1037/a0017364>
- Su, R., Stoll, G. & Rounds, J. (2019). The nature of interests: Toward a unifying theory of Trait-State Interest Dynamics. In C. D. Nye & J. Rounds (Hrsg.), *Vocational interests: Rethinking their role in understanding workplace behavior and practice* (S. 11–38). SIOP Organizational Frontiers Series. New York, NY: Routledge
- Thelwall, M., Bailey, C., Tobin, C. & Bradshaw, N. A. (2019). Gender differences in research areas, methods and topics: Can people and thing orientations explain the results? *Journal of Informetrics*, 13 (1), 149–169. <https://doi.org/10.1016/j.joi.2018.12.002>
- Zimmermann, K. & Matthies, H. (2010). Arbeitsfeld Wissenschaft und Geschlechtersegregation. *WSI Mitteilungen*, 63 (5), 264–270.
<https://doi.org/10.5771/0342-300X-2010-5-264>
- ZPID. (2016). *PSYNDEX Terms* (10. Aufl.). Trier: ZPID. Verfügbar unter:
<https://www.psyndex.de/pub/info/PSYNDEXterms2016.pdf>

Titel zu der Abbildung

Abbildung 1. Positionen der Themen auf den Dimensionen „Prävalenzdifferenz“ und „Konvergenz“. Die x-Achse zeigt die über alle Jahre gemittelte Prävalenzdifferenz der Trends in Prozent. Negative Werte zeigen eine höhere Wahrscheinlichkeit des Themas an, von Männern adressiert zu werden, und vice versa. Die y-Achse zeigt die Koeffizienten des linearen Regressionsmodells mit Prädiktor „Publikationsjahr“ und Kriterium „Differenz der Trends“ in Prozent. Wenn der Unterschied zwischen Frauen und Männern über alle Jahre geringer wird, konvergieren die Trends. Schwarz gedruckte Themen haben einen über alle Jahre hinweg steigenden, grau gedruckte einen fallenden Gesamttrend. Ausgewählte Themen werden im Text beschrieben; eine vollständige Tabelle mit allen Themen findet sich in [PsychArchives-ESM 4](#).

Prävalenzdifferenz und Konvergenz der Thementrends



Anhang D: Supplementmaterialien zu Studie 3

In Anhang D sind enthalten:

Ergebnisse der Analyse mit Abstracts.

Tabelle D1: Themen bei Verwendung von Abstracts als Textquelle.

Tabelle D2: Auflistung aller Themen mit der Anzahl erfolgreicher Reproduktionen.

PsychArchives-ESM 1: Detaillierte Beschreibung des methodischen Vorgehens.

PsychArchives-ESM 3: Abbildung zu Frauenanteilen an den Promotionen.

PsychArchives-ESM 4: Auflistung aller Themen.

PsychArchives-ESM 5: Zeitliche Verläufe aller Themen.

Ergebnisse der Analyse mit Abstracts

Zum Vergleich mit den Ergebnissen in Studie 3, die auf den standardisierten Schlagwörtern beruhen (im Folgenden: CT, *Controlled Terms*), wurde auch ein Topic Model basierend auf Abstracts gerechnet. Bezüglich LDA-Hyperparameter, Anzahl der Themen und Ausschluss unreliabler Themen wurde wie bei dem Modell im Manuskript vorgegangen. Die Themen sind in Tabelle D1 aufgelistet. Wie aus dieser Tabelle ersichtlich, lassen sich zu den Abstract-Themen Äquivalente in Studie 3 (bzw. in PsychArchives-ESM 4) ausmachen. Doch selbst nach „Post-Stemming“, also dem Zusammenfassen von Wörter gleichen Wortstamms nach der Modellinferenz (Schofield & Mimno, 2016) sind Abstract-Themen schwieriger lesbar als CT-Themen. Außerdem sind – trotz Verwendung der Stopword-Liste für deutschsprachige Abstracts aus Studie 2 – einige Themen enthalten, die keine Inhalte abbilden, sondern beispielsweise häufig miteinander vorkommende Begriffe der Durchführung experimenteller Studien.

Am deutlichsten zeigt sich der Nachteil von Abstract-Wörtern am Fehlen des Methodenthemas bei einer Analyse von Abstracts. Zwar war ein halbwegs entsprechendes Thema im Modell enthalten, doch musste es als zufällig angenommen werden: Es konnte bei wiederholten Modellinferenzen nicht reproduziert werden, war also sehr instabil und daher nicht reliabel. Daher fehlt dieses Thema in Tabelle D1. Das Methodenthema mit CT hingegen war zu 100 % reliabel, konnte also nach jeder Modellinferenz bestätigt werden. Nach eingehender Untersuchung dieses Unterschieds (u. a. auch ein Abstract-Modell mit 150 Themen, das Ergänzen weiterer Modellinferenzen für die Reliabilitätsprüfung sowie eine Analyse der häufigsten Begriffe in Abstracts methodischer Arbeiten) wird folgende Schlussfolgerung gezogen: Häufige Abstract-Begriffe in methodischen Arbeiten (z. B. Verfahren, Modelle, Methoden, Variablen, Schätzung) kristallisieren sich deswegen nicht als eigenes Thema heraus, da sie oft auch im Abstract empirischer Arbeiten verwendet werden und dabei mit anderen Wörtern erscheinen, die einen anderen inhaltlichen Bezug aufweisen. Da Topic Models auf Wort-Kookkurenzen basieren, „vermischen“ sich die Methoden-Begriffe mit anderen Wörtern in den Themen. Hinzu kommt, dass anhand von Abstract-Begriffen im Topic Model nicht unterscheidbar ist, ob die Methode selbst Gegenstand der Arbeit war, oder ob sie für eine inhaltliche Fragestellung eingesetzt wurde. Selbst wenn die Methode an sich im Fokus steht, müsste mehrmals die gleiche Methode mit dem gleichen Wort in Abstracts bezeichnet worden sein, damit der Begriff eine Chance hat, einen oberen Rang in einem Topic-Model-Thema einzunehmen. CT wie „Statistical Analysis“ wiederum

fassen alle verschiedene Verfahren in einem Begriff zusammen, und auch nur dann, wenn die Methode selbst im Fokus steht und nicht deren bloße Anwendung. Die Qualität der PSYINDEX-Datenbank und der sorgfältig vergebenen CT erweist sich hier als Vorteil.

Tabelle D1. Themen bei Verwendung von Abstracts als Textquelle.

Tabelle D1. Themen bei Verwendung von Abstracts als Textquelle. Nur reliable Themen wurden eingeschlossen. Die Äquivalente zu den Themen mit den größten Geschlechtsunterschieden laut Abschnitt 6.4 sind fett gedruckt.

Thema (Abstracts)	Äquivalentes Thema im Manuskript	Top 10 Terms (nach post-stemming ⁵⁴)
2	45	lernen, feedback, lernerfolg, studierende, lernleistung, einfluss, vorwissen, lernprozess, motivation, lernprozesse
4	n.a. ⁵⁵	repräsentation, mental, kognitiv, räumlich, modelle, prozesse, informationen, struktur, probanden, vorwissen
5	71	mitarbeiter, unternehmen, führung, führungskräfte, teams, commitment, organisation, vertrauen, vorgesetzte, einfluss
6	35	kognitiv, fähigkeiten, leistungsfähigkeit, leistung, prozesse, komplexität, kontrolle, flexibilität, kognition, funktionen
9	n.a.	einstellungen, fragebogen, studenten, kontakt, sozial, verhalten, beziehung, positiv, faktoren, ausmaß
10	47	alter, ältere, erwachsene, jüngere, junge, altersgruppen, erwachsenenalter, höheren, alters, mittleren
13	n.a.	experiment, versuchspersonen, einfluss, versus, effekt, studierende, teilnehmer, experimentell, manipulation, jeweils
17	n.a.	kontrollgruppe, training, versuchsgruppe, experimentalgruppe, gruppe, teilnehmer, verbesserung, versuchspersonen, erhielt, wochen
19	96	patienten, behandlung, therapie, stationär, wirksamkeit, verbesserung, reduktion, symptomatik, veränderungen, verlauf
20	78	mütter, kinder, entwicklung, eltern, väter, familien, kindlich, verhalten, alter, geburt
22	11	versuchspersonen, eeg, reize, dauer, darbietung, sekunden, musik, minuten, registriert, gemessen
23	33	items, skalen, validität, erfassung, fragebogen, reliabilität, entwicklung, überprüfung, validierung, faktoren
24	88	mädchen, jungen, alter, kinder, geschlecht, selbstkonzept, sozial, jugendliche, einfluss, entwicklung
26	88	erfolg, leistungsmotivation, leistung, misserfolg, kontrolle, motivation, anstrengung, fähigkeit, attribution, gedanken
27	n.a.	veränderungen, verlauf, monate, zeitraum, messzeitpunkte, wochen, monate, hinweg, beginn, stabil

⁵⁴ Zusammenfassen von Wörtern gleichen Wortstammes nach der Modellinferenz (Schofield & Mimno, 2016)

⁵⁵ not available; kein entsprechendes Thema im Manuskript

30	10	psychoanalyse, theorie, entwicklung, psychoanalytisch, arbeit, psychologie, freud, begriff, versuch, überlegungen
35	27	frauen, männer, weiblich, männlich, geschlecht, geschlechtsspezifische, alter, attraktivität, körper, geschlechtsunterschiede
36	66	gesundheit, wohlbefinden, lebenszufriedenheit, subjektiv, ressourcen, alter, psychisch, zufriedenheit, befindlichkeit, erleben
37	98	frauen, partner, paare, partnerschaft, beziehung, schwangerschaft, männer, ehe, geburt, sexualität
40	47	kinder, alter, entwicklung, fähigkeiten, erwachsene, monate, vorschulalter, eltern, leistungen, verständnis
43	83	eltern, kinder, familien, behindert, elterlich, mütter, verhaltensauffälligkeiten, erziehungsverhalten, verhalten, alter
45	n.a.	positiv, negativ, effekt, einfluss, wahrnehmung, affekt, auswirkungen, hypothese, bewertung, erwartungen
46	86	deutsch, kultur, länder, kulturen, bundesrepublik, soziale, integration, identität, studenten, gruppen
49	84	sozial, unterstützung, beziehungen, kompetenz, identität, faktoren, ressourcen, kompetenzen, situation, integration
52	76	intelligenz, test, leistungen, kreativität, fähigkeiten, items, subtests, denken, alter, verfahren
54	96	klienten, therapeuten, therapie, therapeutisch, erwartungen, gespräch, ausmaß, verlauf, psychotherapie, veränderungen
57	94	informationen, entscheidungen, entscheidung, einfluss, unsicherheit, informationsverarbeitung, bewertung, versuchspersonen, präferenzen, sicherheit
60	51	schüler, klassen, schule, mathematik, schulisch, leistung, intelligenz, schulleistung, selbstkonzept, mathematisch
63	nicht reliabel ⁵⁶	depressiv, depression, patienten, symptomatik, depressivität, symptome, gesunde, kontrollpersonen, kontrollgruppe, klinisch
66	nicht reliabel	jugendliche, eltern, entwicklung, alter, jugendalter, kinder, junge, erwachsene, mädchen, familie
68	n.a.	prädiktoren, vorhersage, variablen, merkmale, faktoren, vorhersagen, varianz, arbeit, selbstwirksamkeit, regressionsanalysen
70	84	gruppen, mitglieder, teilnehmer, sozial, stärker, kategorie, versus, unterschiede, status, individuell

⁵⁶ Ein entsprechendes Thema resultierte zwar auch bei Analyse der standardisierten Schlagworte, könnte aber über verschiedene Modellinferenzen hinweg nicht hinreichend oft reproduziert werden.

