

Noah Bubenhofer / Daniel Knuchel / Livia Sutter / Maaïke Kellenberger / Niclas Bodenmann

Von Grenzen und Welten: Eine korpuspragmatische COVID-19-Diskursanalyse

1 Einleitung

Am 13. März, rund vier Wochen nach Semesterstart, beschließt die Schweizer Regierung (der Bundesrat) den „Lockdown“ des Landes angesichts der fortschreitenden COVID-19-Pandemie. Die Universität schließt ihre Türen, die Büros müssen verlassen werden. Das weitere Leben findet für viele „zuhause“ statt und jegliche Interaktion, die den engsten Kreis überspannt, ist medial vermittelt. Man scrollt sich durch Newsticker und verfolgt Medienkonferenzen des Bundesrats komplett, um hinterher auch noch die mediale Berichterstattung darüber zu lesen.

Aus linguistischer Sicht wird deutlich: Da sind auf der einen Seite das Virus, die Kranken und Toten, auf der anderen Seite deren kommunikative Vermittlung, die dieses Virus und alle Handlungen um es herum soziosemiotisch formen und prägen. Dabei finden sich sehr unterschiedliche Prägungen, wie bei Leser*innen-Kommentaren zu Newstexten deutlich wird:

- 1) Das Virus ist ein kluger Zug der Chinesen gewesen. (13.04.2020, 09:50 Uhr, Kommentar auf Blick.ch)
- 2) Dieser Virus ist sehr gefährlich!!!! Bleibt zu Hause!!!! Aufwachen!!!! (10.04.2020, 01:54 Uhr, Kommentar auf 20min.ch)
- 3) Aber das Virus ist republikanisch und wird sich mit seinem geliebten Führer vereinen. (22.05.2020, 06:54, Kommentar auf 20min.ch)
- 4) Ohhh wie schlimm dieser Virus ist, ich mach mir in die Hosen... :-)) (11.04.2020, 19:57, Kommentar auf 20min.ch)

Die private Ohnmacht vor der ungewohnten Situation und den entfesselten diskursiven Kräften weicht schnell dem professionellen Interesse an den sprachlichen Aspekten der Krise. Damit beginnen das Programmieren, Sammeln und nach und nach Analysieren von Daten, um den persönlichen Eindrücken empirisch nachgehen zu können.

2 Daten

Wir sind nicht allein: Das Sammeln von unterschiedlichen Daten über die Pandemie wird allgegenwärtig und sie erscheinen in einer Vielzahl von Dia-

grammen in verschiedenen Medien. Unser Interesse gilt jedoch den *sprachlichen* Daten: Ausgehend von der Hypothese, dass die Pandemie ebenso sehr kommunikativ wie biologisch geprägt ist, interessieren wir uns für die mediale Berichterstattung in der Deutschschweiz und die Reaktionen der Leser*innen darauf. Bis zum 29. Mai 2020 entstehen so zwei Korpora von Newsartikeln zum Thema (Suchstichwort „Corona“ bzw. entsprechende Rubrik) und den dazugehörigen Kommentaren von fünf wichtigen Newsportalen der deutschsprachigen Schweiz, die ohne Bezahl- oder Registrierungsschranke konsumierbar sind (CovArtikel und CovKommentare, Tabelle 1).

	Artikel (CovArtikel)			Kommentare (CovKommentare)		
	Texte	Tokens	Tokens/Text	Texte	Tokens	Tokens/Text
20min.ch	2081	3.041.009	1461	346.095	11.229.956	32
Blick.ch	2087	1.080.553	518	131.347	5.226.056	40
Nau.ch	2576	954.398	370	21.978	673.298	31
SRF.ch	1527	878.101	575	35.642	1.828.596	51
Watson.ch	1011	783.291	775	60.611	2.715.959	45
Total	9282	6.737.352	726	595.673	21.673.865	36

Tabelle 1: Korpuszusammensetzungen CovArtikel und CovKommentare

Gut 9000 Artikel mit 6,7 Mio. Wörtern führen also zu knapp über 60 Mal mehr Kommentaren mit 21,6 Mio. Wörtern. Damit ist nun eine empirische Basis gelegt, um verschiedenen Fragen nachzugehen. Zum gegenwärtigen Stand (Juni 2020) sind unsere Analysen noch nicht abgeschlossen, deshalb berichten wir an dieser Stelle impressionistisch von einigen ersten Einsichten.

