

Auswirkungen der Corona-Pandemie auf deutsche (Online-)Pressesprache: Erfassen, Analysieren, Explorieren

Sascha Wolfer, Alexander Koplenig, Frank Michaelis, Carolin Müller-Spitzer

Leibniz-Institut für Deutsche Sprache, Mannheim

Die Corona-Pandemie beeinflusst fast jede Facette des öffentlichen Lebens, und das praktisch auf der ganzen Welt. Es ist dabei nur allzu verständlich, dass die Pandemie nicht nur in persönlicher Face-to-Face-Kommunikation (soweit diese überhaupt noch möglich ist) einen großen Teil einnimmt, sondern auch die Nachrichten großflächig beherrscht. Als die erste Infektionswelle das deutschsprachige Gebiet erreichte, wurden große Teile des öffentlichen Lebens eingestellt oder starken Restriktionen unterworfen. Dies hat einen Einfluss auf die Berichterstattung: Kulturredaktionen gehen die öffentlichen Veranstaltungen aus, über die sie berichten können. Dasselbe gilt für das Sportressort. Doch auch Politik- und Wirtschaftsressorts müssen auf die Krise reagieren und berichten ausführlich über die Effekte der Pandemie auf die Gesellschaft. Dazu gehören u.a. die Auswirkungen auf das Gesundheits- und Pflegesystem, Kontaktsperrungen, finanzielle Hilfspakete und andere Maßnahmen, die von Regierungsseite aus unternommen werden. In unserem Vortrag möchten wir drei miteinander verbundene Ressourcen vorstellen, die dabei helfen sollen, die Effekte der Corona-Pandemie auf die deutsche Sprache abzubilden, zu analysieren und explorierbar zu machen.

Die grundlegende Ressource ist ein RSS-Korpus deutschsprachiger (Online-)Pressesprache, mit dem wir seit Anfang 2020 anhand elf bundesdeutscher Quellen und je einer Quelle aus Österreich und der Schweiz ein aktuelles Abbild der Verwendung der deutschen Sprache in Online-Presse-Quellen erfassen. Das Korpus umfasst derzeit (Stand 08.10.2020) ca. 20,3 Millionen Token.

Basierend auf diesem Korpus gehen wir in *cOWIDplus Analyse* zwei Hypothesen nach, die die Häufigkeitsverteilung im Korpus über die Zeit hinweg betreffen:

- 1) Die Corona-Pandemie führt zu einer quantitativ messbaren Verengung des Vokabulars in dem Sinne, dass corona-relatiertes Vokabular einen vergleichsweise großen Anteil der Gesamtfrequenz im Korpus einnimmt.
- 2) Sobald die Pandemie und deren Auswirkungen „unter Kontrolle“ sind, entwickelt sich die quantitativ messbare Verengung des Vokabulars zurück auf Prä-Pandemie-Niveau.

Wir werden zeigen, dass Hypothese 1 zutrifft: Informationstheoretische Redundanz, *mean-segmental type-token ratio* und der Frequenzanteil der TOP100-Wortformen zeigen zur Hochzeit der ersten Pandemiewelle (Mitte März bis Mitte April 2020) eine quantitativ messbare Einschränkung des Vokabulars. Wochenbasierte Frequenzlisten der TOP20-Wortformen untermauern dies. Hypothese 2 hingegen scheint sich insofern nicht zu bewahrheiten, dass schon deutlich früher ein Rückgang auf Prä-Pandemie-Niveau messbar ist. Es scheint als hätten wir zwar nicht die Stärke der Auswirkungen der Pandemie auf den von uns erhobenen Sprachausschnitt überschätzt, doch aber die Dauer der Auswirkungen.

cOWIDplus Analyse ist als eine kontinuierlich aktualisierte Webseite konzipiert, auf der die zeitliche Entwicklung der Häufigkeitsverteilung nachvollzogen und kommentiert wird. Außerdem sind hier das zugrundeliegende RSS-Korpus sowie die abgeleiteten Maßzahlen in Form von täglichen Uni- und Bigrammfrequenzlisten frei zum Download verfügbar.

Da die unter *cOWIDplus Analyse* dargestellten Untersuchungen nur eine Möglichkeit sind, den Auswirkungen der Corona-Pandemie auf die deutsche (Online-)Pressesprache nachzugehen, haben wir zusätzlich den *cOWIDplus Viewer* konzipiert, eine Webapplikation zur Exploration der Häufigkeitsentwicklungen einzelner Wortformen und Zweiwortverbindungen. Inspiriert vom

Google Books Ngram Viewer (Michel et al. 2011) können hier auch mehrere Wörter (bzw. Teile von Wörtern und/oder Bigramme) im zugrundeliegenden RSS-Korpus gesucht werden und in ihrer Entwicklung ihrer (relativen) Häufigkeiten verglichen werden. Dabei kann der Datumsbereich eingeschränkt sowie die Glättung (das Fenster für den gleitenden Mittelwert) angepasst werden. Rechercheergebnisse können sowohl in Form von Tabellen als auch als Abbildungen heruntergeladen werden. Zusätzlich ermöglicht es der integrierte „Bigramm-Finder“, Zweiwortverbindungen anhand eines Bestandteils (oder Teilen davon) zu finden.

Wir zeigen fallstudienartig, welche pandemie-relatierten Häufigkeitsveränderungen mit dem *cOWIDplus Viewer* abgebildet werden können, u.a. wie das Coronavirus andere, sonst stets relevante Themen (bspw. Fußball) ablöst, wie schnell und in welchen Verbindungen die Wortform „Klopapier“ plötzlich an Relevanz gewinnt und wie bekannte Virolog*innen (zumindest temporär) in eine ähnliche „Häufigkeitsliga“ aufsteigen wie Spitzenpolitiker*innen.

Ein Korpus mit möglichst aktuellen Daten, wie es das zugrundeliegende RSS-Korpus ist, ist besonders dafür geeignet, um hochaktuelle Entwicklungen im Wortschatz des Deutschen nachzuvollziehen. Wir wollen daher die Ressourcen auch in Zukunft weiterführen. Das bietet der Forschungsgemeinschaft auch nach Abklingen der Corona-Pandemie die Möglichkeit, den Einfluss von weltweit relevanten Großereignissen (so bspw. auch die US-Präsidentschaftswahlen im November 2020) auf den aktuellen Sprachgebrauch abzubilden.

Literatur:

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, ... Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014). 176–182.