71	38	patienten, leistungen, funktionen, kognitiv, defizite, neuropsychologisch, gesund, kontrollpersonen, beeinträchtigungen, gedächtnis
72	n.a.	interviews, qualitativ, auswertung, inhaltsanalyse, quantitativ, arbeit, subjektive, kategorien, verfahren, groundedtheory
74	69	emotionen, ärger, emotional, erleben, musik, angst, reaktionen, gefühle, ausdruck, situationen
76	95	psychologie, arbeit, psychologisch, theorie, entwicklung, forschung, theoretisch, ansätze, praxis, perspektive
78	n.a.	empirisch, arbeit, forschung, theoretisch, grundlagen, entwicklung, zusammenfassung, anschliessend, fragestellungen, einleitung
79	42	parameter, physiologisch, subjektiv, beanspruchung, aktivierung, reaktionen, psychophysiologischen, herzfrequenz, veränderung, angst
80	nicht reliabel	visuell, information, objekte, wahrnehmung, versuchspersonen, experimente, verarbeitung, räumlich, position, aufgabe
82	n.a.	implizit, explizit, motive, wissen, masse, nahe, zudem, trotz, legen, gemessen
83	n.a.	versuchspersonen, variablen, versus, einfluss, bedingungen, abhängige, experimentelle, wirkung, bestätigt, hypothese
84	nicht reliabel	patienten, behandlung, rehabilitation, stationär, psychiatrisch, ärzte, erkrankung, entlassung, versorgung, psychisch
85	33	erfassung, fragebogen, skala, eingesetzt, version, inventory, deutsche, verfahren, items, messung
86	38	patienten, gesund, schizophren, kontrollpersonen, schizophrenie, kontrollgruppe, probanden, gruppen, störungen, chronisch
87	nicht reliabel	störungen, psychisch, symptome, angststörungen, risikofaktoren, auffälligkeiten, interview, erkrankungen, symptomatik, prävalenz
91	21	verhalten, verhaltensweisen, kommunikation, interaktion, phase, verbale, äusserungen, interaktionen, beobachtung, hilfe
93	69	bilder, emotional, stimmung, valenz, stimuli, negativ, affektiv, neutral, positiv, verarbeitung
95	n.a.	submitted, enthält, psychology, beiträge, cognitive, research, social, study, effects, psychologie
96	60	schüler, lehrer, schule, unterricht, lehrkräfte, klassen, schulisch, verhalten, eltern, pädagogisch
98	42	aktivität, aktivierung, neuronal, kortex, funktionell, veränderungen, probanden, korrelate, fmrt, verarbeitung
100	n.a.	modelle, variablen, effekte, empirisch, theorie, form, erklärung, basis, vorhersagen, überprüfung

Tabelle D2. Auflistung aller Themen mit der Anzahl erfolgreicher Reproduktionen.

Tabelle D2. Auflistung aller Themen mit der Anzahl erfolgreicher Reproduktionen. Ein Thema wurde als reliabel eingestuft, wenn es bei mindestens acht von zehn verschiedenen Modellinferenzen reproduziert werden konnte (fett gedruckt).

Them a	Wahrscheinlichste Begriffe	Anzahl erfolgreicher Reproduktionen
1	symptoms, client characteristics, major depression, etiology, psychopathology	0
2	morality, social cognition, justice, attribution, motivation	9
3	aging, quality of life, social support, well being, caregivers	5
4	physicians, therapeutic processes, hospitals, nurses, client attitudes	2
5	group dynamics, group performance, interpersonal interaction, social interaction, group cohesion	10
6	psychosomatic disorders, personality correlates, coping behavior, personality traits, client characteristics	7
7	physiological correlates, heart rate, psychophysiology, physiological arousal, anxiety	10
8	psychotherapeutic outcomes, major depression, treatment outcomes, cognitive behavior therapy, treatment effectiveness evaluation	5
9	cognitive processes, problem solving, models, theory formulation, knowledge level	8
10	psychoanalytic theory, history of psychology, psychoanalysis, psychoanalytic interpretation, theories	9
11	electroencephalography, electrical activity, evoked potentials, brain, physiological correlates	10
12	schizophrenia, drug therapy, psychopharmacology, drugs, cognitive processes	0
13	criminals, juvenile delinquency, crime, prisoners, forensic psychiatry	10
14	human females, psychosexual behavior, sexuality, gender identity, human males	5
15	self concept, self perception, personality, self evaluation, self esteem	6
16	psychodiagnosis, test validity, questionnaires, diagnosis, rating scales	2
17	aggressive behavior, aggressiveness, computer games, television viewing, violence	6

18	career development, unemployment, occupational choice, occupational interests, occupational aspirations	0
19	coping behavior, stress management, stress, personality traits, anxiety	9
20	client characteristics, psychotherapy, psychotherapeutic outcomes, treatment outcomes, psychotherapeutic processes	4
21	verbal communication, nonverbal communication, interpersonal communication, language, interpersonal interaction	9
22	memory, human information storage, retention, cognitive processes, recognition (learning)	10
23	visual perception, stimulus parameters, visual stimulation, cognitive processes, color perception	7
24	methodology, statistical analysis, measurement, models, mathematical modeling	10
25	personality correlates, suicide, life experiences, coping behavior, attempted suicide	0
26	risk factors, major depression, mental disorders, epidemiology, at risk populations	7
27	human females, sex roles, human males, sex role attitudes, human sex differences	8
28	job satisfaction, working conditions, employee attitudes, organizational commitment, employee motivation	7
29	prevention, social skills, evaluation, social skills training, aggressive behavior	9
30	auditory perception, music, auditory stimulation, musical ability, music perception	7
31	preschool education, social interaction, residential care institutions, child welfare, socialization	0
32	cognitive processes, priming, human information storage, associative processes, memory	4
33	questionnaires, test validity, test construction, personality measures, test reliability	10
34	coping behavior, quality of life, neoplasms, client characteristics, well being	10
35	cognitive processes, human information storage, attention, human channel capacity, reaction time	8
36	perceptual motor coordination, motor processes, spatial orientation (perception), perceptual motor processes, motor performance	1
37	hydrocortisone, stress reactions, stress, physiological correlates,	10

	psychological stress	
38	cognitive impairment, cognitive ability, neuropsychological assessment, executive function, schizophrenia	9
39	cognitive development, childhood development, longitudinal studies, language development, intelligence	7
40	behavior therapy, behavior modification, treatment effectiveness evaluation, social skills, self control	6
41	history, history of psychology, fascism, society, germany	1
42	brain, physiological correlates, functional magnetic resonance imaging, neurophysiology, evoked potentials	10
43	motivation, personality traits, individual differences, social interaction, justice	0
44	sleep, physiological correlates, noise effects, emotional states, performance	3
45	learning, computer-assisted instruction, teaching methods, learning strategies, instructional media	10
46	personality traits, extraversion, neuroticism, personality correlates, individual differences	10
47	age differences, short-term memory, cognitive ability, aging, cognitive development	10
48	autism spectrum disorders, mental retardation, special education, autism, learning disorders	1
49	eye movements, visual perception, visual attention, attention, visual search	4
50	social perception, attribution, judgment, stereotyped attitudes, impression formation	9
51	academic achievement, student characteristics, college students, high school students, elementary school students	9
52	sports, athletes, athletic performance, athletic participation, personality correlates	8
53	work teams, human resource management, organizational behavior, group performance, cooperation	1
54	alcoholism, alcohol rehabilitation, treatment outcomes, treatment effectiveness evaluation, relapse (disorders)	7
55	goals, motivation, self-regulation, intention, behavior	9
56	elementary school students, dyslexia, reading skills, mathematical ability, mathematics achievement	5
57	death and dying, religiosity, religious beliefs, coping behavior,	4

	spirituality	
58	genetics, genes, dopamine, polymorphism, genotypes	1
59	working conditions, human factors engineering, occupational stress, occupational safety, human machine systems	0
60	teacher-student interaction, teacher attitudes, teacher characteristics, teachers, teacher education	10
61	brain damage, cerebrovascular accidents, neuropsychological assessment, brain disorders, neuropsychology	3
62	test validity, test construction, measurement, personality measures, projective personality measures	5
63	tobacco smoking, alcohol drinking patterns, drug usage, alcohol abuse, alcoholism	7
64	cognitive development, cognitive processes, reasoning, problem solving, theory of mind	5
65	consumer behavior, consumer attitudes, marketing, consumer research, advertising	7
66	health behavior, health promotion, health, lifestyle, physical activity	9
67	language, semantics, syntax, cognitive processes, evoked potentials	10
68	life experiences, adult development, self concept, personality development, coping behavior	3
69	emotions, emotional states, facial expressions, emotional responses, emotional regulation	9
70	obesity, eating disorders, bulimia, human females, anorexia nervosa	10
71	management personnel, human resource management, business organizations, business management, leadership	10
72	driving behavior, drivers, highway safety, automobiles, motor traffic accidents	10
73	environmental attitudes, conservation (ecological behavior), urban environments, home environment, architecture	8
74	counseling, psychotherapy training, social casework, professional competence, professional supervision	4
75	brain, functional magnetic resonance imaging, lateral dominance, neuroanatomy, magnetic resonance imaging	6
76	test validity, intelligence measures, test construction, measurement, testing	10
77	psychiatric patients, treatment outcomes, schizophrenia, psychiatric hospitals, rehabilitation	5

78	mother-child relations, attachment behavior, infant development, mothers, mother-child communication	8
79	rats, learning, memory, operant conditioning, pigeons	3
80	computer applications, internet, human machine systems, human computer interaction, human factors engineering	3
81	educational program evaluation, teaching methods, academic achievement, educational programs, learning	5
82	treatment effectiveness evaluation, treatment outcomes, coping behavior, psychotherapeutic outcomes, behavioral medicine	3
83	parent-child relations, family relations, childrearing practices, mothers, childhood development	8
84	social perception, ingroup outgroup, social identity, stereotyped attitudes, attribution	9
85	sociocultural factors, cross cultural differences, social influences, culture (anthropological), witnesses	5
86	human migration, cross-cultural differences, acculturation, immigration, cross-cultural communication	10
87	conflict, interpersonal interaction, conflict resolution, interpersonal communication, verbal communication	1
88	attribution, self-concept, emotional states, internal external locus of control, achievement	9
89	personnel selection, aptitude measures, test validity, statistical validity, test construction	10
90	human females, pregnancy, birth, depression (emotion), mothers	7
91	chronic pain, pain perception, pain management, pain, back pain	9
92	age differences, human sex differences, intelligence, individual differences, personality traits	0
93	adolescent development, psychosocial development, personality development, self concept, family relations	6
94	decision making, cognitive processes, judgment, choice behavior, uncertainty	10
95	theories, philosophies, history of psychology, theory formulation, psychology	10
96	psychotherapeutic processes, psychotherapeutic outcomes, client-centered therapy, psychotherapeutic techniques, psychoanalysis	10
97	posttraumatic stress disorder, emotional trauma, coping behavior, trauma, victimization	10
98	marital relations, couples, marital satisfaction, spouses, relationship	10

	quality	
99	occupational stress, stress management, coping behavior, working conditions, health	10
100	attitude change, mass media, messages, persuasive communication, information	2

PsychArchives-ESM 1: Detaillierte Beschreibung des methodischen Vorgehens.

PsychArchives-ESM 1

Elektronisches Supplement zu Bittermann, Greiner & Fischer: "Unterscheiden sich die Forschungsinteressen von Frauen und Männern in der Psychologie?"

Im Folgenden wird das methodische Vorgehen in Ergänzung zum Methodenabschnitt im Artikel ausführlicher beschrieben.