3 Semantische Räume: Grenzen

Methoden der distributionellen Semantik ermöglichen es, die diskursiv geprägten semantischen Räume in den Daten zu identifizieren und machen sie einer qualitativen Analyse zugänglich. Mit dem Ansatz „word2vec“ werden die typischen Kontexte der Ausdrücke im Korpus mit einem neuronalen Lernsystem modelliert und in einen Vektorraum überführt (Mikolov et al. 2013; Lenci 2018). Daraus entsteht dann (vereinfacht ausgedrückt) eine Punktwolke: Die Punkte repräsentieren alle Ausdrücke (Types) im Korpus, und die Nähe der Punkte zueinander zeigt an, wie ähnlich ihre Kontexte sind. Je ähnlicher ihre Kontexte, desto näher stehen sie semantisch (seien das synonymische, antonymische, hypo-/hyperonymische o.a. Relationen).

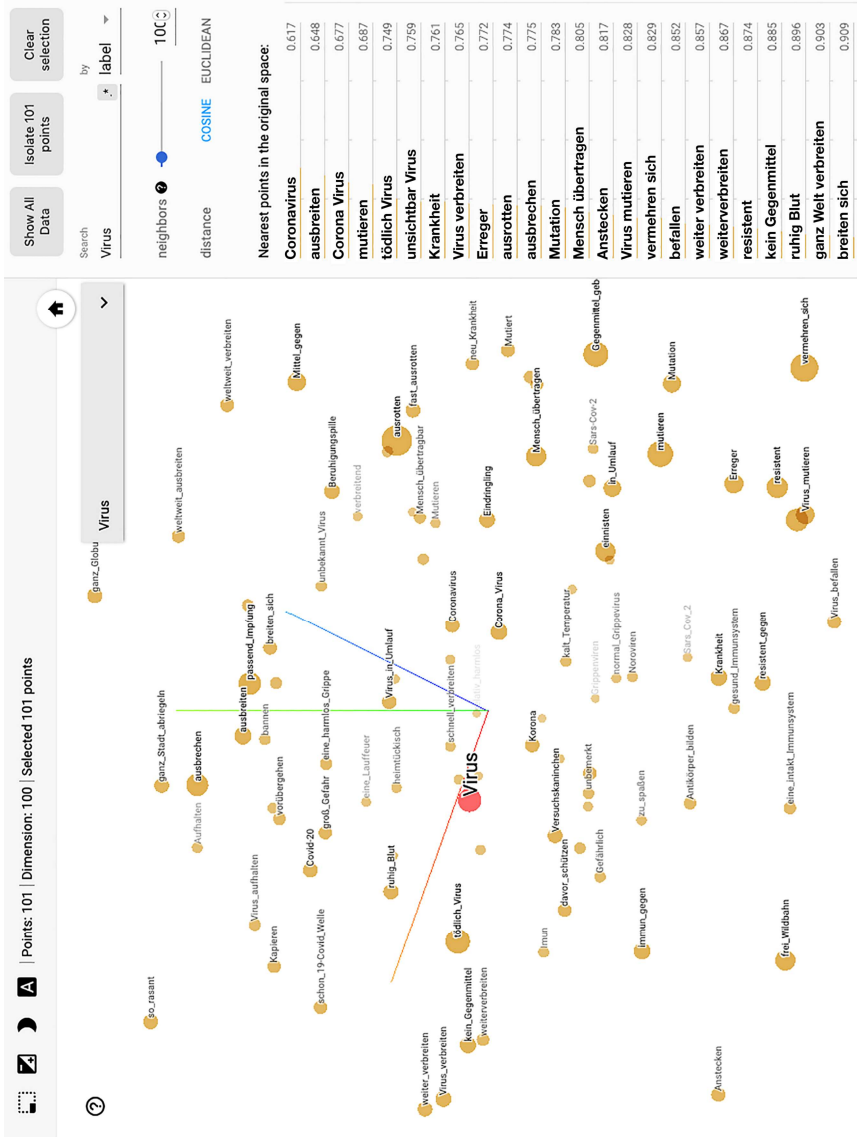


Abbildung 1: Word Embedding CovKommentare-Korpus mit den nächsten Nachbarn von „Virus“

Wir sehen in dieser Methode die Möglichkeit, eine gebrauchsssemantische, diskurslinguistische Perspektive einzunehmen und anstelle klassischer semantischer Relationen eher diskursiv geprägte „funktionale Äquivalenz“ zu modellieren (Bubenhof im Druck; Bubenhof/Calleri/Dreesen 2019).

