



# Raumgeographische Verteilung von Twitter- Hashtags im deutschen Sprachraum

**STEVEN COATS**

**Abstrakt** Diese Studie untersucht die räumliche Verteilung von Hashtags in einem Korpus deutschsprachiger Tweets unter Berücksichtigung dreier Arten von Nutzerstandortinformationen: exakter Standort, kodiert als Breitengrad-Längengrad-Koordinaten, ein „place“-Attribut, ausgewählt aus einer von Twitter geführten Liste von Orten oder ein freier Eintrag im Nutzerprofil. Hashtags in Tweets mit exakten Ortsangaben weisen mit etwas höherer Wahrscheinlichkeit eine räumliche Konzentration auf als Hashtags mit Orts- oder Nutzerangaben, was möglicherweise auf die Verwendung von Mobilgeräten zur Veröffentlichung von Tweets zurückzuführen ist. Die Analyse der räumlichen Autokorrelation zeigt zwar, dass die meisten Hashtags keine starke räumliche Tendenz aufweisen, aber bei denjenigen, bei denen dies der Fall ist, handelt es sich meistens um Toponyme, Appellativa oder Eigennamen, die mit bestimmten Orten in Verbindung gebracht werden, wie eine Kartierung der Autokorrelationswerte veranschaulicht. Darüber hinaus beschreiben einige Hashtags, die eine räumliche Tendenz aufweisen, lokalisierte geografische oder meteorologische Phänomene. Hashtags können somit, trotz ihrer universeller Verschlagwortungsfunktion, wie andere lexikalische Elemente als Mittel zur Erzeugung lokaler Bedeutungen verwendet werden.

**Keywords** Twitter, Hashtags, German, spatial analysis

## 1. Einleitung

Das Hashtag ist in den letzten 15 Jahren zu einem wichtigen und vielseitigen Informations- und Diskursorganisationsmittel avanciert beim internetbasierten Austausch von Text, Fotos, Videos, Spielinhalten und anderen Dateien auf sozialen Netzwerken. Hashtags werden verwendet, um eine Vielfalt von unterschiedlichen kommunikativen Funktionen wahrzunehmen: auf sozialen Netzwerken werden Hashtags eingesetzt als Mittel zur Diskurskontextualisierung oder -strukturierung, indem sie Verbindungen herstellen von einem Beitrag zu anderen Beiträgen eines spezifischen Diskursstranges. Auf Twitter und anderen sozialen

Netzwerken werden Hashtags automatisch in Hyperlinks umgewandelt, die beim Anklicken andere Tweets anzeigen, die das Hashtag enthalten. Hashtags können auch bewertend eingesetzt werden, um zum Beispiel Wertschätzung für die Lieblingsfußballmannschaft oder Antipathie gegen einen unliebsamen politischen Akteur zu signalisieren. Sie zeigen ein breites Spektrum möglicher kommunikativer Funktionen, und selbst innerhalb eines bestimmten Diskurses können Hashtags unterschiedliche Funktionen haben (dazu Page 2012; Zappavigna 2012, 2018; Dang-Anh et al. 2013; Wikström 2014; Squires 2015).

Hashtags werden nicht nur verwendet, um den Inhalt eines Beitrags zu indizieren, ihn auf einen bestimmten Diskurskontext auszurichten oder die Zugehörigkeit eines Nutzers zu einer Wertegemeinschaft zu signalisieren: Sie können auch verwendet werden, um verschiedene Arten von geographischen Informationen explizit zu kennzeichnen, z. B. den Standort des Nutzers, den Standort eines in einem Text, Foto oder Video dargestellten Gegenstandes oder Ereignisses oder den Standort eines in einem Beitrag kommentierten Sachverhaltes, der möglicherweise mit zusätzlichen Hashtags kontextualisiert wird (z. B. „Zweiter Verhandlungstag gescheitert bei Trump-Kim Gipfeltreffen #Trump #KimJongUn #Singapur“). Als Hashtag verwendet, kann eine Ortsbezeichnung auf Twitter Verschlagwortungsfunktionen übernehmen, aber auch andere Arten von Standortinformationen können in einem Tweet vorkommen, z. B. der vom Nutzer angegebene Wohnort oder ein genauer Standort als Breite- und Längenkoordinaten angezeigt.

Hinsichtlich seiner Funktion ist der Hashtag ein quasi-lexikalisches Element, das verschiedene Arten von semantischen Inhalten tragen und in unterschiedlichen pragmatischen Kontexten verwendet werden kann (Schudy 2021), aber auch ein technisches Mittel zur Bündelung von Diskursfragmenten (vgl. Bernard 2021), um sie für die Visualisierung und Auswertung durch algorithmische Verfahren verfügbar zu machen. Im Hinblick auf die geografische Verbreitung von Hashtags stellt sich die Frage: Gibt es einen Zusammenhang zwischen der Semantik des Hashtags als Verschlagwortungsmittel und seinem Bezug zu lokalen Entitäten? Oder werden Hashtags als Indizierungs- und Bündelungsinstrumente unabhängig von geografischen Gesichtspunkten verwendet?

Aus dieser Konstellation ergeben sich die Forschungsfragen dieser Arbeit: Erstens, wie sind Hashtags in geolokalisierbaren Tweets im deutschen Sprachraum räumlich verteilt, und zweitens, wie sieht die räumliche Verteilung für spezifische Hashtags aus? Um diesen Fragen nachzugehen, wird im folgenden Abschnitt ein Überblick über den Stand der Erforschung von raumbezogenen Sozialmedien- und Twitterdaten geboten. Im dritten Abschnitt werden die unterschiedlichen Arten von geographischen Informationen in Tweets kurz skizziert, das Untersuchungskorpus von Tweets beschrieben, das Kartenwerk für die Analyse erläutert und die Methoden der Analyse kurz vorgestellt. Im vierten Abschnitt werden die Analyse und die Ergebnisse vorgestellt. In der anschließenden Diskussion im fünften Abschnitt werden zunächst einige Vorbehalte in Bezug auf Methoden und Interpretationen geäußert und dann eine linguistische Interpretation der Beziehung zwischen der geografischen Verteilung von Hashtags und ihren möglichen Bedeutungen vorgeschlagen. Zumindest für eine Teilmenge von Hashtags sind lokale Bedeutungen wichtig: Dies gilt nicht nur für Toponyme, die direkt Orte bezeichnen, sondern auch für eine Vielzahl von Hashtags, die indirekt lokale Phänomene, Organisationen, Ereignisse, Begebenheiten oder Einstellungen indizieren. Das Potenzial von Hashtags, lokale Bedeutungen zu tragen, hat in dieser Hinsicht eine gewisse Ähnlichkeit mit dem indexikalischen Bedeutungspotenzial einiger lexikalischer Elemente in traditionellen Dialekten.

## 2. Forschungsüberblick

Hashtags in sozialen Medien können eine Reihe von Text-, Diskurs- und Kommunikationsfunktionen aufweisen und nach verschiedenen kommunikativen, pragmatischen, und strukturellen Parametern klassifiziert werden (Page 2012; Wikström 2014; Squires 2015; De Cock & Pedraza 2018; Zappavigna 2018; Schudy 2021). Das Verwenden von Hashtags kann als spezifische, mediatisierte Kommunikationsstrategie im Rahmen der Online-Öffentlichkeitsinszenierung verstanden werden, die für viele Persönlichkeiten des öffentlichen Lebens wie Politiker oder Unterhaltungsindustriefiguren immer wichtiger geworden ist (Thimm et al. 2012; Dang-Anh et al. 2013).

Hashtags werden auch als Marketingmittel eingesetzt, um Produkte, Events, oder Kampagnen Profil zu geben (Bauer & Goetz 2021). Diskurs, der durch die Verwendung von Hashtags organisiert wird, kann sowohl qualitativ, im Sinne einer Beschreibung der kommunikativen Funktionen und Wirkungen der Hashtagverwendung einzelner Akteure, als auch quantitativ, durch korpusbasierte Analyse von Hashtags mit spezifischen thematischen Inhalten, ausgewertet werden (z. B. zum Thema *Brexit*: Amador Diaz Lopez et al. 2017; Grčar et al. 2017; Hürlimann et al. 2017; Bastos & Mercea 2018; Georgiadou et al. 2020). In einem breiteren Kontext kann die korpusbasierte Analyse der Hashtag-Benutzung in sozialen Medien je nach Untersuchungsgegenstand der *computational social science* (Lazer et al. 2009) oder *computational sociolinguistics* (Nguyen et al. 2016) zugeordnet werden.