Software

Die Analysen wurden mit RStudio Version 1.1.456 (RStudio Team, 2016) basierend auf R Version 3.5.1 (R Core Team, 2018) durchgeführt. Für die Aufbereitung der Textdaten wurde *quanteda* 1.4.1 (Benoit et al., 2018), für die Bestimmung des Geschlechts *genderizeR* 2.1.0 (Wais, 2006), für Topic Modeling *topicmodels* 0.2-8 (Grün & Hornik, 2011) und für die Ermittlung der zeitlichen Trends wurden die Pakete *nnet* 7.3-12 (Venables & Ripley, 2002) sowie *quantqual* 0.0.2 (Fischer, 2019) verwendet.

Bestimmung des Geschlechts

Die Datengrundlage bildeten $N = 18\,050$ in der psychologischen Referenzdatenbank PSYNDEX nachgewiesene Dissertationen aus dem deutschen Sprachraum und den Jahren 1968 bis 2017. Dissertationen wurden für die Untersuchung von Forschungsinteressen deshalb ausgewählt, da ihr Inhalt klar der wissenschaftlichen Arbeit einer Person zugeordnet werden kann. Selbst bei kumulativen, publikationsbasierten Promotionen formt die promovierende Person den thematischen Zusammenhang.

Das Geschlecht der verfassenden Person wurde anhand des Vornamens (Naldi, Luzi, Valente & Parenti, 2004) mithilfe des *genderizeR*-Klassifikationsalgorithmus (Wais, 2006) bestimmt. Fell und König (2016) berichten in ihrer szientometrischen Anwendung dieses Algorithmus eine Korrelation von $r = .95$ zwischen wahrem Geschlecht und der automatischen Zuordnung. Wir haben in den 1787 Fällen, in denen der Algorithmus mit einer Sicherheit von weniger als 95 % oder gar nicht zuordnen konnte, sowie bei Unisex-Vornamen (z. B. Dominique, Eike, Luca), das Geschlecht durch manuelle Recherche bestimmt. Insgesamt 79 Personen konnten nach diesem Vorgehen keinem Geschlecht zugeordnet werden, womit der finale Datensatz aus $n = 17\,971$ Dissertationen nachweisen bestand.

Topic Modeling

Controlled Terms

Der Inhalt der Dissertationen wurde wie bei Bittermann und Fischer (2018) anhand der standardisierten Schlagworte in PSYNDEX erfasst (englisch: „Controlled Terms“, im Folgenden: CT). Diese CT sind nicht die freien Schlagworte der Autorinnen und Autoren, sondern sind im *Thesaurus of Psychological Index Terms* (Tuleya, 2007) der American Psychological Association definiert¹. Aus einem Pool von aktuell über 6500 Begriffen werden die CT vom wissenschaftlichen Personal der Informations- und Recherchedienste am ZPID vergeben. Damit wird sichergestellt, dass die zentralen Inhalte einer Publikation durch diese CT abgebildet sind (Gerards, Gerards, Kuhberg-Lasson, Singleton & Trierweiler, 2014, S. 13). Die CT korrespondieren somit mit den Inhalten der Abstracts. Daher repräsentieren CT genau das, was Text Mining von Freitexten zum Ziel hat: Die Extraktion bedeutsamer Information aus Texten (Allahyari et al., 2017). Die Verwendung von CT löst außerdem viele Probleme der Datenaufbereitung von Freitexten (etwa Synonyme, Flexionen, sog. *Stop Words*² usw.) und bietet bei einem Korpus, das die thematische Breite der gesamten Psychologie abdeckt³, einen optimalen Kompromiss aus Detailgrad und Zusammenfassung von Inhalten (für weitere Vorteile siehe Bittermann & Fischer, 2018, S. 5 f).

Von den Vorteilen der CT für unsere Fragestellung abgesehen hätte die Verwendung von Abstracts Mehrsprachigkeit zum Problem gehabt. Mehrsprachigkeit kann bei Topic Modeling zu verschiedenen Themen trotz gleicher Inhalte führen (z. B. englisches "Therapy"-Thema und deutsches "Therapie"-Thema), bzw. die semantische Kohärenz der Themen (Mimno, Wallach, Talley, Leenders & McCallum, 2011) senken (etwa, wenn bei kumulativen Dissertationen in einem deutschsprachigen Abstract die englischen Titel der Einzelbeiträge aufgeführt sind). Daher wird für Topic Modeling in der Regel ein sprachlich homogenes Korpus verwendet, was in unserem Fall zum Ausschluss von 10.3 % der Dissertationen geführt hätte (weder ein deutsches Abstract noch eine Übersetzung wurden publiziert). Dabei ist erstens festzustellen, dass der englischsprachige Ausschluss eine andere thematische Verteilung aufweist als der Rest, vor allem höhere Anteile von Dissertationen aus den Feldern der Neuropsychologie, Elektrophysiologie und Allgemeinen Psychologie⁴. Zweitens ist der Frauenanteil im Ausschluss höher: 65 % vs. 52 % im Rest⁵. Schließlich ist drittens entsprechend des zunehmenden Trends englischsprachiger Publikationen aus den deutschsprachigen Ländern (Schui & Krampen, 2016) zu betonen, dass bei Ausschluss der Dissertationen mit nur englischsprachigen Abstracts ab Publikationsjahr 2001 für manche

¹ Vor allem bei methodischen Arbeiten ist hier sichergestellt, dass ein Verfahren wie etwa "Regressionsanalyse" nicht nur in einer empirischen Analyse eingesetzt wurde, sondern die Methode selbst im Fokus der Publikation stand. Der Thesaurus kann bei ZPID (2016) eingesehen werden.

² Dies sind Wörter, die wenig inhaltliche Relevanz haben (z. B. "die", "ist", "und").

³ Abstracts oder Volltexte sind CT natürlich dann vorzuziehen, wenn das Korpus bereits thematisch eingeschränkt ist und die Detailtiefe von CT nicht mehr ausreichend ist (siehe z. B. Bittermann & Klos, 2019).

⁴ thematische Grobanalyse anhand des PSYNDEX-Klassifikationssystems (ZPID, 2016); $\chi^2(93, 28074) = 1808.98, p < .0001, \text{Cramer's } V = .25$

⁵ $\chi^2(1, 17971) = 105.12, p < .0001, \text{Phi} = .08$

Jahre bis zu 51 % der Arbeiten in unserer Analyse gefehlt hätten⁶. Somit wäre der Ausschluss keine zufällige Stichprobe, vielmehr wären unsere Ergebnisse thematisch und zeitlich verzerrt gewesen.

Topic-Modeling-Verfahren für mehrsprachige Textquellen (z. B. Mimno, Wallach, Naradowsky, Smith & McCallum, 2009; Vulić, De Smet, Tang, & Moens, 2015) erzielen nach unseren Erfahrungen keine befriedigenden Ergebnisse und stellen hohe Anforderungen der Datenvorbereitung, die in unserem Fall nicht erfüllt werden konnten: Thematisch passende Dokumente müssen sprachübergreifend vorab gepaart werden, was bei der sehr ungleichen Verteilung in unserem Korpus nicht adäquat möglich gewesen wäre.

Latent Dirichlet Allocation

Um die den Dissertationen zugrunde liegenden Themen zu bestimmen, wurde Topic Modeling basierend auf *Latent Dirichlet Allocation* (LDA; Blei, Ng & Jordan, 2003) mit Gibbs-Sampling⁷ angewandt. Ziel dieses Verfahrens aus dem maschinellen Lernen ist es, solche Gruppen von Wörtern zu identifizieren, die häufig miteinander in einem Dokument enthalten sind und damit die Themen im Korpus repräsentieren. Zur Identifikation von Themen der Fachliteratur erwies sich Topic Modeling der Verwendung eines Klassifikationssystems (wie z. B. bei König et al., 2015; Preckel & Krampen, 2016) hinsichtlich Detailgrad und Flexibilität überlegen (Bittermann & Fischer, 2018; Ding, 2011), da anhand der Klassifikationskategorien von Literaturdatenbanken nur solche Themen gefunden werden können, die vorab festgelegt wurden.

Bei der Bestimmung der Themen folgten wir den Best-Practice-Empfehlungen von Maier et al. (2018). Es wurden Modelle mit verschiedenen LDA-Hyperparametern⁸ und unterschiedlichen Anzahlen von Themen (75 bis 200) geprüft. Das finale Modell ($k = 100$, $\alpha = 0.01$, $\delta = 0.01$) wurde unter Einbezug von Evaluationsmetriken, anhand qualitativer Prüfung von Dokument-Thema-Zuordnungen und angesichts der besten intersubjektiven Interpretierbarkeit gewählt. Die Themen dieses finalen Modells wurden anschließend hinsichtlich ihrer Reliabilität untersucht (Niekler, 2016). Das heißt, es wurde geprüft, ob sie bei wiederholten Modellinferenzen reproduzierbar sind. Dieses Vorgehen stellt sicher, dass die gefundenen Themen nicht nur aufgrund lokaler Optima der LDA-Modellinferenz zustande kamen. Entsprechend des Ansatzes von Niekler (2016) wurde die Themen-Reliabilität mittels Kosinus-Ähnlichkeit analoger Themen-Paare von mindestens .8 bei acht von zehn wiederholten Modellinferenzen bestimmt. Dieses Vorgehen verringert die Anzahl der Themen zugunsten robusterer Ergebnisse. Für weitere Details verweisen wir auf den kommentierten Code, der in PsychArchives-ESM 2 einsehbar ist.

⁶ $\chi^2(16, 8846) = 776.88, p < .0001, \text{Cramer's } V = .30$

⁷ Mit default-Kontrollparametern des *topicmodels*-Pakets; allerdings wurde das beste Modell aus zehn Random Starts mit unterschiedlichen Seeds über das Produkt aus Semantischer Kohärenz (Mimno et al., 2011) und Exklusivität (Roberts et al., 2014) bestimmt, nicht über die heldout-Likelihood.

⁸ alpha-Werte von 0.0001, 0.001 und 0.01, fixed delta = 0.01

Einordnung von Prävalenzdifferenzen

Die Prävalenz beschreibt die durchschnittliche Wahrscheinlichkeit eines Themas im Korpus und wird über die mittlere Dokument-Thema-Wahrscheinlichkeit berechnet. Bei Gleichverteilung aller Themen hat jedes Thema eine Wahrscheinlichkeit von $1/k$, da sich die Wahrscheinlichkeitswerte aller Themen zu 1 addieren. So beträgt bei $k = 100$ Themen die durchschnittliche Prävalenz $1/k = 1/100 = 1\%$ (wenn also jedes Dokument jedes Thema gleich stark behandelt). Dies bedeutet also, dass die beobachteten Prävalenzen der einzelnen Themen umso kleiner ausfallen, je mehr Themen im Modell enthalten sind. Gleiches gilt entsprechend für die Prävalenzdifferenzen zweier Themen oder zweier Gruppen.

Um nun zu beurteilen, wann eine absolute Prävalenzdifferenz ($|D|$) als groß oder klein angenommen werden kann, ist $1/k$ ein hilfreicher Ausgangspunkt. Aufgrund fehlender Standards in der Topic-Modeling-Literatur haben wir folgendes Kriterium festgelegt: Eine Prävalenzerhöhung, die die Prävalenz eines anderen Themas – unter Konstanzhaltung aller anderen Themenprävalenzen – mehr als halbiert (also die halbe durchschnittliche Prävalenz $= 0.5 \cdot 1/k$), wurde als ein größerer Unterschied angesehen. Differenzen $\leq 0.5 \cdot 1/k$ erachten wir entsprechend als einen geringen Unterschied. Im Falle von $k = 100$ Themen bedeutet dies, dass $|D| = 0.5\%$ der Cut-Off zwischen größeren und geringen Unterschieden ist. Dieser Cut-Off gilt auch nach Ausschluss unreliabler Themen (s. o.), da die Prävalenzen der eingeschlossenen Themen während der Inferenz eines Modells mit $k = 100$ Themen berechnet wurden.