Abbildung 1 zeigt die „nächsten Nachbarn“ zum Ausdruck *Virus*, also diejenigen Ausdrücke, die im CovKommentare-Korpus in ähnlichen Kontexten erscheinen. Die Liste auf der rechten Seite zeigt die Ausdrücke geordnet nach Distanz zu *Virus*. *Coronavirus* ist demnach der Ausdruck, der *Virus* am ähnlichsten ist; es folgen *ausbreiten*, *Corona Virus*, *mutieren*, *tödliches/r Virus*, *unsichtbares/r Virus*, *Krankheit* etc. (die Daten sind lemmatisiert, so dass immer die Grundformen sichtbar sind).

Je nach Datengrundlage zeigen sich anders geprägte semantische Räume und damit andere nächste Nachbarn zu *Virus*. Z.B. finden wir in einem Gesundheitsforen-Korpus (symptome.ch) als nächste Nachbarn zu *Virus*: *Protozoen*, *Herpesvirus*, *Bakterium*, *Erreger*, *Parasit*, *Grippevirus*, *parasitär*, *pathogen* (usw.); in einem Zeitungskorpus (Spiegel, Bild und Zeit von 2013-2018) sind es hingegen: *infizieren*, *Erreger*, *Infektion*, *infiziert*, *Ebola-Virus*, *Grippevirus*.

Im Unterschied zu Diskursen in der Zeit vor Corona steht *Virus* im COVID-19-Diskurs in semantischen Zusammenhängen von Tod, Mutation, Ausrottung, Unsichtbarkeit, Vermehrung, Verbreitung und Globalität. Das *Virus* ist *unbekannt*, *heimtückisch*, *verbreitet* sich *unbemerkt* und wie *ein Lauffeuer* und lässt nicht mit sich *spassen*. Während der Corona-Krise ist der semantische Raum um *Virus* also deutlich dichter als beispielsweise in der Zeit von 2013 bis 2018, in der kein bestimmtes *Virus* so prägend war wie jetzt.