In Bezug auf Geographie sind Twitterdaten in diversen Studien zu unterschiedlichen Zwecken verwendet worden. Einen wichtigen Teilbereich der raumbezogenen Sozialmedienforschung stellt die Analyse (und ggf. Vorhersage) von geographisch verteilten Ereignissen anhand deren Repräsentation auf Twitter oder in anderen sozialen Medien dar. Solche Studien haben eine breite thematische Reichweite und erstrecken sich von der Lokalisierung von Erdbeben, Epidemien, oder Unruhen (Murzintcev & Chen 2017; Shibuya 2017) über die Vermarktung von Tourismuszielen (Andéhn et al. 2014) bis zur geographischen Dynamik von Massenmigrationen (Hübl et al. 2017). Prädiktives Modellieren von Nutzerinformationen oder demographischen Eigenschaften anhand von Metadaten und Textinhalten, zum Beispiel Wohnort, ist ein lebendiges Forschungsfeld. Für Twitterdaten können zum Beispiel die Zeit der Veröffentlichung eines Tweets, die im Profil angegebene Zeitzone oder das Verwenden von *location indicative words* wie Ortsnamen, Eigennamen, und Dialektwörter, zum Beispiel *Picadilly*, *Lakers*, oder *yinz*, aussagekräftig im Hinblick auf den Nutzerstandort sein (vgl. Eisenstein et al. 2014; Han, Cook & Baldwin 2014).<sup>1</sup>

In der Sprachwissenschaft haben z. B. Eisenstein et al. (2014) die geographische und temporale Verbreitung von neuen Englischen Wortformen in einem Korpus von geolokalisierten Tweets aus den USA untersucht, und

---

1 Diese Wörter können assoziiert werden mit London, Los Angeles, und Pittsburgh.

Coats (2019) hat die Sprachwahl von mehrsprachigen Twitter-Benutzern in Europa analysiert. Grieve et al. (2019) haben die Raumverteilung von lexikalischer und grammatischer Variation in einem Korpus von geolokalisierten Englischen Tweets aus der UK mit Mustern aus klassischen dialektologischen Datenbeständen verglichen. In beiden Studien wurde festgestellt, dass geographische Nähe zumindest bis zu einem gewissen Grad mit sprachlicher Ähnlichkeit korreliert.

Nur wenige Studien haben sich speziell mit der Beziehung zwischen der Semantik von Hashtags und ihrer geografischen Verteilung befasst. Die geografische Analyse von Hashtags wird zwar von Unternehmen, die sich auf die Analyse sozialer Medien spezialisiert haben, als Dienstleistung angeboten, um beispielsweise die Resonanz einer Werbekampagne besser zu verstehen, doch geschieht dies in der Regel im Dienste der Kundenforschung und des Marketings und nicht, um semantische Fragen auf einer grundlegenden Ebene zu behandeln.<sup>2</sup>

Die Ergebnisse von Studien, die speziell die Raumverteilung von Hashtags analysieren, sind gemischt. Bei einer Analyse der globalen Verteilung von Hashtags fanden Kamath et al. (2013), dass Hashtags eher lokal (d.h. in einem Land) als global verbreitet sind, wobei die Wahrscheinlichkeit einer globalen Verbreitung höher ist für englischsprachige Tweets. Leetaru et al. (2013) fanden bei einer globalen Analyse von Tweets mit Geometadaten heraus, dass Tweetinhalte wie Hashtags oder tweetabgeleitete Informationen wie Netzwerke von Followern nur wenig standortsabhängig sind. Analysen der Raumverteilung von Hashtags in auf politische Ereignisse bezogenen Tweets innerhalb eines Landes zeigen nicht unbedingt eine klar nachvollziehbare geographische Verteilung. Bastos et al. (2013) fanden, dass die Geographie von politischen Protesten in Brasilien 2013 sich nur teilweise deckt mit der Geoinformation in darauf bezogenen Tweets. Auf ähnliche Weise stimmt das politische Signal in Geo-Metadaten-enthaltenden Tweets mit brexitbezogenen Hashtags nur partiell überein mit den Abstimmungsergebnissen des Brexitreferendums 2016 in der UK (Bastos & Mercea 2018). Hashtags, welche

---

2 Z.B. Keyhole, <https://keyhole.co> (letzter Zugriff 1. Dezember 2022); siehe dazu Bernard (2021).

die Klimakrise thematisieren, weisen im Vereinigten Königreich kein klares geografisches Muster auf (Schepanski & Seppelfricke 2021).

Dass lexikalische, phonetische/phonologische, und morpho-syntaktische Variabilität geographisch konditioniert ist und dass Dialektähnlichkeit meist mit zunehmender räumlicher Entfernung nachlässt, ist von Nerbonne und Kleiweg als *fundamental postulate of dialectology* formuliert worden (2007: 149). Das Postulat beruht auf der Tatsache, dass Sprachgemeinschaften traditionellerweise raumverbunden sind und dass Sprecher deswegen weniger Kontakt haben zu Menschen in entfernten Orten. Obwohl diese Bedingung nicht für soziale Netzwerke gilt, sind Sprachmerkmale wie Lexik und Syntax auch in geolokalisierten Daten von sozialen Netzwerken nachweisbar, vermutlich aufgrund der Tatsache, dass die Nutzer diese Sprachmerkmale nicht nur sprechen, sondern auch in Online-Kommunikation verwenden.

Hashtags, obwohl sie schriftsprachlich lexikalisiert werden und somit gelegentlich zu eigenständigen, in Wörterbüchern verzeichneten Wörtern ohne Rautenzeichen aufsteigen können, werden hauptsächlich als Kontextualisierungszeichen verwendet im Socialmediendiskurs und eher selten als phonologische Wörter realisiert: Es stellt sich die Frage, ob Hashtags kohärente raumgeographische Verteilungsmuster aufweisen.

### 3. Daten und Methoden

#### 3.1 Standortinformation in Tweets

Die Tweets in diesem Korpus können verschiedene Arten von Standortinformationen enthalten. Tweets von Nutzern, die im Profil eine Standortangabe eingetragen haben, enthalten im Tweet-Datenobjekt ein *user:location*-Feld, dessen Wert eine beliebige Unicode-Zeichenfolge ist (z. B. „Berlin“, „auf dem Mond“, „“), und die jederzeit geändert werden kann. Wenn ein Nutzer im Profil die Option „Standortinformationen zu meinen Tweets hinzufügen“ ausgewählt hat, kann bei jeder Tweet-Veröffentlichung ein Standort zu dem Tweet hinzugefügt werden; diese Information wird im *place*-Attribut des Tweet-Datenobjekts gespeichert. Standorte werden gewählt aus einem Twitter-internen Standortverzeichnis, welches dem Nutzer eine Liste von Stand-

orten zeigt aufgrund seiner IP-Adresse oder des GPS-Signals seines Mobilgeräts; Nutzer können aber auch einen beliebigen Standort auswählen aus dem Verzeichnis. Diese Art von Standortinformation, eine *point-of-interest*-Metadatei, die von vielen Onlinediensten verwendet wird (vgl. Hochmair et al. 2018), wird in Twitter-Metadaten als vier Längen- und Breitenkoordinaten gespeichert, welche den gewählten Standort im kleinsten umschreibenden Rechteck (dem *bounding box*) umfassen. Drittens, bei der Auswahl „Standortinformationen hinzufügen“ auf GPS-fähigen Geräten können Tweets mit genauen Breiten- und Längenkoordinaten versehen werden; diese werden im *coordinates*-Feld gespeichert und automatisch dem entsprechenden Standort des Standortverzeichnisses zugewiesen.<sup>3</sup> Der Nutzer kann diesen Standort ändern, indem er bei der Tweetveröffentlichung einen anderen Standort aus dem Standortverzeichnis wählt.

Die Werte dieser drei Metadatenfelder müssen nicht übereinstimmen: Ein Nutzer, der „Berlin“ als *user:location* angegeben hat, kann einen Tweet mit einem anderen geographischen Tag versehen, um zum Beispiel ein Ereignis an einem anderen Ort zu kommentieren, oder kann einen Tweet auf einem Mobiltelefon veröffentlichen bei einem Aufenthalt in Hamburg oder woanders. Obwohl nicht alle Nutzer einen Standort in dem *user:location* Feld angeben (Hecht et al. 2011), gibt es bei Nutzern, deren Tweets sowohl diese Metadaten als auch *place* Metadaten enthalten, eine hohe Übereinstimmung zwischen dem Mittelwert der *place*-Angaben und dem ermittelten *user:location* (Johnson et al. 2016: 520). Es gibt ebenfalls in der Regel eine hohe Korrelation zwischen den Koordinaten des Flächenmittelpunktes des *place*-Rechtecks und der *coordinates*-Angabe, obwohl relativ wenige Tweets *coordinates*-Metadaten enthalten (Leetaru et al. 2013; Sloan & Morgan 2015; Laylavi et al. 2016).

---

3 Seit Juni 2019 können vom Twitterklienten nur noch *place*-Metadaten, und nicht mehr genaue GPS-Koordinaten zu den meisten Tweetarten hinzugefügt werden (<https://twitter.com/TwitterSupport/status/1141039841993355264>, letzter Zugriff 1. Dezember 2022).