Frauenanteile

Die Frauenanteile an den Themen können nicht unmittelbar aus dem Topic Model abgeleitet werden. Dazu muss erst festgelegt werden, wann ein Dokument zu einem Thema gezählt werden kann. Wir haben dies wie folgt operationalisiert: Der Frauenanteil wurde anhand der Anzahl derjenigen Dissertationen bestimmt, in denen das jeweilige Thema das insgesamt Dominierende ist (d. h., der Inhalt der Dissertation lässt sich zu mindestens 50 % diesem Thema zuordnen und zu höchstens 50 % anderen Themen⁹). Wenn also etwa von 100 Dissertationen 40 eine Dokument-Thema-Wahrscheinlichkeit von mindestens 50 % aufweisen, dann werden diese 40 Dissertationen als zu diesem Thema zugehörig gezählt. Anschließend wird der Frauenanteil an diesen 40 Dissertationen ausgezählt. Zwar geht durch diesen Cut-Off Information verloren, für illustrative Zwecke ist der Frauenanteil an einem Thema allerdings intuitiver erfassbar als Prävalenzdifferenzen.

⁹ z. B.: Die Behandlung der Depression ist Hauptthema der Dissertation, thematisch werden in geringerem Umfang auch Diagnostik und Epidemiologie angeschnitten.

Zeitliche Trends

Um die Zeitverläufe der Themen von Schwankungen zu befreien, wurden Trendkurven getrennt nach Geschlecht mithilfe von *Multilayer Perceptrons* (MLP) berechnet. Die MLPs ermöglichen nichtlineare Regressionsfunktionen mit einer minimalen Summe von quadratischen Residuen für jedes Thema. Für jedes MLP wurde eine Struktur mit zwei Hidden Units berechnet¹⁰, um die durchschnittliche Themenwahrscheinlichkeit als nichtlineare Funktion des Erscheinungsjahres zu modellieren (vgl. Bittermann & Fischer, 2018). Zwei Hidden Units wurden gewählt, um einerseits – im Gegensatz zu weniger komplexen Strukturen – auch nicht-monotone Funktionen zu ermöglichen und andererseits – im Vergleich zu komplexeren Strukturen – das Risiko (und das Ausmaß) einer Überanpassung (engl. *Overfitting*) des Modells an die Daten zu reduzieren (Fischer, 2015). Die zeitlichen Trends wurden schließlich auf drei Dimensionen untersucht:

- (1) Dimension “Geschlecht”, das heißt die mittlere Differenz zwischen der weiblichen und männlichen Trendkurve,
- (2) Dimension “Konvergenz” (im Sinne einer über die Zeit zunehmenden Nivellierung zweier Trends) anhand eines lineares Regressionsmodells mit Prädiktor Publikationsjahr und Kriterium Differenz der weiblichen und männlichen Trendkurve, sowie
- (3) Dimension “Trend”, das heißt der lineare Gesamttrend des Themas, unabhängig vom Geschlecht.

¹⁰ Um im Rahmen des Schätzprozesses lokale Minima zu vermeiden, wurde analog zu Fischer (2015) der Schätzprozess zehn mal wiederholt und das beste Modell zur Analyse der jeweiligen Trendkurve herangezogen.

Liste der verwendeten R-Pakete

```
> sessionInfo()
R version 3.5.1 (2018-07-02)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 7 x64 (build 7601) Service Pack 1

Matrix products: default

locale:
[1] LC_COLLATE=German_Germany.1252 LC_CTYPE=German_Germany.1252 LC_MONETARY=German_Germany.1252
LC_NUMERIC=C
[5] LC_TIME=German_Germany.1252

attached base packages:
[1] stats graphics grDevices utils datasets methods base

other attached packages:
[1] nnet_7.3-12 quantqual_0.0.2 genderizeR_2.1.0 data.table_1.12.0 stringr_1.4.0 dplyr_0.8.0.1
topicmodels_0.2-8
[8] quanteda_1.4.1

loaded via a namespace (and not attached):
 [1] modeltools_0.2-22 tidyselect_0.2.5 NLP_0.2-0 slam_0.1-45 reshape2_1.4.3 purrr_0.3.1
 [7] splines_3.5.1 lattice_0.20-38 colorspace_1.4-0 generics_0.0.2 stats4_3.5.1 yaml_2.2.0
[13] survival_2.43-3 prodlim_2018.04.18 rlang_0.3.1 ModelMetrics_1.2.2 pillar_1.3.1 glue_1.3.1
[19] withr_2.1.2 foreach_1.4.4 plyr_1.8.4 lava_1.6.5 timeDate_3043.102 munsell_0.5.0
[25] gtable_0.2.0 recipes_0.1.4 codetools_0.2-16 caret_6.0-81 parallel_3.5.1 tm_0.7-6
[31] class_7.3-15 Rcpp_1.0.0 spacyr_1.0 scales_1.0.0 ipred_0.9-8 RcppParallel_4.4.2
[37] fastmatch_1.1-0 stopwords_0.9.0 ggplot2_3.1.0 stringi_1.3.1 grid_3.5.1 tools_3.5.1
[43] magrittr_1.5 lazyeval_0.2.1 tibble_2.0.1 crayon_1.3.4 pkgconfig_2.0.2 MASS_7.3-51.1
[49] Matrix_1.2-15 xml2_1.2.0 lubridate_1.7.4 gower_0.2.0 assertthat_0.2.0 rstudioapi_0.9.0
[55] iterators_1.0.10 R6_2.4.0 rpart_4.1-13 nlme_3.1-137 compiler_3.5.1
```

Literatur

- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint*. <https://arxiv.org/abs/1707.02919v2>
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3 (30), 774. <http://doi.org/10.21105/joss.00774>
- Bittermann, A. & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, 226 (1), 3–13. <https://doi.org/10.1027/2151-2604/a000318>
- Bittermann, A. & Klos, E. M. (2019). Ist die psychologische Forschung durchlässig für aktuelle gesellschaftliche Themen? Eine szientometrische Analyse am Beispiel Flucht

- und Migration mithilfe von Topic Modeling. *Psychologische Rundschau*, 70 (4), 239–249. <https://doi.org/10.1026/0033-3042/a000426>
- Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics*, 5 (1), 187–203. <https://doi.org/10.1016/j.joi.2010.10.008>
- Fell, C. B. & König, C. J. (2016). Is there a gender difference in scientific collaboration? A scientometric examination of co-authorships among industrial–organizational psychologists. *Scientometrics*, 108 (1), 113–141. <https://doi.org/10.1007/s11192-016-1967-5>
- Fischer, A. (2015). How to determine the unique contributions of input-variables to the nonlinear regression function of a multilayer perceptron. *Ecological Modelling*, 309, 60–63. <https://doi.org/10.1016/j.ecolmodel.2015.04.015>
- Fischer, A. (2019). *quantqual: Software package for analyzing quantitative and qualitative data*. [Computer software]. Verfügbar unter <https://github.com/AndreasFischer1985/quantqual>
- Gerards, A., Gerards, M., Kuhberg-Lasson, V., Singleton, K. & Trierweiler, L. (2014). Dokumentationsmanual PSYINDEX (aktualisierte Ausgabe). Trier: ZPID. Verfügbar unter <https://www.psyindex.de/download/PSYINDEXmaterial/psyindex-dokumentationsmanual.pdf>
- Grün, B. & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40, 1–30. <https://doi.org/10.18637/jss.v040.i13>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri H. & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12 (2–3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- Mimno, D., Wallach, H. M., Naradowsky, J., Smith, D. A., & McCallum, A. (2009, August). Polylingual topic models. In P. Koehn & R. Mihalcea (Hrsg.), *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2* (S. 880–889). Stroudsburg, PA: Association for Computational Linguistics. <https://www.aclweb.org/anthology/D09-1092.pdf>
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011, July). Optimizing semantic coherence in topic models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (262–272). Association for Computational Linguistics. Chicago.

- Naldi, F., Luzi, D., Valente, A., & Parenti, I. V. (2004). Scientific and technological performance by gender. In H. F. Moed, W. Glänzel, & U. Schmoch (Hrsg.), *Handbook of quantitative science and technology research* (S. 299–314). Dordrecht: Kluwer.
<http://dx.doi.org/10.1007/1-4020-2755-9>
- Niekler, A. (2016). *Automatisierte Verfahren für die Themenanalyse nachrichtenorientierter Textquellen*. Dissertation, Universität Leipzig. Verfügbar unter:
http://asv.informatik.uni-leipzig.de/publication/file/350/Niekler_Diss.pdf
- Preckel, F. & Krampen, G. (2016). Entwicklung und Schwerpunkte in der psychologischen Hochbegabungsforschung. *Psychologische Rundschau*, 67 (1), 1–14.
<https://doi.org/10.1026/0033-3042/a000289>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [Computer software]. Verfügbar unter:
<https://www.R-project.org/>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58 (4), 1064–1082.
<https://doi.org/10.1111/ajps.12103>
- RStudio Team. (2016). *RStudio: Integrated development for R* [Computer software]. Boston, MA: RStudio, Inc. Verfügbar unter: <http://www.rstudio.com/>
- Schui, G. & Krampen, G. (2016). ZPID-Monitor 2014 zur Internationalität der Psychologie aus dem deutschsprachigen Bereich: Der ausführliche Bericht. *ZPID Science Information Online*, 16 (2). Verfügbar unter
https://www.zpid.de/pub/research/zpid-monitor_2014_lang.pdf
- Tuleya L. G. (Hrsg.). (2007). *Thesaurus of psychological index terms* (11th ed.). Washington, DC: American Psychological Association.
- Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth Edition). Springer, New York.
- Vulić, I., De Smet, W., Tang, J., & Moens, M. F. (2015). Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. *Information Processing & Management*, 51, 111–147. <https://doi.org/10.1016/j.ipm.2014.08.003>
- Wais, K. (2006). Gender Prediction Methods Based on First Names with genderizeR. *The R Journal*, 8 (1), 17–37. <http://doi.org/10.32614/RJ-2016-002>
- ZPID – Leibniz-Zentrum für Psychologische Information und Dokumentation (Hrsg.). (2016). *PSYINDEX Terms* (10. Aufl.). Trier: ZPID. Verfügbar unter:
<https://www.psyindex.de/pub/info/PSYINDEXterms2016.pdf>

PsychArchives-ESM 2: R-Code der Analysen

Der R-Code der Analysen ist in digitaler Form bei PsychArchives.org verfügbar:
<http://dx.doi.org/10.23668/psycharchives.2677>

PsychArchives-ESM 3: Abbildung zu Frauenanteilen an den Promotionen.

PsychArchives-ESM 3

Elektronisches Supplement zu Bittermann, Greiner & Fischer: "Unterscheiden sich die Forschungsinteressen von Frauen und Männern in der Psychologie?"

Im Folgenden findet sich eine ergänzende Abbildung 2 zu den Ergebnissen (Abbildung 1 ist im Artikel enthalten).