Die Exploration des Vektorraums offenbart nun einige wichtige Aspekte des Diskurses – vor allem auch in Differenz zwischen den Kommentaren und den Artikeln. Ein wichtiger Topos des Diskurses ist die *Grenze*. Da eine Epidemie ein Phänomen der Ausbreitung darstellt, mag es eine erwartbare Reaktion sein, sich der Ausbreitung zu entziehen zu versuchen. Nachdem im südlichen Nachbarland der Schweiz, Italien – und zwar: Norditalien! –, die Zahl der Erkrankungen stark zunahm, wurde das Schließen der Südgrenzen zu einem wichtigen Thema. Das semantische Feld sieht dabei so aus: *Grenze – schließen*, *Grenzgänger*, *Grenzschließung*, *Gotthard*, *Grenze schließen*, *nach Italien*, *Grenze öffnen*, *einreisen*, *Landesgrenze*, *Schengen*, *ausreisen*, *Grenzöffnung*, *zumachen*, *dicht machen*, *Riegel vorschieben*, *offen lassen*, *Deutschschweiz*, *reinlassen*, *Grenze dicht*, *Brenner*, *Nachbarland* etc. Vieles ist nicht überraschend – neben der Grenzschließung wird natürlich auch die erneute Öffnung diskutiert –, bemerkenswert sind jedoch *Gotthard* und *Deutschschweiz*: Das Erscheinen dieser Toponyme lässt erahnen, dass nicht nur eine Schließung der Landesgrenze diskutiert, sondern die Losung „Gotthard zu!“ thematisiert wird:

Schliesst den Gotthard, deklariert das Tessin als rote Zone. Subito Grenze schliessen. Jetzt. Mehr muss nicht gesagt werden. (10.03.2020, 18:50 Uhr, Kommentar auf 20min.ch)

Das südlich des Gotthards liegende, italienischsprachige Tessin soll also von der Deutschschweiz abgeriegelt werden, um eine weitere Verbreitung des Virus zu verhindern. Die Alpen markieren damit eine zusätzliche diskursive (und auch sprachbedingte) Grenze innerhalb eines Landes, das sich ansonsten als mehrsprachige „Willensnation“ geriert.

Ein paar Wochen später, um Ostern herum, ist der Gotthard erneut Thema einer Schließungsdebatte. Allerdings geht es nun nicht nur um das ‚Nicht-Einlassen‘ des Virus in die Deutschschweiz, sondern um die Bedrohung des Tessins durch die Deutschschweiz; deren Bewohner*innen besuchen an den Feiertagen bekanntermaßen gern das Tessin:

Ganz einfach, alle Pässe und die Gotthard Röhre schliessen und abriegeln, mal schauen wie dämlich diese Egoisten aus der Wäsche schauen. (07.04.2020, 10:28 Uhr, Kommentar auf 20min.ch)

Die Befürchtung vieler Kommentarverfasser*innen war also, dass die Appelle der Regierung, über Ostern auf die traditionelle Tessinfahrt zu verzichten, nicht fruchten würden (eine unbegründete Befürchtung, wie sich später herausstellen sollte).

Im Vergleich mit dem Word-Embedding-Modell der Artikel (CovArtikel-Korpus) zeigt sich zudem, dass in den Kommentaren eine Vielfalt von Ausdrücken zur Tätigkeit der Grenzschießung vorkommt – *zumachen, dicht machen, Riegel vorschieben* etc. – in Kombination mit Kritik an dem angeblich nachlässigen Verhalten der Regierung: *absolut verantwortungslos, hätte man, lange offen*.

Anders in den Artikeln: Dort gibt es einerseits Ausdrücke, die Unannehmlichkeiten der Grenzschießung thematisieren (der für ein Binnenland unerreichbare *Strand, Verwandte besuchen wollen, eine grenzüberschreitende Beziehung, Reise, Zweitwohnung*). Andererseits beschränken sich Ausdrücke für die Tätigkeit der Grenzschießung auf *Grenze schliessen* und dann aber auf die Wiedereröffnung mit *lockern, wieder offen, lockern werden, Lockerungen* etc.