### 3.2 Korpusdaten

Das für die Studie verwendete Korpus besteht aus 36 240 530 deutschsprachigen Tweets (534 211 366 Token), die von Nutzern aus Deutschland, Österreich und der Schweiz stammen. Das Korpus wurde in mehreren Schritten erstellt. Zunächst hat der Autor dieses Beitrags 653 457 659 Tweets mit *place*-Metadaten von November 2016 bis Juni 2017 mit Hilfe von Tweepy (Roesslein 2015) weltweit von der Twitter Streaming API gesammelt. Aus diesen Daten wurden Nutzer identifiziert, deren Profile *place*-Metadaten in Deutschland, Österreich oder der Schweiz enthielten und die mindestens einen deutschsprachigen Tweet gesendet hatten (70 986 Nutzer). Alle Tweets dieser Nutzer (bzw. die letzten 3 250, die Obergrenze für das Herunterladen von Tweets eines einzelnen Nutzers aus der REST-API von Twitter) wurden im April 2018 heruntergeladen, um ein vorläufiges Korpus zu erstellen von insgesamt 61 118 733 Tweets. Auf Deutsch geschriebene Tweets (59,3%) wurden beibehalten.<sup>4</sup> Diesen Tweets wurden aufgrund Standort-Metadaten Längen- und Breitengradkoordinaten zugewiesen: Bei Tweets mit genauen Standortangaben (*coordinates*) wurden die Längen- und Breitengradkoordinaten verwendet. Falls *place*-, aber keine *coordinates*-Metadaten vorhanden waren, wurde der berechnete Mittelpunkt des Umschreibungsrechteckes verwendet. Bei Tweets, die weder *coordinates*- noch *place*-Metadaten enthielten, wurde die *user:location*-Angabe mit dem Geocoder geopy und einer Geocoding-API in Längen- und Breitengradkoordinaten aufgelöst (Esmukov et al. 2020).<sup>5</sup> *Place*-Standorte, die mindestens bis auf Kreisebene auflösbar waren, wurden beibehalten.<sup>6</sup> Insgesamt konnten 77.2% der Tweets einem Standort zugeordnet werden.

---

4 Die Tweet-IDs sind verfügbar bei <https://github.com/stcoats/Germananglicisms> (letzter Zugriff 1. Dezember 2022). Das Korpus kann durch Herunterladen der Tweets von der Twitter-API erstellt werden.

5 <https://developers.google.com/maps/documentation/geocoding/overview> (letzter Zugriff 1. Dezember 2022).

6 Tweets mit grobkörnigeren *place*-Angaben wie "Deutschland", "Bayern" oder "Austria" wurden nicht berücksichtigt, weil deren berechnete Umschreibungsrechteckmittelpunkte eine sehr große Bandbreite von möglichen Standorten abdecken.

Um die Genauigkeit der Methode zu prüfen, wurden *user:location*-Angaben mit *coordinates*-Angaben verglichen für Nutzer und Tweets mit beiden Metadaten-Einträgen in Deutschland, Österreich und der Schweiz. Der Mittelwert für die Entfernung zwischen diesen zwei Metadaten-Feldern war 34,5 km, ein Auflösungsgrad, der nicht präzise genug wäre für eine feinkörnige Analyse z. B. der Tweetinhalte innerhalb einer einzigen Stadt (vgl. Hecht et al. 2011), aber eine räumliche Analyse gewährleistet, bei der Tweets zu geographischen Gebieten in der Größenordnung eines Verwaltungskreises zugeordnet werden.

Von den 47,2 Millionen standort aufgelösten Tweets konnten 39 238 144 einer der 730 in der folgenden geographischen Analyse berücksichtigten geographischen Flächen in Deutschland, Österreich, Belgien, Italien, Luxemburg, Frankreich, Dänemark, Liechtenstein oder in der Schweiz zugewiesen werden. Diese Tweets enthielten insgesamt 2 479 412 Hashtag-Typen und 24 353 452 Hashtag-Token. Tweet- und Hashtagaufkommen sind am höchsten in den bevölkerungsreichsten Flächen: Berlin, Hamburg, München, Köln, und Wien (Stadt) (Tabelle 1). In einigen Kartenflächen, zum Beispiel kleineren, nicht zusammenhängenden Teilgebieten eines Landkreises, wurden keine Tweets oder Hashtags verzeichnet, so zum Beispiel für die ostfriesischen Inseln Juist, Norderney und Baltrum im Kreis Aurich oder für die Gemeinde Planken in Liechtenstein.

**Tabelle 1: Flächen mit den meisten Tweets und Hashtags.**

Kartenfläche	Tweets	Hashtags
Berlin	5 547 649	3 385 121
Hamburg	2 237 406	1 339 642
München	1 866 308	1 236 280
Köln	1 570 435	996 105
Wien (Stadt)	1 533 771	874 544

Zu den häufigsten Hashtags im Korpus (Tabelle 2) gehören Städtenamen (#Berlin, #berlin, #Hamburg, und #München), politische Parteien und Politiker (#SPD, #AfD, #Merkel) und Fernsehunterhaltung (#Tatort, #tatort, und

#ibes, eine Abkürzung für die deutsche Fernsehshow *Ich bin ein Star – Holt mich hier raus!*). Die Hashtags #fb, #Facebook, #socialmedia und #ff beziehen sich auf soziale Netzwerke – Letzteres als Abkürzung von *follow Friday*, ein auf Twitter seit vielen Jahren verwendeter Aufruf zum gegenseitigen Folgen. #effzeh, eine Lautschreibung für „FC“, wird meist in Bezug auf den Fußballverein 1. FC Köln verwendet. #Digitalisierung dient oft der thematischen Zuordnung und #fail wird verwendet, um die Haltung eines Nutzers zum Tweetinhalt zu kommentieren (vgl. Zappavigna 2012, Kap. 8). Die Hashtags #Runtastic, #SoundCloud, und #gameinsight werden von Apps automatisch einem Tweet hinzugefügt, um Sport- oder Trainingsdaten mitzuteilen, den Inhalt einer Musik-Playlist zu veröffentlichen oder Ereignisse in einem Computer- oder Mobiltelefonspiel zu verkünden.

**Tabelle 2: Häufigste Hashtags.**

Rang	Hashtag	Anzahl	Rang	Hashtag	Anzahl
1	#Berlin	118 869	11	#Runtastic	31 395
2	#fb	58 607	12	#effzeh	30 705
3	#berlin	56 292	13	#SoundCloud	29 856
4	#Hamburg	52 417	14	#ff	29 854
5	#Tatort	46 748	15	#München	28 243
6	#Facebook	38 526	16	#socialmedia	26 878
7	#ibes	36 520	17	#Merkel	26 743
8	#SPD	34 606	18	#tatort	26 702
9	#Digitalisierung	34 491	19	#fail	26 561
10	#AfD	32 000	20	#gameinsight	26 389

In der Analyse im Abschnitt 4 wird das Vorkommen von häufigen Hashtag-Typen (Mindestaufkommen 100) ausgewertet.

### 3.3 Karten

Tweets wurden aufgrund ermittelter Standorte zu Flächen in Deutschland, Österreich, Liechtenstein, der Schweiz, und Luxemburg zugewiesen

sowie angrenzenden Teilgebieten in Italien, Frankreich, Belgien, und Dänemark, in denen die deutsche Sprache entweder einen offiziellen Status als Amtssprache hat oder traditionell von einer Minderheit gesprochen wird: Südtirol, den französischen Départements Haut-Rhin, Bas-Rhin, und Moselle, Luxemburg, dem Gebiet der deutschsprachigen Gemeinschaft in Belgien und dem dänischen Nordschleswig. Die 730 Kartenflächen, die 677 Verwaltungsgebieten hauptsächlich auf Kreis- oder Bezirksebene entsprechen, wurden auf der Grundlage von GIS-Daten aus verschiedenen Quellen zusammengestellt.<sup>7</sup>

### 3.4 Geographische- und Räumlichkeitsstatistik

Mit der Dialektometrie (Séguy 1971; Goebel 1982) wurde eine systematische Methode zur Erfassung von Dialektähnlichkeiten bzw. -unterschieden entwickelt, die zunächst zur Auswertung von aggregierten Daten aus historischen Sprachatlanten diente. Im traditionellen dialektometrischen Verfahren wird eine Ähnlichkeitsmatrix konstruiert, um die standortgebundenen Mess- oder Vergleichswerte in einem Datensatz paarweise zusammenzufassen und somit auf indirekte Weise die räumliche Verteilung der Variablen zu berücksichtigen. Bei der Geostatistik dagegen werden räumliche Relationen als quantitative Größen miteinbezogen bei der Berechnung der Verteilungswahrscheinlichkeit einer Variablen.