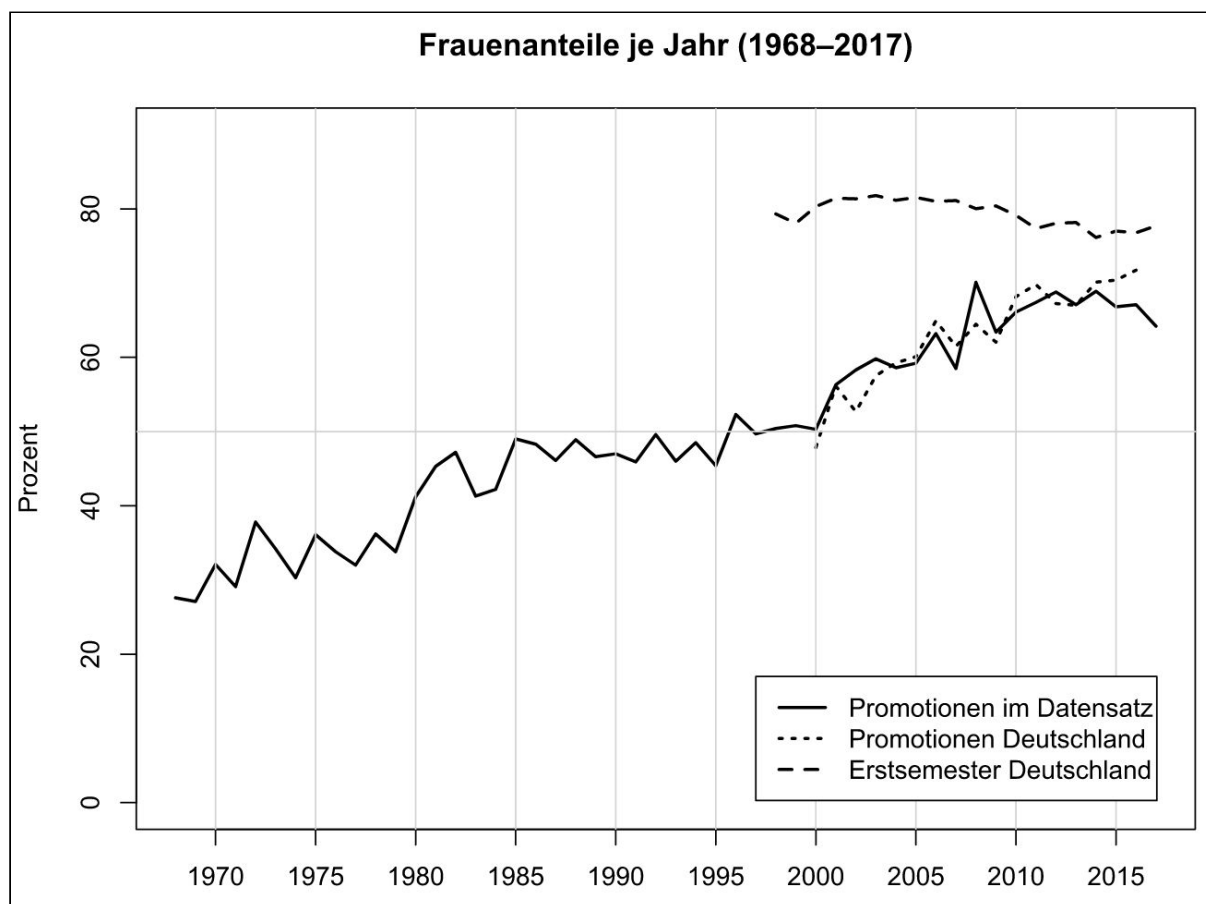


Abbildung 2. Frauenanteile an den Promotionen im Datensatz von 1968 bis 2017. Zum Vergleich sind die von Antoni (2019) berichteten Anteile an den Promotionen in Deutschland sowie die vom Statistischen Bundesamt (Destatis, 2019) verfügbaren Daten zu Anteilen an den Erstsemestern eingezeichnet.

Antoni, C. H. (2019). Zur Lage der Psychologie. *Psychologische Rundschau*, 70, 4-26.

<https://doi.org/10.1026/0033-3042/a000429>

Destatis. (2019). Statistisches Bundesamt. *Studienanfänger: Deutschland, Semester, Nationalität, Geschlecht, Studienfach*. Abgerufen von <https://www-genesis.destatis.de> (14.02.2019)

PsychArchives-ESM 4: Auflistung aller Themen.

PsychArchives-ESM 4

Elektronisches Supplement zu Bittermann, Greiner & Fischer: "Unterscheiden sich die Forschungsinteressen von Frauen und Männern in der Psychologie?"

Im Folgenden sind die jeweils fünf wahrscheinlichsten Begriffe alle Themen sowie die jeweiligen Prävalenzdifferenzen D und Frauenanteile aufgelistet. Unreliable Themen wurden ausgeschlossen. Tabelle 1 ist nach Thema, Tabelle 2 nach Prävalenzdifferenz (absteigend) und Tabelle 3 nach Frauenanteil (absteigend) sortiert.

Die Frauenanteile an den Themen können nicht unmittelbar aus dem Topic Model abgeleitet werden. Dazu muss erst festgelegt werden, wann ein Dokument zu einem Thema gezählt werden kann. Wir haben dies wie folgt operationalisiert: Der Frauenanteil wurde anhand der Anzahl derjenigen Dissertationen bestimmt, in denen das jeweilige Thema das insgesamt Dominierende ist (d. h., der Inhalt der Dissertation lässt sich zu mindestens 50 % diesem Thema zuordnen und zu höchstens 50 % anderen Themen¹). Wenn also etwa von 100 Dissertationen 40 eine Dokument-Thema-Wahrscheinlichkeit von mindestens 50 % aufweisen, dann werden diese 40 Dissertationen als zu diesem Thema zugehörig gezählt. Anschließend wird der Frauenanteil an diesen 40 Dissertationen ausgezählt. Zwar geht durch diesen Cut-Off Information verloren, für illustrative Zwecke ist der Frauenanteil an einem Thema allerdings intuitiver erfassbar als Prävalenzdifferenzen. Abweichungen der Ränge zwischen Prävalenzdifferenzen und Frauenanteil ergeben sich dadurch, dass in die Berechnung von Letzterem weniger Information einfließt.

Tabelle 1 (sortiert nach Thema):	Seite 2
Tabelle 2 (sortiert nach Prävalenzdifferenz):	Seite 5
Tabelle 3 (sortiert nach Frauenanteil):	Seite 8

¹ z. B.: Die Behandlung der Depression ist Hauptthema der Dissertation, thematisch werden in geringerem Umfang auch Diagnostik und Epidemiologie angeschnitten.

Tabelle 1. Übersicht über die Themen – sortiert nach Thema.

Thema	Wahrscheinlichste Begriffe	D* (%)	Frauenanteil (%)
2	morality, social cognition, justice, attribution, motivation	-0.10	47.83
5	group dynamics, group performance, interpersonal interaction, social interaction, group cohesion	-0.28	37.96
7	physiological correlates, heart rate, psychophysiology, physiological arousal, anxiety	-0.15	47.09
9	cognitive processes, problem solving, models, theory formulation, knowledge level	-0.76	32.00
10	psychoanalytic theory, history of psychology, psychoanalysis, psychoanalytic interpretation, theories	-0.19	39.83
11	electroencephalography, electrical activity, evoked potentials, brain, physiological correlates	-0.27	50.89
13	criminals, juvenile delinquency, crime, prisoners, forensic psychiatry	-0.10	49.67
19	coping behavior, stress management, stress, personality traits, anxiety	0.17	59.63
21	verbal communication, nonverbal communication, interpersonal communication, language, interpersonal interaction	-0.13	44.44
22	memory, human information storage, retention, cognitive processes, recognition (learning)	0.07	47.46
24	methodology, statistical analysis, measurement, models, mathematical modeling	-1.01	23.12
27	human females, sex roles, human males, sex role attitudes, human sex differences	0.99	81.90
29	prevention, social skills, evaluation, social skills training, aggressive behavior	0.24	78.57
33	questionnaires, test validity, test construction, personality measures, test reliability	-0.13	52.11
34	coping behavior, quality of life, neoplasms, client characteristics, well being	0.48	68.90
35	cognitive processes, human information storage, attention, human channel capacity, reaction time	-0.37	48.54
37	hydrocortisone, stress reactions, stress, physiological correlates, psychological stress	0.08	66.67
38	cognitive impairment, cognitive ability, neuropsychological assessment, executive function, schizophrenia	0.34	66.50
42	brain, physiological correlates, functional magnetic resonance imaging, neurophysiology, evoked potentials	-0.11	61.21
45	learning, computer-assisted instruction, teaching methods, learning strategies, instructional media	-0.12	55.00
46	personality traits, extraversion, neuroticism, personality correlates, individual differences	0.13	50.00

47	age differences, short-term memory, cognitive ability, aging, cognitive development	0.06	62.60
50	social perception, attribution, judgment, stereotyped attitudes, impression formation	0.48	67.59
51	academic achievement, student characteristics, college students, high school students, elementary school students	0.03	51.76
52	sports, athletes, athletic performance, athletic participation, personality correlates	-0.33	32.88
55	goals, motivation, self-regulation, intention, behavior	-0.20	57.98
60	teacher-student interaction, teacher attitudes, teacher characteristics, teachers, teacher education	-0.01	50.66
66	health behavior, health promotion, health, lifestyle, physical activity	0.16	67.39
67	language, semantics, syntax, cognitive processes, evoked potentials	-0.10	51.38
69	emotions, emotional states, facial expressions, emotional responses, emotional regulation	0.08	61.06
70	obesity, eating disorders, bulimia, human females, anorexia nervosa	0.51	81.44
71	management personnel, human resource management, business organizations, business management, leadership	-0.24	51.30
72	driving behavior, drivers, highway safety, automobiles, motor traffic accidents	-0.69	34.03
73	environmental attitudes, conservation (ecological behavior), urban environments, home environment, architecture	-0.16	44.34
76	test validity, intelligence measures, test construction, measurement, testing	-0.61	36.78
78	mother-child relations, attachment behavior, infant development, mothers, mother-child communication	0.72	74.32
83	parent-child relations, family relations, childrearing practices, mothers, childhood development	0.91	72.50
84	social perception, ingroup outgroup, social identity, stereotyped attitudes, attribution	-0.03	54.17
86	human migration, cross-cultural differences, acculturation, immigration, cross-cultural communication	0.00	51.92
88	attribution, self-concept, emotional states, internal external locus of control, achievement	0.00	45.11
89	personnel selection, aptitude measures, test validity, statistical validity, test construction	-0.47	39.24
91	chronic pain, pain perception, pain management, pain, back pain	0.18	71.08
94	decision making, cognitive processes, judgment, choice behavior, uncertainty	-0.65	37.67
95	theories, philosophies, history of psychology, theory formulation, psychology	-1.06	25.31
96	psychotherapeutic processes, psychotherapeutic outcomes, client-centered	0.03	49.27

	therapy, psychotherapeutic techniques, psychoanalysis		
97	posttraumatic stress disorder, emotional trauma, coping behavior, trauma, victimization	0.27	73.73
98	marital relations, couples, marital satisfaction, spouses, relationship quality	0.26	64.71
99	occupational stress, stress management, coping behavior, working conditions, health	-0.01	60.80

Anmerkung: **D* = Prävalenzdifferenz (weiblich - männlich)

Tabelle 2. Übersicht über die Themen – sortiert nach Prävalenzdifferenz (*D*) (absteigend).