Die Daten lassen die Hypothese formulieren, dass die Grenzschießung ein Thema ist, das primär für die News-Kommentierenden wichtig ist, weit weniger aber für die Medienschaffenden. Dazu muss man berücksichtigen, dass die rechtskonservative Schweizerische Volkspartei SVP im Zusammenhang von Migration und Europapolitik die Grenze seit Jahrzehnten als angebliche Garantin für Selbstbestimmung und Schutz bewirtschaftete, was sich in den Kommentaren widerspiegelt. So scheint, dass die Grenzschießung vor allem einen symbolischen Akt darstellte, während aus epidemiologischer Perspektive eher von Mobilitätsbeschränkung die Rede ist. Gene-

relle Mobilitätsbeschränkungen müssen jedoch nicht immer mit geschlossenen Landesgrenzen einhergehen, denn inländische sind mancherorts viel intensiver als die grenzüberschreitenden Relationen. Oder landesübergreifende Wirtschaftsräume wie etwa das Tessin, Genf oder Basel sind gerade im Gesundheitssystem hochgradig abhängig von den Grenzgänger*innen.

Wir müssen es an dieser Stelle bei diesen Explorationen des semantischen Raumes belassen – die kurzen Streifzüge zeigen, wie datengeleitet Diskursstrukturen freigelegt werden können, und den meisten Hinweisen sind wir noch gar nicht nachgegangen.

4 Diskursdynamiken: Was ist die Welt?

Die Analysen der Diskursräume beschränkten sich auf eine statische Sicht: Die Räume ergeben sich aus einem bestimmten Datensample. Ein zweiter Ansatz versucht, die Dynamik des Diskurses besser zu erfassen und verwendet als Grundlage das sog. LDA-Topic-Modeling (Graham/Weingart/Milligan 2012; Blei/Ng/Jordan 2003). Es handelt sich dabei um ein Textklassifikationsverfahren, das ebenso datengeleitet vorgeht: Eine Menge von Texten wird aufgrund ihres Vokabulars bestmöglich in Cluster aufgeteilt, sodass sich die Texte eines Clusters bezüglich des verwendeten Vokabulars gleichen und gleichzeitig gegenüber anderen Clustern abgrenzen. Das spezifische Vokabular des jeweiligen Clusters bildet dann ein „Topic“. Wichtig ist hier zu betonen, dass grundsätzlich in jedem Text jedes Topic vorkommt, einfach unterschiedlich stark. Und jeder Ausdruck ist bis zu einem gewissen Grad Topic-spezifisch, jedoch ebenso unterschiedlich stark. Wir haben auf das CovKommentare-Korpus ein LDA-Topic-Modeling angewandt und den Algorithmus 20 „Topics“ identifizieren lassen. Die Liste in Tabelle 2 (s.u.) gibt die gefundenen „Topics“ mittels ihres spezifischsten Vokabulars wieder.

Die Übersicht offenbart aus linguistischer Sicht ein paar methodologische Probleme. Vor allem wirft sie die Frage auf, was denn eigentlich unter „Topic“ zu verstehen ist – wir glauben, die Gleichsetzung mit „Thema“, wie das in den Sozialwissenschaften oft geschieht, schlägt fehl. Trotzdem ist die Methode interessant, da sie neben Diskursthemen auch textsortenspezifische Handlungen (kommentieren, kritisieren, unterstellen, Statements abgeben etc.) aufgrund des Vokabulars identifiziert, wie unsere Interpretation der Topic-Cluster zeigt. Dies stellt eine interessante Erweiterung thematischer Schwerpunkte um *Diskurspraktiken* dar.