In den letzten zehn Jahren wurden zunehmend geostatistische Verfahren auch bei der Untersuchung von Sprachdaten verwendet, um räumliche Variationsmuster zu erkennen und Analysen von räumlich verteilten Variablen zu ermöglichen. Eine geostatistische Methode hat sich als besonders fruchtbar erwiesen: die Verwendung von Autokorrelationsstatistiken, welche das Ausmaß gegenseitiger Beeinflussung von benachbarten Gebieten für einen gegebenen Messwert quantifizieren, um Muster in räumlich verteilten Daten zu identifizieren (siehe, z.B., Grieve, Speelman & Geeraerts 2011; Siblinger 2011; Lameli 2013; Stoeckle 2016; Coats 2020). Im folgenden Abschnitt wird die

---

7 Die Anzahl von Flächen ist höher als die Anzahl von Verwaltungsgebieten, weil manche Verwaltungsgebiete mehrere nicht-zusammenhängende Kartenflächen umfassen.

globale Statistik Morans  $I$  verwendet (Moran, 1950; Cliff & Ord 1973, 1981; Getis 2010) um die raumgeographische Verteilung von Hashtags zu quantifizieren, d.h. um festzustellen, ob ein Hashtag geographisch gleichmäßig oder zufällig verteilt ist oder zusammengeballt in bestimmten Gebieten.<sup>8</sup> Dazu noch wird die lokale Autokorrelationsstatistik Getis-Ord  $G_i^*$  berechnet (Getis & Ord, 1992; Ord & Getis, 1995), um Standorte mit höheren oder niedrigeren Relativaufkommen für einen Hashtag zu identifizieren und kartographisch zu veranschaulichen. Autokorrelationen wurden mit dem Python-Modul PySAL unternommen (Rey & Anselin 2007).

## 4. Analyse und Resultate

### 4.1 Hashtagdichte

Die Hashtagdichte (Durchschnittsanzahl von Hashtags pro Tweet) wird auf Abb. 1 dargestellt: Flächenwerte variieren von 0 Hashtags (mehrere Flächen) bis zu knapp über 3 Hashtags pro Tweet (im Distrikt Diekirch, Luxemburg und im Bezirk Val-de-Travers, Kanton Neuenburg, Schweiz), mit einem Durchschnittswert von 0,65 Hashtags pro Tweet. Wie die Karte auf Abb. 1 veranschaulicht, weisen die Durchschnittswerte kein erkennbares geographisches Muster auf. Morans globale  $I$ -Statistik, berechnet für die Hashtagdichte aufgrund unterschiedlicher Gewichtungsmatrizen (Choroplethen-Kontinuitätswerte, verkehrte Distanzwerte, und von 5 bis 50-*Nearest-Neighbors*), bestätigt das Fehlen von Clusterbildung mit Werten, je nach Konfiguration der Gewichtungsmatrix, von 0,02 bis 0,059 — nur leichte Abweichungen von einem für eine zufällige Verteilung erwarteten Wert von -0,001. Somit weist die Neigung zur Verwendung von Hashtags im Allgemeinen kein besonderes geographisches Muster in den Daten auf — Hashtags werden in Städten und ländlichen Gebieten im gesamten betrachteten Gebiet verwendet.

---

<sup>8</sup> Für einen Überblick der Verwendung geostatistischer Methoden in der Sprachforschung mit Schwerpunkt auf dem deutschen Sprachraum siehe Pickl und Pröll (2019).

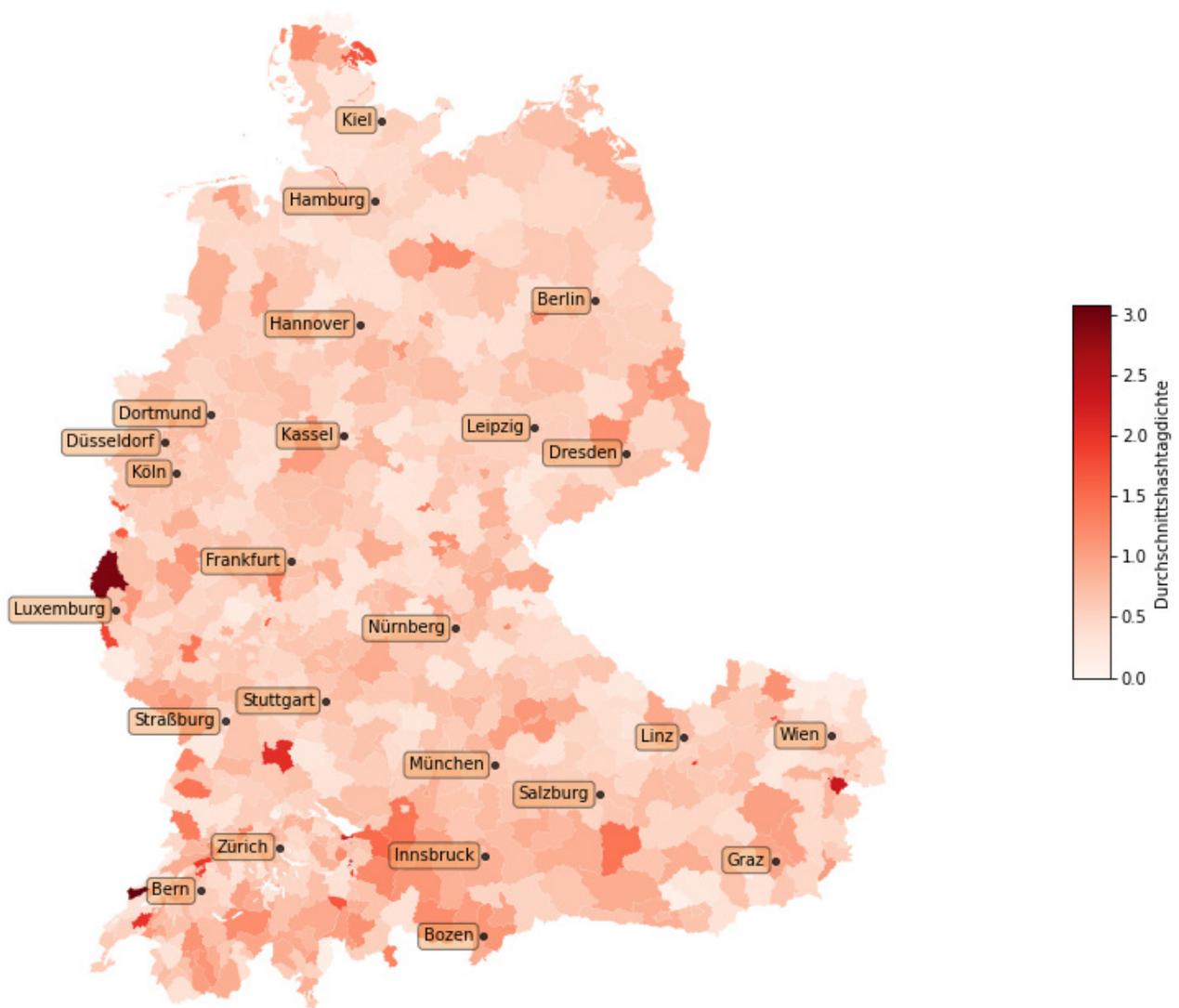
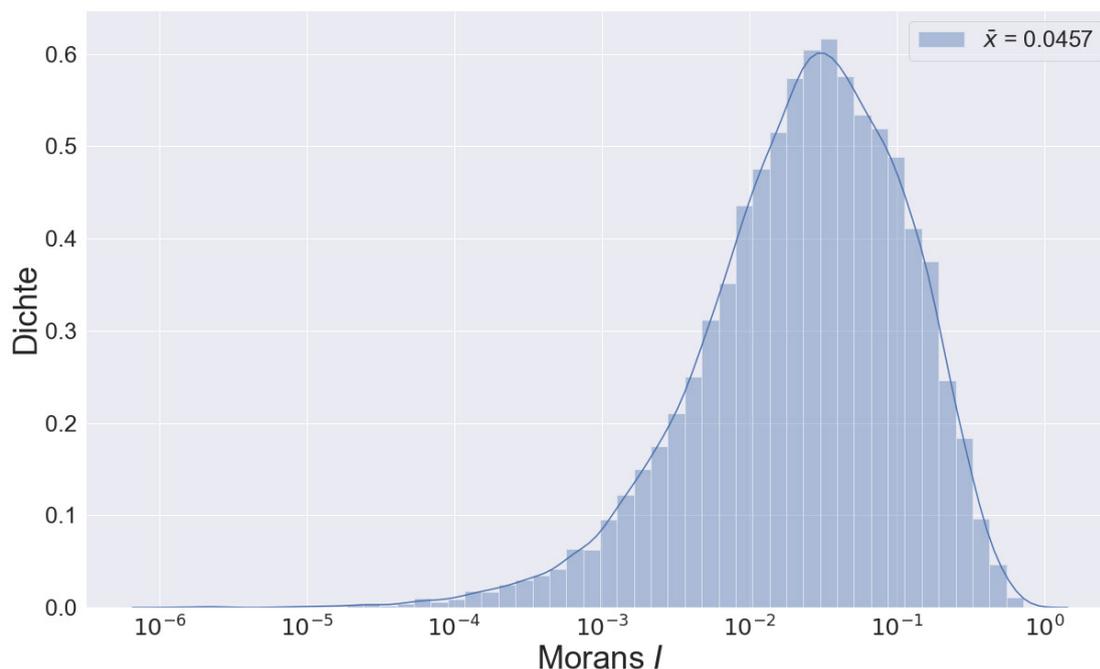


Abbildung 1: Hashtagdichte.