Thema	Wahrscheinlichste Begriffe	<i>D</i> * (%)	Frauenanteil (%)
27	human females, sex roles, human males, sex role attitudes, human sex differences	0.99	81.90
83	parent-child relations, family relations, childrearing practices, mothers, childhood development	0.91	72.50
78	mother-child relations, attachment behavior, infant development, mothers, mother-child communication	0.72	74.32
70	obesity, eating disorders, bulimia, human females, anorexia nervosa	0.51	81.44
34	coping behavior, quality of life, neoplasms, client characteristics, well being	0.48	68.90
50	social perception, attribution, judgment, stereotyped attitudes, impression formation	0.48	67.59
38	cognitive impairment, cognitive ability, neuropsychological assessment, executive function, schizophrenia	0.34	66.50
97	posttraumatic stress disorder, emotional trauma, coping behavior, trauma, victimization	0.27	73.73
98	marital relations, couples, marital satisfaction, spouses, relationship quality	0.26	64.71
29	prevention, social skills, evaluation, social skills training, aggressive behavior	0.24	78.57
91	chronic pain, pain perception, pain management, pain, back pain	0.18	71.08
19	coping behavior, stress management, stress, personality traits, anxiety	0.17	59.63
66	health behavior, health promotion, health, lifestyle, physical activity	0.16	67.39
46	personality traits, extraversion, neuroticism, personality correlates, individual differences	0.13	50.00
37	hydrocortisone, stress reactions, stress, physiological correlates, psychological stress	0.08	66.67
69	emotions, emotional states, facial expressions, emotional responses, emotional regulation	0.08	61.06
22	memory, human information storage, retention, cognitive processes, recognition (learning)	0.07	47.46
47	age differences, short-term memory, cognitive ability, aging, cognitive development	0.06	62.60
51	academic achievement, student characteristics, college students, high school students, elementary school students	0.03	51.76
96	psychotherapeutic processes, psychotherapeutic outcomes, client-centered therapy, psychotherapeutic techniques, psychoanalysis	0.03	49.27
86	human migration, cross-cultural differences, acculturation, immigration, cross-cultural communication	0.00	51.92

88	attribution, self-concept, emotional states, internal external locus of control, achievement	0.00	45.11
60	teacher-student interaction, teacher attitudes, teacher characteristics, teachers, teacher education	-0.01	50.66
99	occupational stress, stress management, coping behavior, working conditions, health	-0.01	60.80
84	social perception, ingroup outgroup, social identity, stereotyped attitudes, attribution	-0.03	54.17
2	morality, social cognition, justice, attribution, motivation	-0.10	47.83
13	criminals, juvenile delinquency, crime, prisoners, forensic psychiatry	-0.10	49.67
67	language, semantics, syntax, cognitive processes, evoked potentials	-0.10	51.38
42	brain, physiological correlates, functional magnetic resonance imaging, neurophysiology, evoked potentials	-0.11	61.21
45	learning, computer-assisted instruction, teaching methods, learning strategies, instructional media	-0.12	55.00
21	verbal communication, nonverbal communication, interpersonal communication, language, interpersonal interaction	-0.13	44.44
33	questionnaires, test validity, test construction, personality measures, test reliability	-0.13	52.11
7	physiological correlates, heart rate, psychophysiology, physiological arousal, anxiety	-0.15	47.09
73	environmental attitudes, conservation (ecological behavior), urban environments, home environment, architecture	-0.16	44.34
10	psychoanalytic theory, history of psychology, psychoanalysis, psychoanalytic interpretation, theories	-0.19	39.83
55	goals, motivation, self-regulation, intention, behavior	-0.20	57.98
71	management personnel, human resource management, business organizations, business management, leadership	-0.24	51.30
11	electroencephalography, electrical activity, evoked potentials, brain, physiological correlates	-0.27	50.89
5	group dynamics, group performance, interpersonal interaction, social interaction, group cohesion	-0.28	37.96
52	sports, athletes, athletic performance, athletic participation, personality correlates	-0.33	32.88
35	cognitive processes, human information storage, attention, human channel capacity, reaction time	-0.37	48.54
89	personnel selection, aptitude measures, test validity, statistical validity, test construction	-0.47	39.24
76	test validity, intelligence measures, test construction, measurement, testing	-0.61	36.78
94	decision making, cognitive processes, judgment, choice behavior, uncertainty	-0.65	37.67

72	driving behavior, drivers, highway safety, automobiles, motor traffic accidents	-0.69	34.03
9	cognitive processes, problem solving, models, theory formulation, knowledge level	-0.76	32.00
24	methodology, statistical analysis, measurement, models, mathematical modeling	-1.01	23.12
95	theories, philosophies, history of psychology, theory formulation, psychology	-1.06	25.31

Anmerkung: *D = Prävalenzdifferenz (weiblich - männlich)

Tabelle 3. Übersicht über die Themen – sortiert nach Frauenanteil (absteigend).

Thema	Wahrscheinlichste Begriffe	D* (%)	Frauenanteil (%)
27	human females, sex roles, human males, sex role attitudes, human sex differences	0.99	81.90
70	obesity, eating disorders, bulimia, human females, anorexia nervosa	0.51	81.44
29	prevention, social skills, evaluation, social skills training, aggressive behavior	0.24	78.57
78	mother-child relations, attachment behavior, infant development, mothers, mother-child communication	0.72	74.32
97	posttraumatic stress disorder, emotional trauma, coping behavior, trauma, victimization	0.27	73.73
83	parent-child relations, family relations, childrearing practices, mothers, childhood development	0.91	72.50
91	chronic pain, pain perception, pain management, pain, back pain	0.18	71.08
34	coping behavior, quality of life, neoplasms, client characteristics, well being	0.48	68.90
50	social perception, attribution, judgment, stereotyped attitudes, impression formation	0.48	67.59
66	health behavior, health promotion, health, lifestyle, physical activity	0.16	67.39
37	hydrocortisone, stress reactions, stress, physiological correlates, psychological stress	0.08	66.67
38	cognitive impairment, cognitive ability, neuropsychological assessment, executive function, schizophrenia	0.34	66.50
98	marital relations, couples, marital satisfaction, spouses, relationship quality	0.26	64.71
47	age differences, short-term memory, cognitive ability, aging, cognitive development	0.06	62.60
42	brain, physiological correlates, functional magnetic resonance imaging, neurophysiology, evoked potentials	-0.11	61.21
69	emotions, emotional states, facial expressions, emotional responses, emotional regulation	0.08	61.06
99	occupational stress, stress management, coping behavior, working conditions, health	-0.01	60.80
19	coping behavior, stress management, stress, personality traits, anxiety	0.17	59.63
55	goals, motivation, self-regulation, intention, behavior	-0.20	57.98
45	learning, computer-assisted instruction, teaching methods, learning strategies, instructional media	-0.12	55.00
84	social perception, ingroup outgroup, social identity, stereotyped attitudes, attribution	-0.03	54.17
33	questionnaires, test validity, test construction, personality measures, test reliability	-0.13	52.11

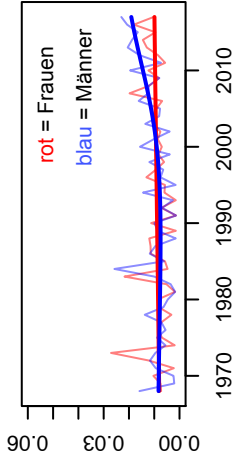
86	human migration, cross-cultural differences, acculturation, immigration, cross-cultural communication	0.00	51.92
51	academic achievement, student characteristics, college students, high school students, elementary school students	0.03	51.76
67	language, semantics, syntax, cognitive processes, evoked potentials	-0.10	51.38
71	management personnel, human resource management, business organizations, business management, leadership	-0.24	51.30
11	electroencephalography, electrical activity, evoked potentials, brain, physiological correlates	-0.27	50.89
60	teacher-student interaction, teacher attitudes, teacher characteristics, teachers, teacher education	-0.01	50.66
46	personality traits, extraversion, neuroticism, personality correlates, individual differences	0.13	50.00
13	criminals, juvenile delinquency, crime, prisoners, forensic psychiatry	-0.10	49.67
96	psychotherapeutic processes, psychotherapeutic outcomes, client-centered therapy, psychotherapeutic techniques, psychoanalysis	0.03	49.27
35	cognitive processes, human information storage, attention, human channel capacity, reaction time	-0.37	48.54
2	morality, social cognition, justice, attribution, motivation	-0.10	47.83
22	memory, human information storage, retention, cognitive processes, recognition (learning)	0.07	47.46
7	physiological correlates, heart rate, psychophysiology, physiological arousal, anxiety	-0.15	47.09
88	attribution, self-concept, emotional states, internal external locus of control, achievement	0.00	45.11
21	verbal communication, nonverbal communication, interpersonal communication, language, interpersonal interaction	-0.13	44.44
73	environmental attitudes, conservation (ecological behavior), urban environments, home environment, architecture	-0.16	44.34
10	psychoanalytic theory, history of psychology, psychoanalysis, psychoanalytic interpretation, theories	-0.19	39.83
89	personnel selection, aptitude measures, test validity, statistical validity, test construction	-0.47	39.24
5	group dynamics, group performance, interpersonal interaction, social interaction, group cohesion	-0.28	37.96
94	decision making, cognitive processes, judgment, choice behavior, uncertainty	-0.65	37.67
76	test validity, intelligence measures, test construction, measurement, testing	-0.61	36.78
72	driving behavior, drivers, highway safety, automobiles, motor traffic accidents	-0.69	34.03
52	sports, athletes, athletic performance, athletic participation, personality	-0.33	32.88

	correlates		
9	cognitive processes, problem solving, models, theory formulation, knowledge level	-0.76	32.00
95	theories, philosophies, history of psychology, theory formulation, psychology	-1.06	25.31
24	methodology, statistical analysis, measurement, models, mathematical modeling	-1.01	23.12

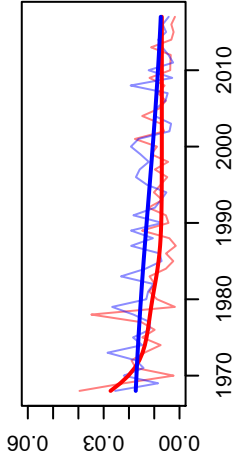
Anmerkung: **D* = Prävalenzdifferenz (weiblich - männlich)

PsychArchives-ESM 5: Zeitliche Verläufe aller Themen.

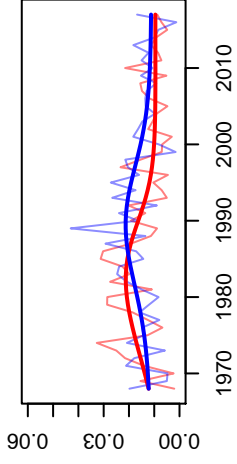
2 morality, social cognition



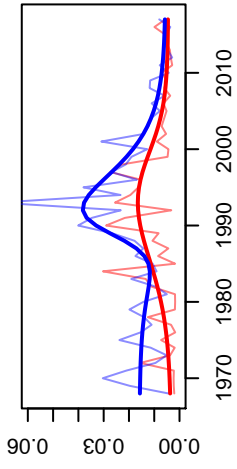
5 group dynamics, group performance



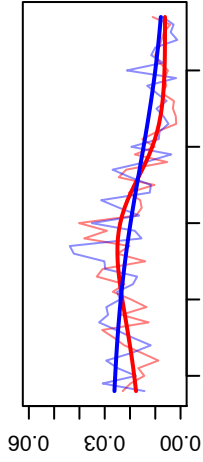
7 physiological correlates, heart rate



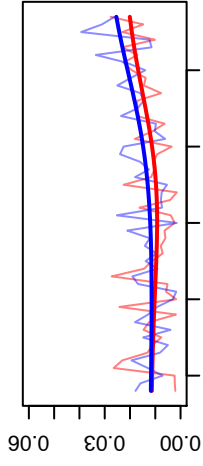
9 cognitive processes, problem solving



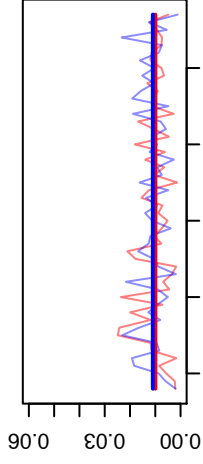
10 psychoanalytic theory, history of psychology