Nr.	Spezifisches Vokabular	Interpretation
0	eine haben von Bundesrat für dies zu Herr SVP BR unser sich mit Volk Politiker ihr gegen Frau Partei	Innenpolitik / Regierung
1	App mit the Datum Handy bei von eine of neu to Google all 5G Tracing and TV Internet Apple	Mobilfunk (5G-Verschwörungstheorie), App Contact Tracing
2	! ? was jetzt warum wer wo sollen doch denn wieso endlich noch wie oder alle genau Dank mal	Textsortenspezifik Kommentare: Fragen stellen, antworten
3	man nicht können eine wenn auch es haben werden dann aber so ja müssen oder dass was er machen	Besserwisserei I gesellschaftliche Perspektiven
4	ich haben mein auch aber für gut finden als kein machen so sehr dass da denken sehen mögen hoffen	Besserwisserei II individuelle Perspektiven
5	card bei Zahl von mit werden Tote eine Jahr Test Schweiz Tag testen Corona 000 als sterben Fall bis	Statistik, Verlauf der Epidemie
6	es nicht geben haben kein noch auch viel mehr zu aber werden so gehen nur immer da wie ja	Textsortenspezifik Kommentare: Statements
7	Virus eine Mensch werden Corona mit sterben sich dies Grippe von bei nicht oder als Angst zu alt es	Virus
8	zu nicht Kind bleiben müssen arbeiten mit gehen Haus auch sich sollen können halten Arbeit dürfen Abstand kein oder	Gesellschaftliche und individuelle Praktiken, Home-Office, Kinderbetreuung
9	Geld für werden eine von bezahlen müssen mit zahlen kein Firma Lohn sollen bei Staat Jahr dann dies Steuer	Wirtschaftliche Aspekte
10	ihr zu sich nicht sie mit nur dies alle um es was mal hier einfach Leute von machen so	Textsortenspezifik Kommentare: Kommentierung
11	eine ich mit nach von card Jahr dann werden noch fahren bis für Auto Tag vor bei Woche oder	Gesellschaftliche und individuelle Praktiken, Mobilität, Zukunft
12	Schweiz haben wir Land Grenze zu aus Schweizer Italien von EU schließen hier werden unser nach Deutschland wie für	Grenze
13	eine zu werden nicht dies dass für sich auch mit können es als von aber viel um sehr gut	Textsortenspezifik Kommentare: Kommentierung
14	... ja eine so aber dann mal gut mit Ja doch schon auch nur alle noch kommen da immer	Textsortenspezifik Kommentare: Kommentierung
15	eine von dies zu mit dass ... hier werden als Artikel Medium sich lesen über wie aus oder	Thematisierung Medien, Kommentierungen
16	Maske mit eine bei tragen kaufen ich Hand für kein auch zu oder schützen waschen Laden vor bestellen Post	Maske, Mund-Nasen-Schutz
17	wir werden haben zu noch alle wieder jetzt es eine unser kommen nicht dies müssen so Wirtschaft bis mehr	Zukunft, wirtschaftliche Entwicklung
18	er von haben China eine Trump USA mit als werden Welt Land aus WHO Präsident dies Chinese wie Europa	internationale Politik, China, Trump
19	sie Sie du ? haben was nicht ihr dein dann oder wenn Herr schreiben wie ich mal wissen meinen	Textsortenspezifik Kommentare: Kommentierung

Tabelle 2: Topic Modeling – jeweils die 19 spezifischsten Ausdrücke für die 20 Topics

Um einen Blick auf die Diskursdynamik zu ermöglichen, betrachten wir nun aber die zeitliche Verteilung der einzelnen Topics: Zu welchem Zeitpunkt finden wir besonders viele Kommentare, in denen die jeweiligen Topics stark ausgeprägt sind?

Abbildung 2 (s.u.) zeigt die Verläufe vier ausgewählter Topics, aggregiert nach Wochen. In Woche 11 (leicht links der Mitte) befindet sich der 13. März 2020: der Tag, an dem die ersten weitgehenden Maßnahmen des Schweizer Bundesrates kommuniziert worden sind.

Zunächst ein Blick auf das textsortenspezifische Topic „Fragen stellen, antworten“. Die meisten textsortenspezifischen Topics zeigen keine ausgeprägten Verwendungsveränderungen über die Zeit (mit Ausnahme der Topics 13 und 15, die gegen Ende zunehmen). Anders dieses Topic, das um Woche 11/12 (Mitte März) häufiger vorkommt und die Unsicherheit um den Start des Lockdowns zeigt. Dass ein Wort wie *endlich* (neben Frage- und Ausrufezeichen) dafür typisch ist, ist ein Indiz für den Streit in den Kommentaren darüber, ob dieser Schritt nun angemessen ist oder nicht.