## 4.2 Häufige Hashtags

Um die räumliche Verteilung von einzelnen Hashtags zu untersuchen, wurden für die 30 025 häufigsten Hashtag-Typen (d.h. alle mit einem Mindestaufkommen von 100) Autokorrelationsstatistiken berechnet. Der Variablenwert, auf dessen Basis die Autokorrelationswerte kalkuliert wurde, war dabei das relative Aufkommen des jeweiligen Hashtags als Proportion aller in dem entsprechenden Choropleth aufkommenden Hashtags, unter Verwendung einer binären Kontiguitäts-Gewichtungsmatrix. Die Verteilung der absoluten *I*-Werte für diese 30 025 Hashtag-Typen wird in Abb. 2 dargestellt:



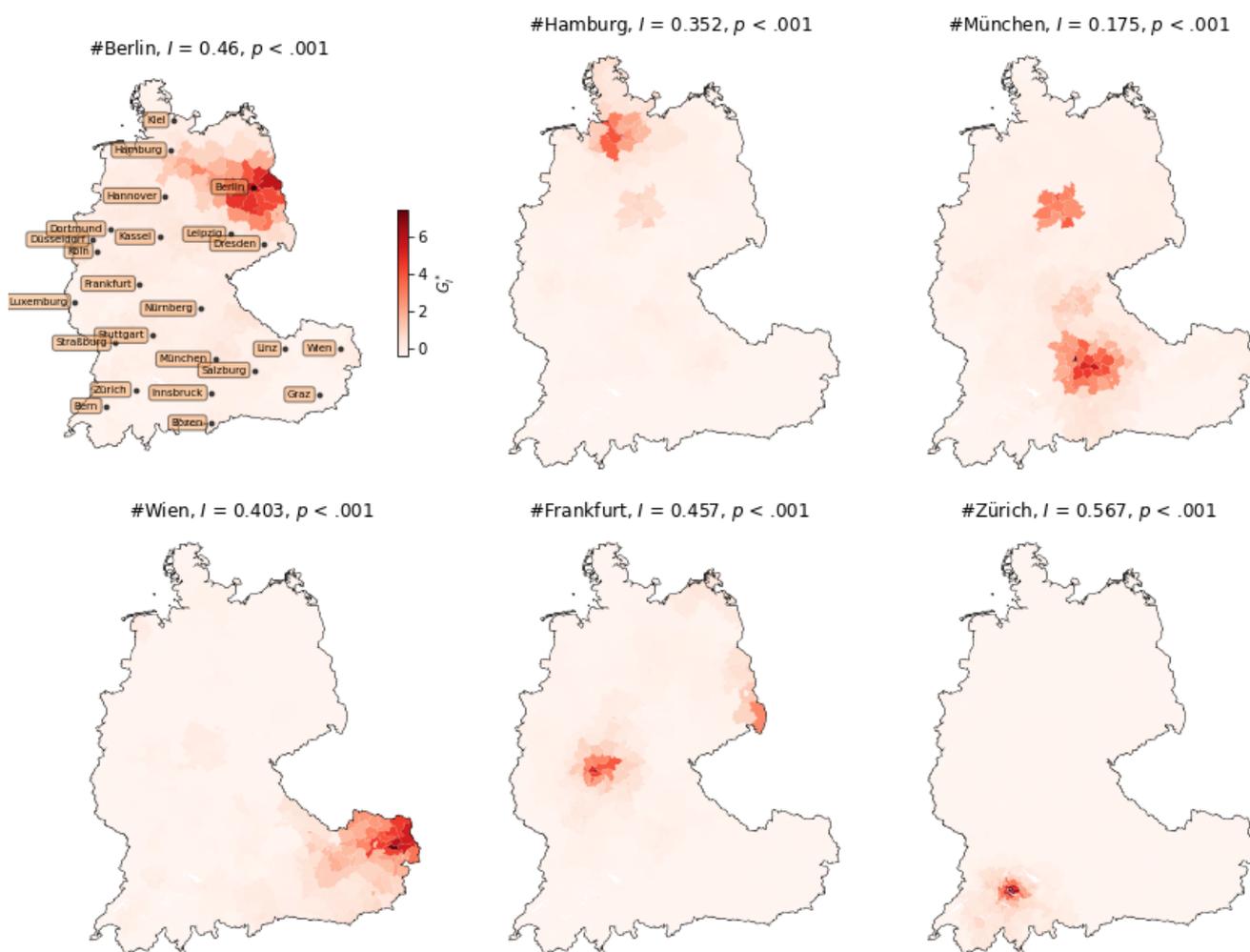
**Abbildung 2: Verteilung von Morans-I-Absolutwerten für 30 025 Hashtagtypen mit Mindestaufkommen von 100.**

Wie ersichtlich, zeigen die Werte eine leichte linksschiefe Verteilung mit einer Reichweite von  $6.47e-6$  bis  $0.74$  und einem Durchschnittswert von  $0,0457$ , nur geringfügig über dem Wert für eine willkürliche räumliche Verteilung. Insofern zeigen die meisten häufigen Hashtags nur ein schwaches räumliches Verteilungsmuster auf.

Obwohl das Gesamtaufkommen von Hashtags im untersuchten Raum kein erkennbares räumliches Muster aufweist und die Verteilung von Morans *I*-Werten für die Relativhäufigkeit einzelner Hashtags darauf hindeutet, dass die meisten Hashtags keine räumlichen Cluster bilden, ist für eine Anzahl von Hashtags eine klare Tendenz zu lokaler oder regionaler Clusterbildung erkennbar. In den folgenden Abschnitten werden zwei Kategorien von Hashtags berücksichtigt: erstens, Hashtags mit einem spezifisch geographischen Bezug (meist Städtenamen oder andere Eigennamen), und zweitens häufige Hashtags, deren Semantik keinen direkten Rückschluss auf geographisch lokalisierbare Einheiten zulässt. Um die räumliche Verbreitung dieser Hashtags zusammenzufassen wird Morans globale *I* berechnet; die Getis-Ord  $G_i^*$ -Statistik wird berechnet für die 730 Kartenflächen, um Ballungen von Werten zu identifizieren und kartographisch zu veranschaulichen.

#### 4.2.1 Hashtags mit geographischem Bezug

Ortsbezeichnungen wie Städtenamen sind als Hashtags meistens klar geclustert im Raum: Sie haben relativ hohe Morans  $I$ -Werte und zeigen höhere  $G_i^*$ -Werte in der unmittelbaren Umgebung der bezeichneten Stadt (Abb. 3).



**Abbildung 3: Morans- $I$  und Getis-Ord  $G_i^*$ -Werte für 6 Städtenamen-Hashtags.**

Die geographische Verteilung von Städtenamen-Hashtags lässt darauf schließen, dass Nutzer diese Hashtags eher verwenden, um einen lokalen Bezug herzustellen, als um Ereignisse an anderen Orten zu kommentieren. Die Tendenz von Nutzern, mit Hashtags eher lokale als ferne Entitäten zu indizieren, ist auch bei anderen Toponymen ersichtlich, wie Ländernamen auf Deutsch und auf Englisch (Abb. 4 und 5).