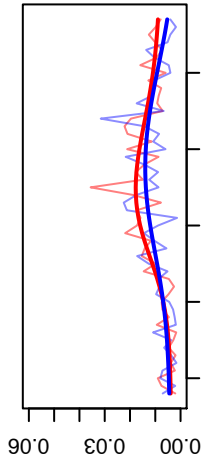
11 electroencephalography, electrical activity



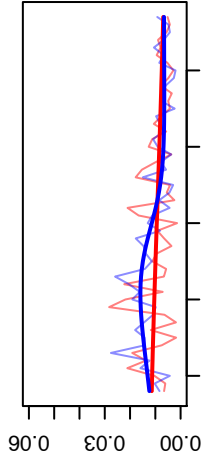
13 criminals, juvenile delinquency



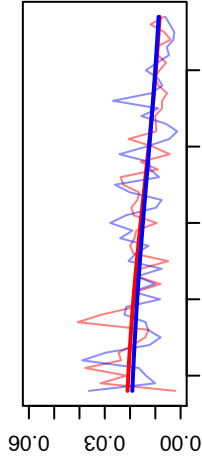
19 coping behavior, stress management



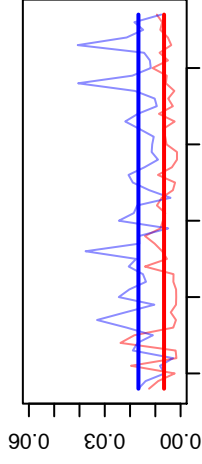
21 verbal communication, nonverbal communication



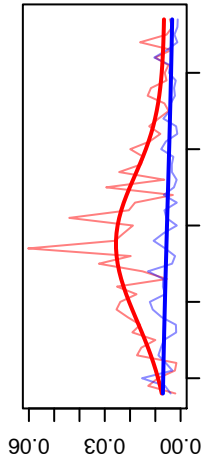
22 memory, human information storage



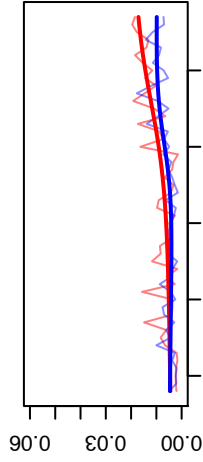
24 methodology, statistical analysis



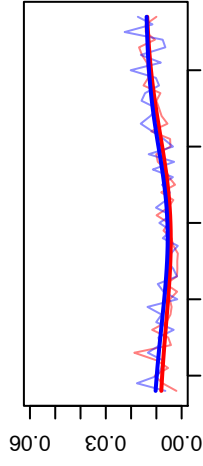
27 human females, sex roles



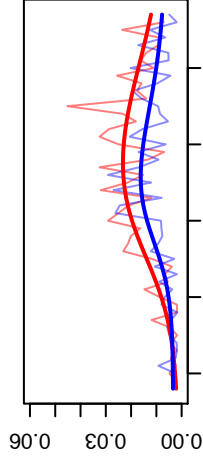
29 prevention, social skills



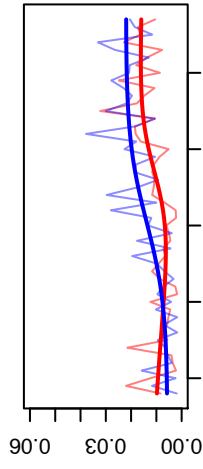
33 questionnaires, test validity



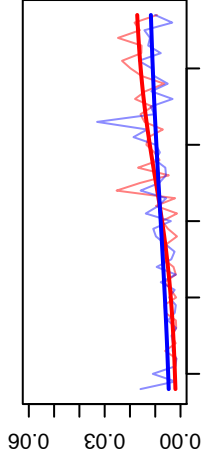
34 coping behavior, quality of life



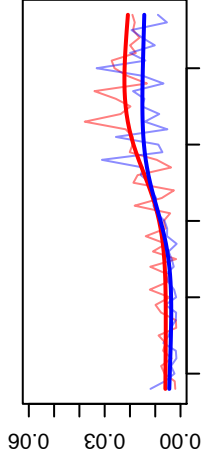
35 cognitive processes, human information storage



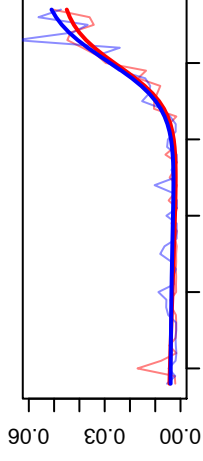
37 hydrocortisone, stress reactions



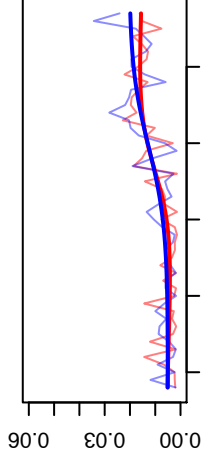
38 cognitive impairment, cognitive ability



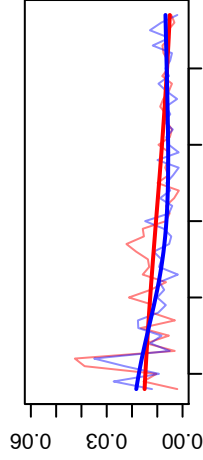
42 brain, physiological correlates



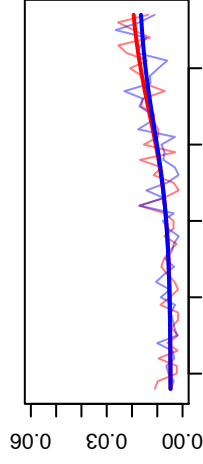
45 learning, computer assisted instruction



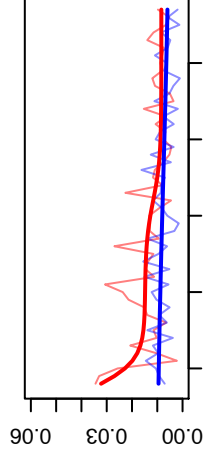
46 personality traits, extraversion



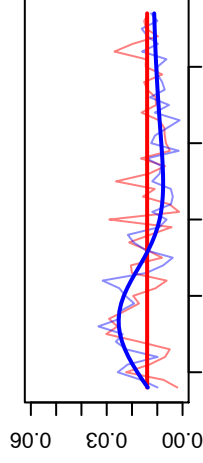
47 age differences, short term memory



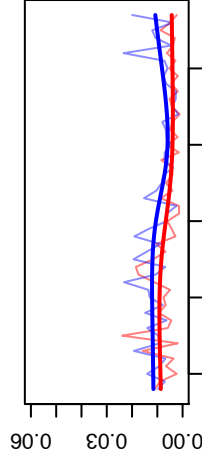
50 social perception, attribution



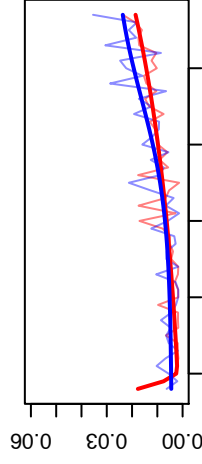
51 academic achievement, student characteristics



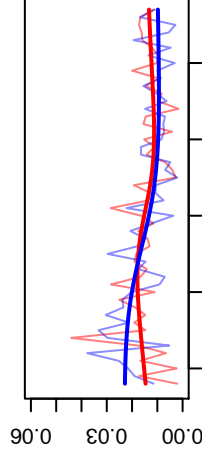
52 sports, athletes



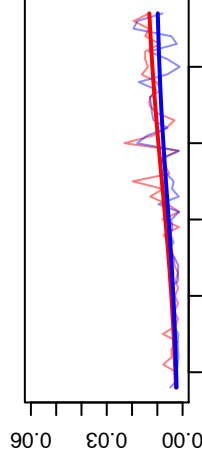
55 goals, motivation



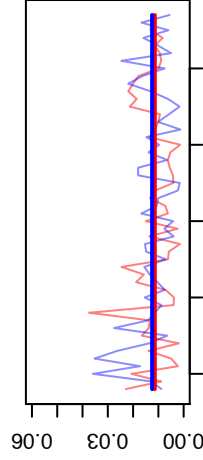
60 teacher student interaction, teacher attitudes



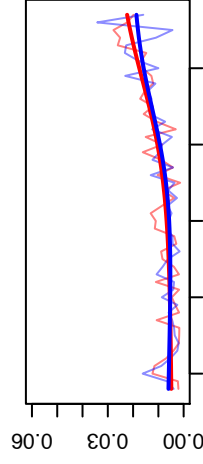
66 health behavior, health promotion



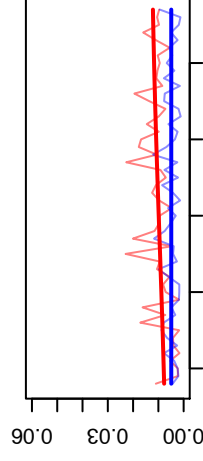
67 language, semantics



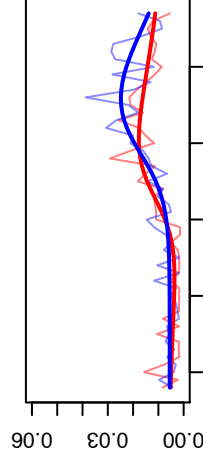
69 emotions, emotional states



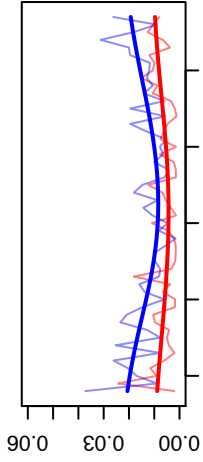
70 obesity, eating disorders



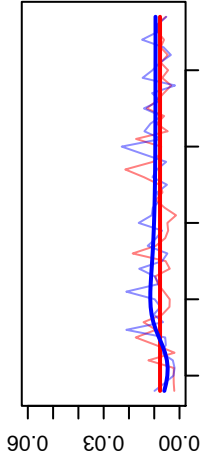
71 management personnel, human resource management



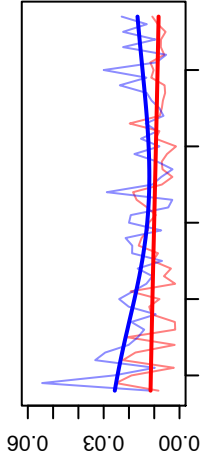
72 driving behavior, drivers



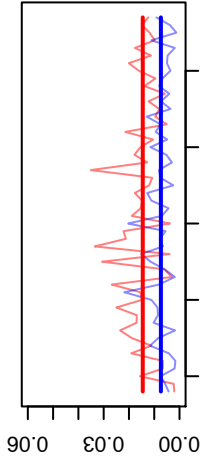
3 environmental attitudes, conservation (ecological beh



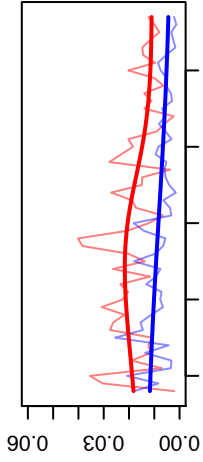
76 test validity, intelligence measures



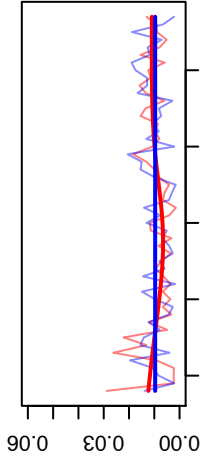
78 mother child relations, attachment behavior



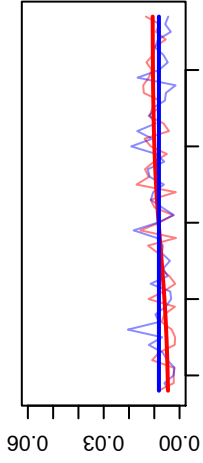
83 parent child relations, family relations