Die drei stärker thematischen Cluster 7, 8 und 18 zeigen ganz unterschiedliche Dynamiken: Das Virus wird hauptsächlich in der Zeit weit vor dem 13. März diskutiert, danach ist es viel weniger Thema. Dafür sind es die praktischen Folgen im Alltag: Kinderbetreuung, Home-Office, Abstandsregeln. Cluster 18 (internationale Politik) zeigt einen verblüffenden Effekt des Lockdowns: Er verengt den Blick der Kommentierenden auf das Inland – internationale Ereignisse wie zu Beginn (Ausbruch der Epidemie in China) und am Ende des Beobachtungszeitraums (Trumps Reaktionen) bleiben wenig beachtet.

Damit, so unsere Hypothese, ergibt sich ein bemerkenswerter Widerspruch im Diskurs: Unsere Daten zeigen, dass der COVID-19-Diskurs zumindest in der Deutschschweiz sehr national geprägt ist und die Welt an den Grenzen der Schweiz endet. Auf der anderen Seite ist das Spezielle an der Corona-Pandemie ihr globaler Charakter (Jarren 2020): Praktisch jedes Land ist davon betroffen und die Reaktionen darauf sind sehr vielfältig. Dies wären ideale Voraussetzungen für einen global geprägten Diskurs, der aber so zumindest in der Deutschschweiz kaum stattfindet.