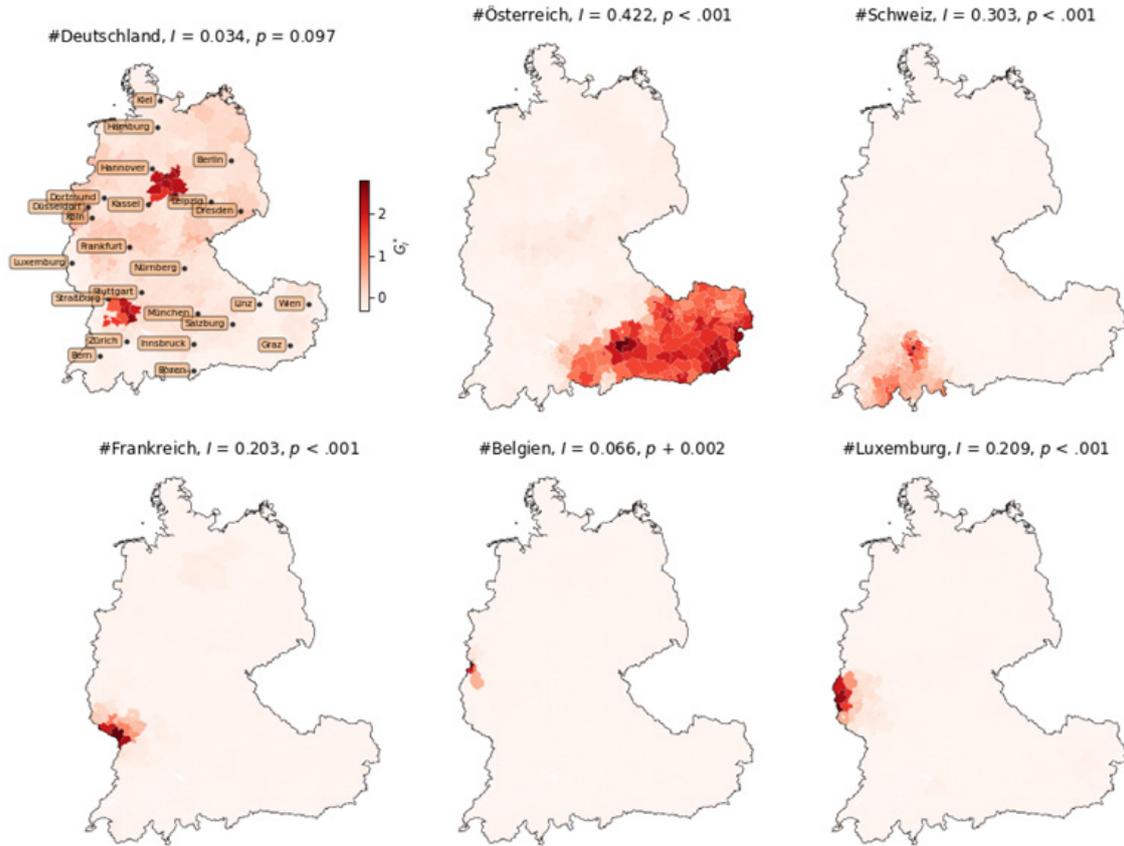


Abbildung 4: Morans-I und Getis-Ord  $G_i^*$ -Werte für Ländernamen auf Deutsch.

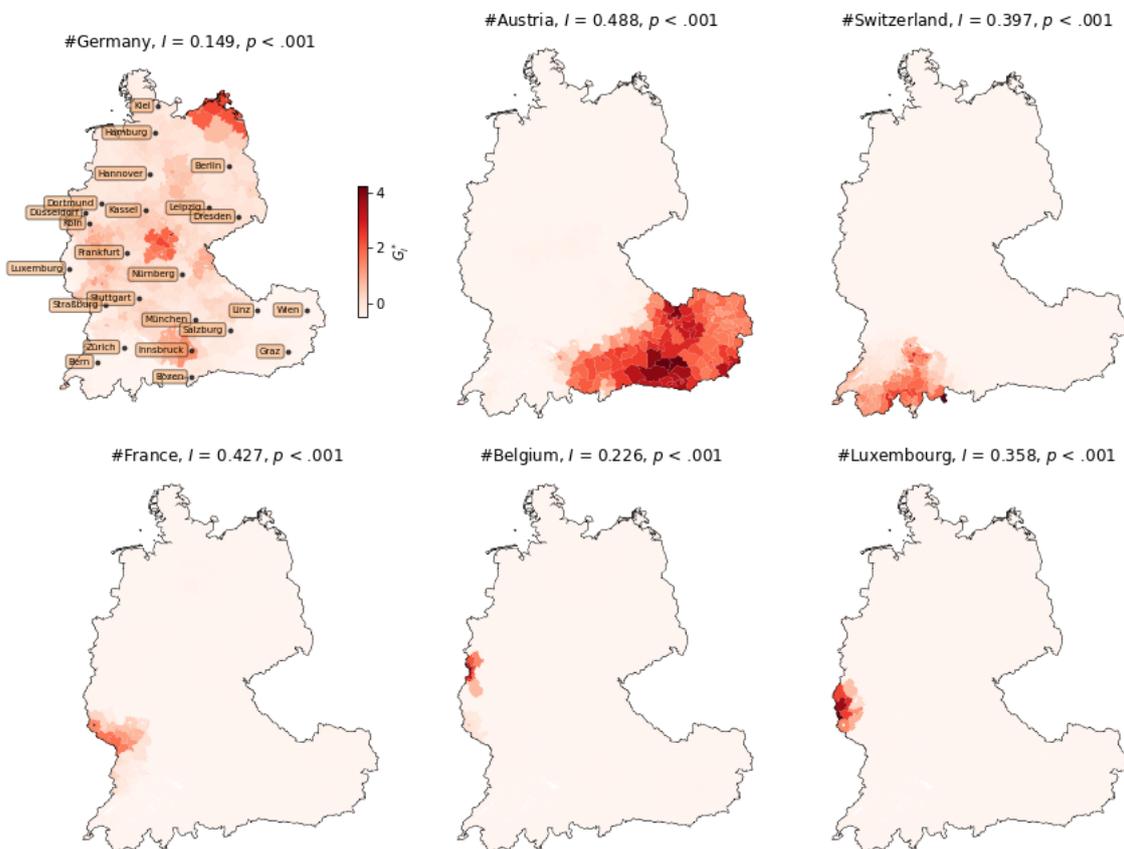
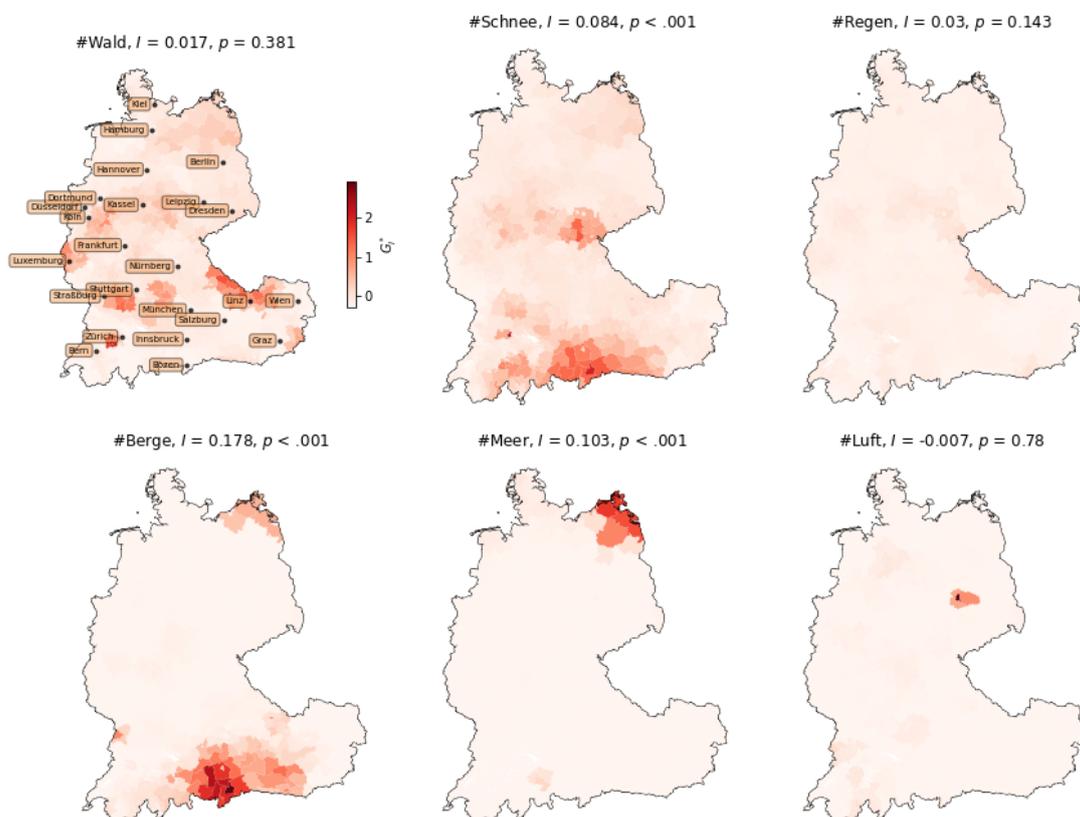


Abbildung 5: Morans-I und Getis-Ord  $G_i^*$ -Werte für Ländernamen auf Englisch.

Die höheren Werte für englische Ländernamen gehen möglicherweise auf die Tatsache zurück, dass die auf Englisch verfassten Tweets im Korpus einen höheren Anteil an *coordinates*-Standorten haben (siehe Diskussion im Abschnitt 5).