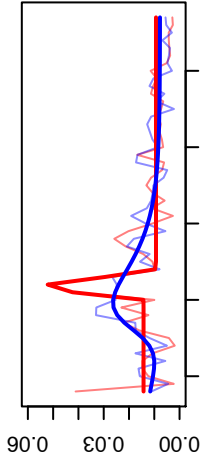
84 social perception, ingroup outgroup



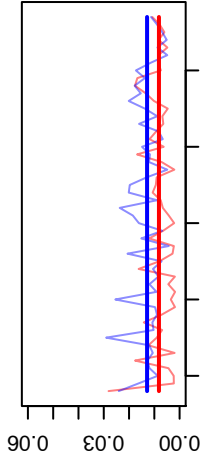
86 human migration, cross cultural differences



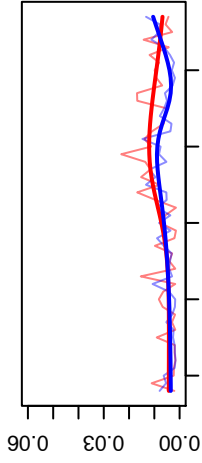
88 attribution, self concept



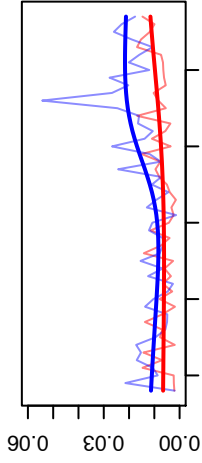
89 personnel selection, aptitude measures



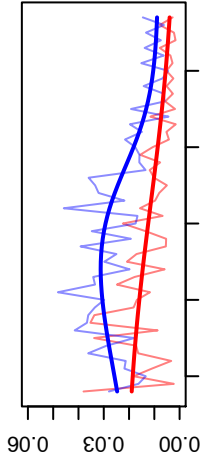
91 chronic pain, pain perception



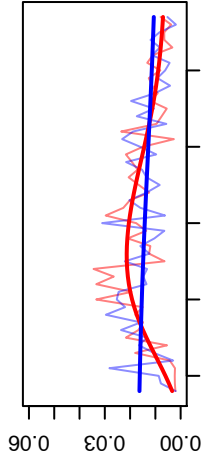
94 decision making, cognitive processes



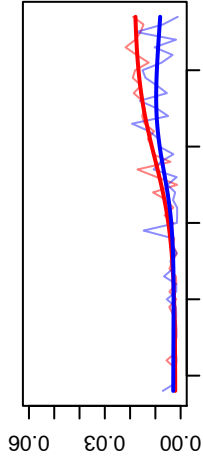
95 theories, philosophies



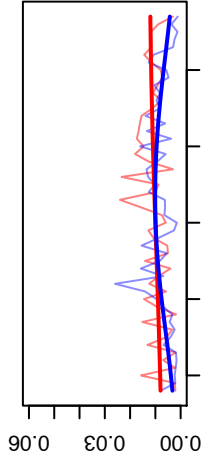
6 psychotherapeutic processes, psychotherapeutic out



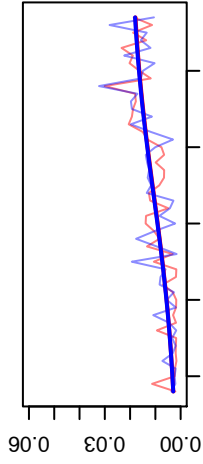
97 posttraumatic stress disorder, emotional trauma



98 marital relations, couples



99 occupational stress, stress management



Anhang E: Originalpublikation zu Abschnitt 7

Development of a user-friendly app for exploring and analyzing research topics in psychology

André Bittermann

abi@leibniz-psychology.org

ZPID – Leibniz Institute for Psychology Information, Universitätsring 15, D-54296 Trier (Germany)

Background

Keeping track of the developments in a scientific field can be challenging. Regarding increasing numbers of publications, summarizing the contents of hundreds of thousands of scientific publications on specific topics is necessary to gain insights into the processes of a scientific field.

Many databases offer classifications, i.e., broad subject headings for categorizing the publications' contents. Past research has treated these categories as research topics (e.g., Krampen, 2016), but regarding the level of detail, topicality, and flexibility this approach has been criticized (e.g., Bittermann & Fischer, 2018).

Techniques for automated content analysis represent a promising approach for getting insight into large text corpora. Topic modeling (e.g., Blei, Ng, & Jordan, 2003), in particular, is gaining in popularity in scientometrics. In their well-known paper, Griffiths and Steyvers (2004) demonstrated how to find scientific topics by applying topic models to a corpus of scientific abstracts.

A topic-guided and user-friendly interface for databases of scientific literature can open publication trends to a broader audience with various user scenarios: exploring the current “hot topics,” investigating the ups and downs of topic popularity over time, or comparing publication trends concerning societal processes (e.g., the increasing trend of a topic referring to refugees and emotional trauma in psychological publications from the German-speaking countries after 2015).

Aim

The goal of this project was to develop a user-friendly web-based application for exploring and analyzing research topics in psychology. This app is considered as an entry point to further research of scientific literature by informing the user about past and current developments of publication topics. To this end, the topics are directly linked to search queries in a database for psychological literature.

Method

Data

The psychological research topics were derived from PSYINDEX – the comprehensive database

containing references for German- and English-language publications in psychology and closely related disciplines from the German-speaking countries. It is developed and hosted by the Leibniz Institute for Psychology Information (ZPID; Trier, Germany). In April 2019, there were more than 350,000 psychological articles, book chapters, reports, and dissertations indexed in PSYINDEX. In the development of the app, documents published between 1980 and 2017 were included ($N = 329,240$ in early 2019).

The PSYINDEX editorial staff assigns controlled terms from the *Thesaurus of Psychological Index Terms* published by the American Psychological Association (Tuleya, 2007). This standardized vocabulary of keywords the input for topic modeling. Main advantages compared to abstract texts are, inter alia, the direct usability for efficient literature search for this topic, the avoidance of stemming, stop words, and synonyms, as well as faster computation time (Bittermann & Fischer, 2018).

Software

All analyses were conducted in R version 3.5.1 (R Core Team, 2018). For inference of research topics, the package *topicmodels* 0.2-8 (Grün & Hornik, 2011) was employed. The user interface was built as a *Shiny app* using *shiny* 1.2.0 (Chang, Cheng, Allaire, Xie, & McPherson).

Topic Modeling

Topic modeling based on *latent Dirichlet allocation* (Blei et al., 2003) was applied. Following the best-practice recommendations by Maier et al. (2018), several candidates for the alpha hyperparameter (0.0001, 0.0005, 0.001) and the number of topics k (250–550) were examined. Delta was fixed to 0.01. The final model (alpha = 0.0005 and $k = 325$) was selected regarding interpretability and document–topic assignments. Finally, only reliable topics with stability across multiple inference runs were included to increase the robustness of the results.

Results

Topic Model

The final model comprised 213 topics. The five terms with highest probabilities were included in the app. The topic with the highest prevalence overall was “psychoanalysis, psychotherapeutic processes, psychotherapeutic transference, counter-transference, psychoanalytic theory.” The most strongly increasing trend over the whole range of years was shown by the topic “functional magnetic resonance imaging, cerebral blood flow, brain, prefrontal cortex, neuroanatomy.” The “hottest” topic during the last three years was “posttraumatic stress disorder, emotional trauma, refugees, trauma, war.”

Features of the App

Users can set the range of years from which popularity and trends are dynamically calculated. For exploring the topics, they can switch between “popular topics,” “hot topics” (see Fig. 1), “cold topics,” and an overview of “all topics.” Each topic entry has a search button that forwards a search query to PSYNDEX for literature relevant to this topic. For optimizing results, the terms in this query are weighted according to the term probability in the topic model.

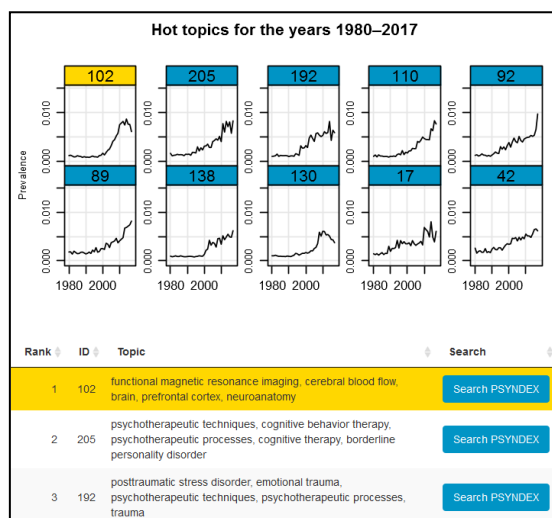


Figure 1. Hot topic view showing topics with most strongly increasing trends from 1980–2017.

A demo version of the app can be accessed for free via <https://abitter.shinyapps.io/psychtopics/>

Conclusions and Future Developments

Initial user experiences confirm the app’s ease of use. The implemented search queries help to clarify the topics’ contents and offer a low-threshold starting point to literature search. Topic inference is data-driven and independent from prior knowledge about a database’s contents. Since the standardized vocabulary used in PSYNDEX is updated on a

regular basis, the topic model can be updated as well and kept up to date with low maintenance efforts.

Future features will include current developments from our research group. For instance, topics with a high/low degree of empirical evidence can be shown, which may be of interest for research synthesis or explorative research. Using forecasting techniques, the observed trends can be compared to expected courses over time and help to quantify sudden increases and decreases in publication numbers. Author information could be included for investigating topical author networks.

The app can be applied to other databases with only few modifications necessary. The code is available for free on PsychArchives (<http://dx.doi.org/10.23668/psycharchives.2410>).

References

- Bittermann, A. & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, 226(1), 3–13. <https://doi.org/10.1027/2151-2604/a000318>
- Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1162/jmlr.2003.3.4-5.993>
- Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2018). shiny: Web Application Framework for R. R package version 1.2.0 [Computer software].
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl. 1), 5228–5235. <https://doi.org/10.1073/pnas.0307752101>
- Grün, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13). <https://doi.org/10.18637/jss.v040.i13>
- Krampen, G. (2016). Scientometric trend analyses of publications on the history of psychology: Is psychology becoming an unhistorical science? *Scientometrics*, 106, 1217–1238. <https://doi.org/10.1007/s11192-016-1834-4>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... & Schmid-Petri, H. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2–3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [Computer software].
- Tuleya L. G. (Ed.). (2007). *Thesaurus of psychological index terms* (11th ed.). Washington, DC: American Psychological Association.

Elektronische Supplementmaterialien (ESM)

ESM zu Studie 1: R-Code der Analysen und Tabelle aller Themen.

<https://doi.org/10.1027/2151-2604/a000318>

ESM zu Studie 2: R-Code der Analysen und Liste der stopwords.

<http://dx.doi.org/10.23668/psycharchives.2499>

ESM zu Studie 2: Eine Liste mit den Nachweisnummern der zehn repräsentativsten Publikationen je Thema.

<http://dx.doi.org/10.23668/psycharchives.2498>

PsychArchives-ESM 2 zu Studie 3: R-Code der Analysen.

<http://dx.doi.org/10.23668/psycharchives.2677>

ESM zu Abschnitt 7: R-Code für PsychTopics.

<http://dx.doi.org/10.23668/psycharchives.2410>

Eidesstattliche Erklärung

Hiermit erkläre ich, dass die vorliegende Dissertationsschrift von mir eigenständig angefertigt wurde und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet wurden. Wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte wurden als solche kenntlich gemacht. Die Arbeit oder Teile davon wurden bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Trier, 11.02.2020

André Bittermann