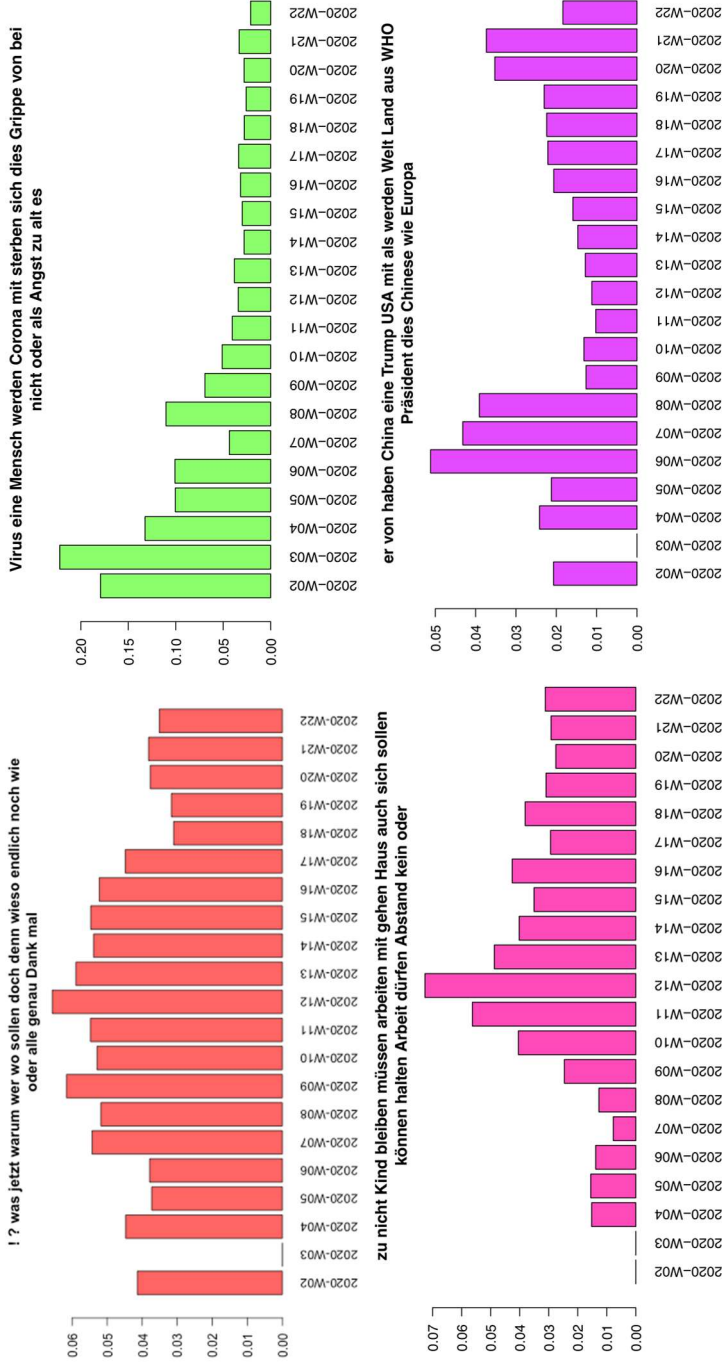


Abbildung 2: Dynamiken ausgewählter Topics (von links nach rechts und oben nach unten): Topic 2 (Textsortenspezifik Kommentare: Fragen stellen, antworten), Topic 7 (Virus), Topic 8 (Gesellschaftliche und individuelle Praktiken, Home-Office, Kinderbetreuung) und Topic 18 (internationale Politik, China, Trump).

5 Fazit

Die COVID-19-Krise zeigt deutlich, dass deren Bewältigung nicht nur medizinisch-epidemiologisch angegangen werden kann. Es sind diskursive Konstruktionen, die das gesellschaftliche Handeln („Meinen und Verstehen“, Feilke 1994) beeinflussen. Wie wird das Virus semantisiert? Welche Bedeutung hat die Landesgrenze für die Gesellschaft angesichts einer globalen Krise? Welche Horizonte werden im Diskurs überhaupt begünstigt: Nationale, internationale oder globale Horizonte? Der Erfolg von epidemiologischen Maßnahmen ist auch davon abhängig, welche diskursiven Prägungen der Krise wirkmächtig sind.

Wir haben versucht aufzuzeigen, wie die Analyse eines Diskurses datengeleitet angegangen werden kann und wo der Mehrwert für ein umfassendes Verständnis eines Krisendiskurses liegt. Die ersten impressionistischen Einblicke in die Analysen zeigen das Potenzial der Methoden.

Literatur

- Blei, David M./Ng, Andrew Y./Jordan, Michael I. (2003): Latent Dirichlet allocation. In: *Journal of Machine Learning Research*, S. 993–1022.
- Bubenhof, Noah (im Druck): Semantische Äquivalenz in Geburtserzählungen: Anwendung von Word Embeddings. In: *Zeitschrift für Germanistische Linguistik*.
- Bubenhof, Noah/Calleri, Selena/Dreesen, Philipp (2019): Politisierung in rechtspopulistischen Medien: Wortschatzanalyse und Word Embeddings. In: *OBST. Osnabrücker Beiträge zur Sprachtheorie*, S. 211–242.
- Feilke, Helmuth (1994): *Common sense-Kompetenz. Überlegungen zu einer Theorie „sympathischen“ und „natürlichen“ Meinens und Verstehens*. Frankfurt a.M.
- Graham, Shawn/Weingart, Scott/Milligan, Ian (2012): *Getting Started with Topic Modeling and MALLET. The Programming Historian*. Online: <http://programminghistorian.org/lessons/topic-modeling-and-mallet>.
- Jarren, Otfried (2020): Die Corona-Pandemie ist eine besondere Krise. In: *higgs*. Online: <https://www.higgs.ch/die-corona-pandemie-ist-eine-besondere-krise/33283/>.
- Lenci, Alessandro (2018): Distributional Models of Word Meaning. In: *Annual Review of Linguistics* 4, S. 151–171.
- Mikolov, Tomas et al. (2013): Efficient Estimation of Word Representations in Vector Space. In: arXiv:1301.3781 [cs].