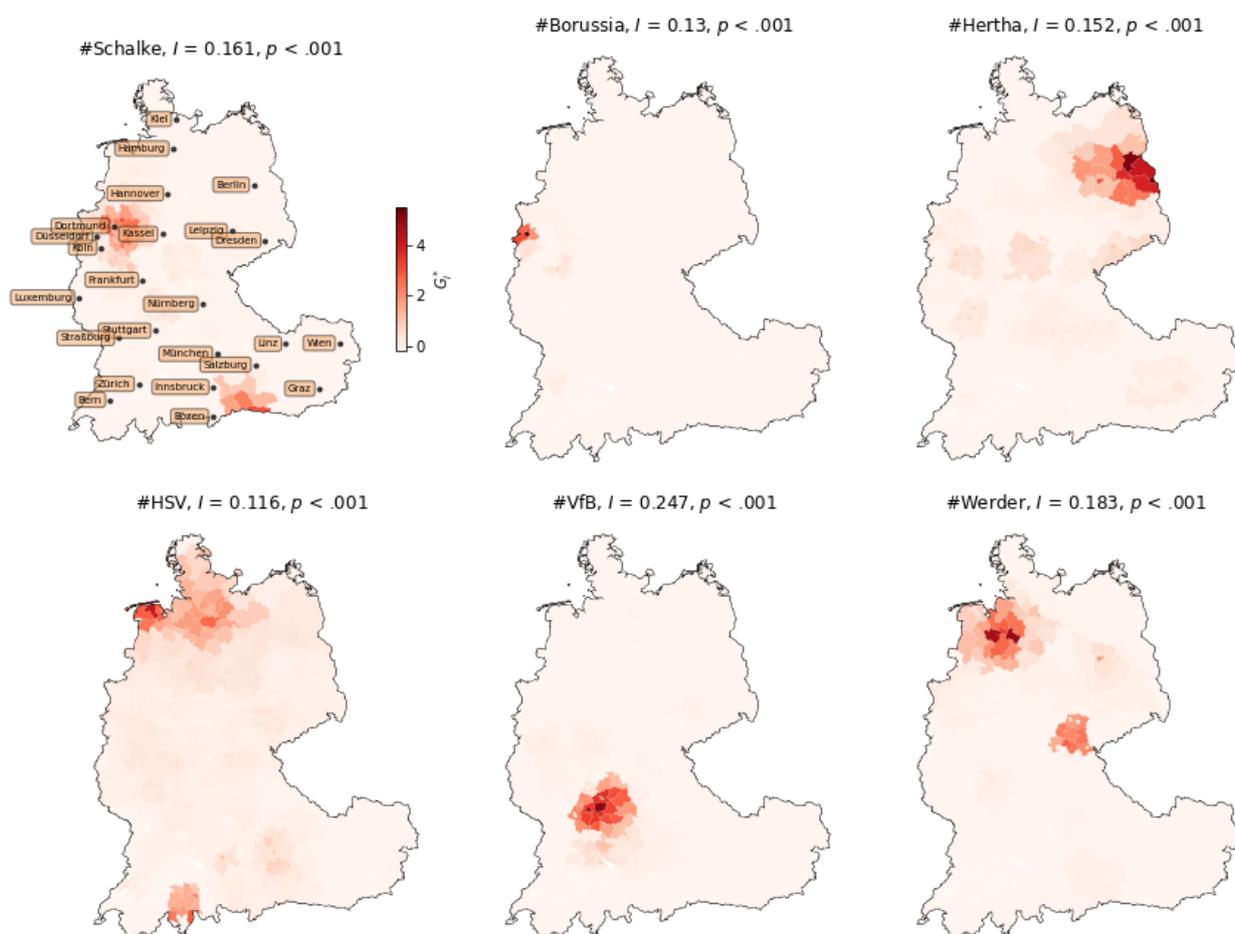
Es ist zwar durchaus nachvollziehbar, dass Toponyme wie Städte- oder Ländernamen geografisch geclustert sind, aber auch andere Arten von Hash-tags weisen eine Tendenz zur Clusterbildung auf, die anhand der berechneten Morans-*I*-Werte ermittelt wurde.

Viele Hash-tags, die nicht Toponyme sind, haben einen geographischen Bezug, zum Beispiel weil sie Landschaften oder Naturphänomene bezeichnen, die ungleichmäßig verteilt sind im Raum. Auf Abb. 6 werden die Getis-Ord  $G_i^*$ -Werte für die Hash-tags #Wald, #Schnee, #Regen, #Berge, #Meer, und #Luft dargestellt. #Wald ist ungefähr zufällig verteilt, mit einem Morans-*I*-Wert von 0,017, aber mit leicht erhöhten  $G_i^*$ -Werten für ein Gebiet, das sich vom Bayerischen Wald bis zum österreichischen Waldviertel erstreckt. #Schnee und #Berge kommen mit höheren Relativhäufigkeiten in den Alpen vor, und #Meer wird fast ausschließlich an der Ostsee zu Tweets hinzugefügt. Bei #Regen und #Luft hingegen gibt es keine erkennbaren Muster.



**Abbildung 6: Morans-*I* und Getis-Ord  $G_i^*$ -Werte für Naturphänomene.**

Die Namen von ortsverbundenen Organisationen sind ein weiteres Beispiel für Hashtags, die mit bestimmten Orten assoziiert werden können. Auf Abb. 7 werden die Namen oder Namensabkürzungen von 6 deutschen Fußballvereinen gezeigt. Alle diese Hashtags haben signifikante Morans- $I$ -Werte und Getis-Ord  $G_i^*$ -Werte, die um die Vereinsstandorte herum geballt sind.

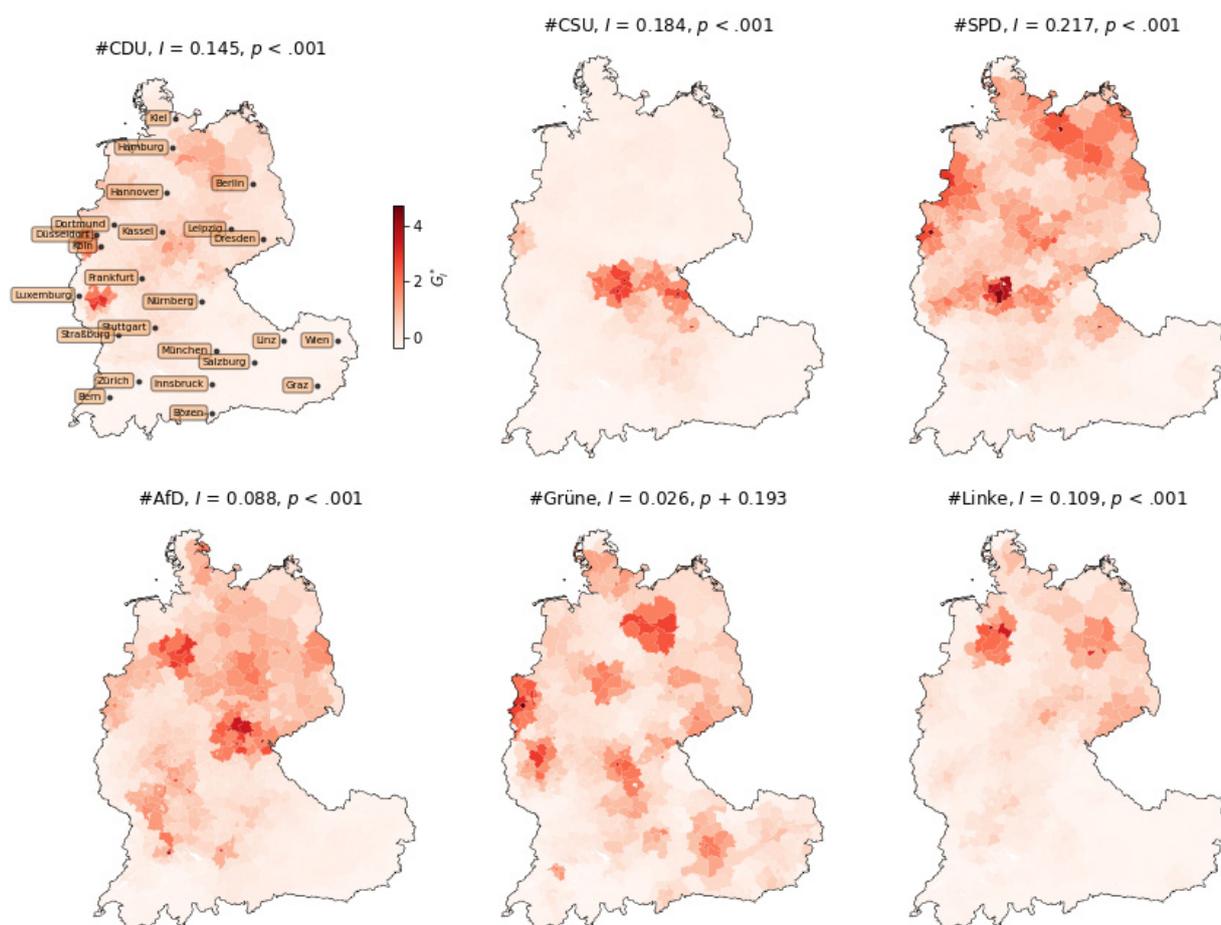


**Abbildung 7: Morans- $I$  und Getis-Ord  $G_i^*$ -Werte für Fußballvereine.**

#### 4.2.2. Hashtags ohne geographischen Bezug

Namen oder Abkürzungen von politischen Parteien zeigen laut ihrer Morans- $I$ -Werte räumliches Clustering auf, sind aber in Bezug auf ihre geographische Verteilung nicht eindeutig interpretierbar, im Gegensatz zu Toponymen- oder Naturphänomene-Hashtags (Abb. 8). Außer dass die Namen deutscher politischer Parteien in Österreich und in der Schweiz kaum getweetet werden, und abgesehen von der Tatsache, dass #CSU relativ hohe Werte erzielt im Norden von Bayern, entsprechen die Verteilungen der nach Hashtag-Relativhäufigkeit berechneten Getis-Ord  $G_i^*$ -Werte nicht Wahlergebnissen oder

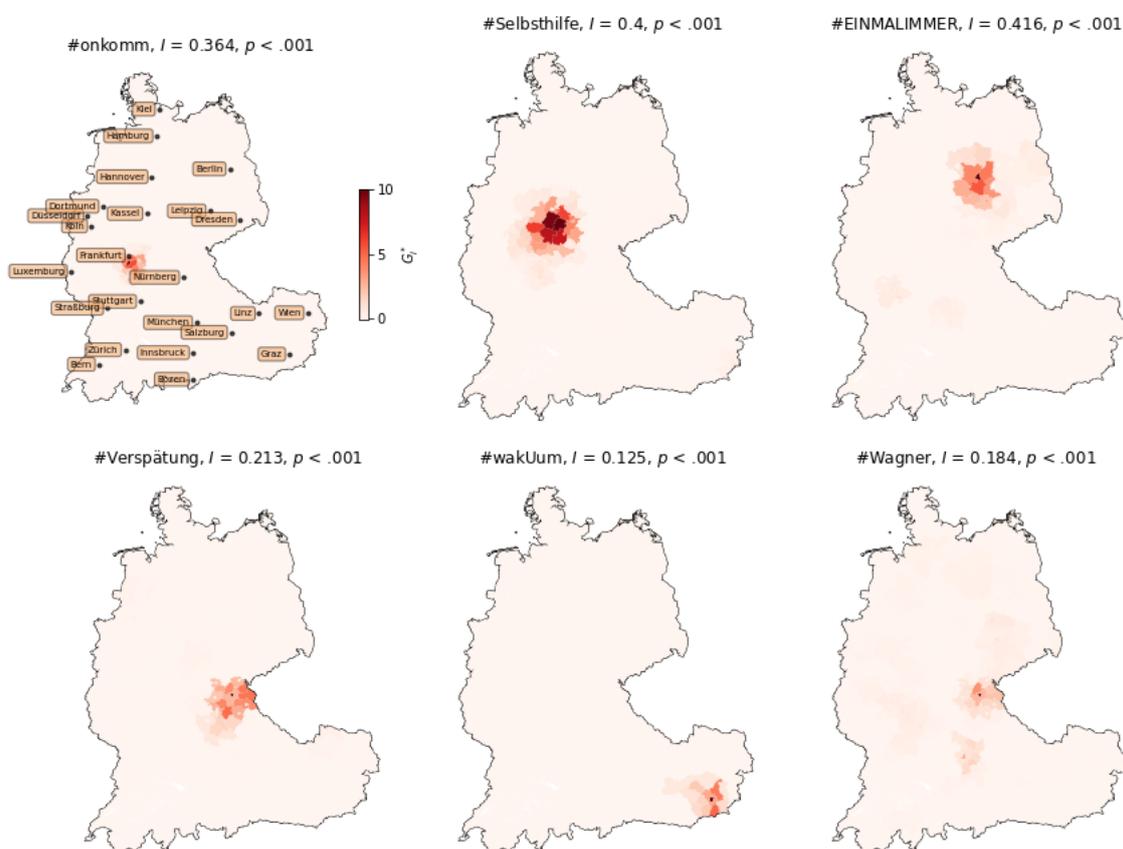
Umfrageergebnissen zur Beliebtheit politischer Parteien. Insofern liefern in diesen Daten Tweetinhalte wie Hashtags keine repräsentative Widerspiegelung aggregierter politischer Präferenzen, zumindest nicht in Bezug auf ihre räumliche Verteilung (vgl. Bastos & Mercea 2018, die eine ähnliche Uneinigkeit feststellten zwischen den Ergebnissen des Brexitreferendums und dessen geographischem Signal in Tweets).



**Abbildung 8: Morans-I und Getis-Ord  $G_i^*$ -Werte für politische Parteien.**

Der Hauptteil der 30 035 häufigen Hashtags hat eher niedrigere Morans-I-Werte, aber es gibt jedoch einige häufig verwendete Hashtags, die trotz des Fehlens eines unmittelbar erkennbaren geographischen Bezugs eine geographische Clusterbildung (und entsprechend erhöhte Morans I-Werte) zeigen, wie auf Abb. 9, welche die Getis-Ord  $G_i^*$  Werte für die Hashtags #onkomm, #Selbsthilfe, #EINMALIMMER, #Verspätung, #wakuUm, und #Wagner zeigt. Obwohl diese Begriffe nicht unmittelbar einem geographischen Ort zugeordnet werden können, zeigt eine manuelle Überprüfung der Inhalte die-

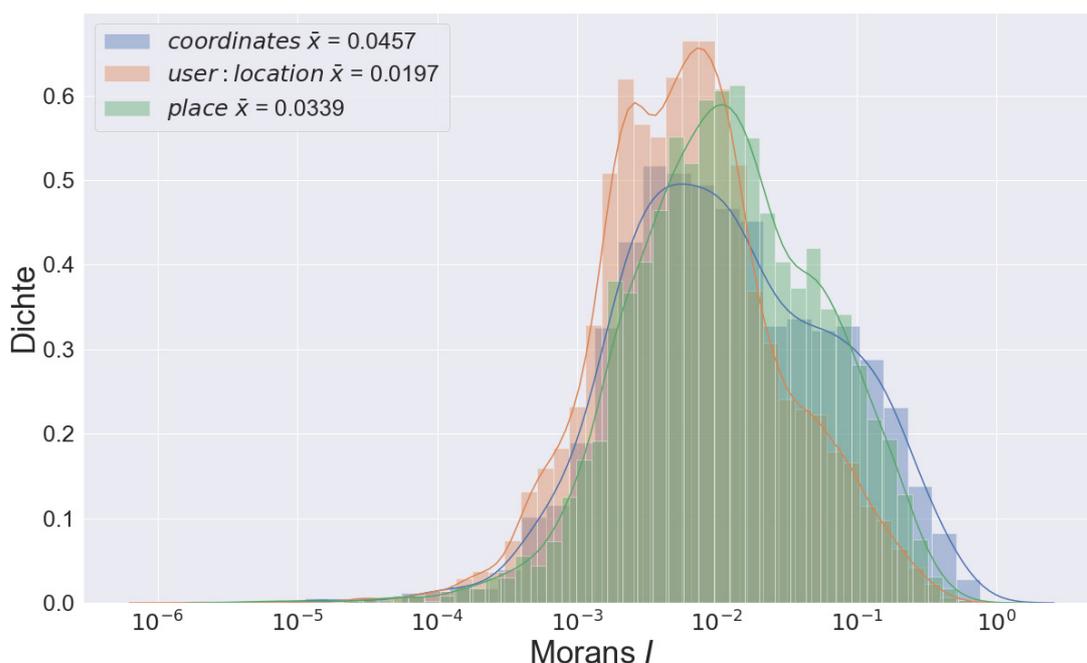
ser Tweets eine Assoziation mit Standorten: #onkomm („Online-Kommunikation“) als Studienfach an einer Darmstädter Hochschule, #Selbsthilfe als häufig benutztes Hashtag eines hessischen Vereins, #EINMALIMMER als Unterstützungsaufwurf von Fans des Fußballvereins 1. FC Magdeburg, #Verspätung als häufig benutztes Hashtag eines auf Zugverspätungsinformation spezialisierten Twiternutzers im Raum Nürnberg, #wakuUm als Name eines Nachtclubs in Graz, und #Wagner als Erläuterung zu den Bayreuther Festspielen. Hashtags, die eine geografische Häufung aufweisen, können somit in drei Kategorien unterteilt werden: Erstens: Hashtags, die spezifisch einen Ort benennen (Toponyme), zweitens: Hashtags, deren semantischer Wert keinen bestimmten Ort bezeichnet, die aber offensichtlich indirekt mit bestimmten Orten assoziiert sind, und drittens: Hashtags, die weder bestimmte Orte bezeichnen noch allgemein bekannt sind, dass sie semantisch direkt mit bestimmten Orten assoziiert sind, die aber dennoch eine indirekte semantische Assoziation mit einer Örtlichkeit haben. In Abschnitt 5 wird diese Unterscheidung ausführlicher erörtert.



**Abbildung 9: Morans-I und Getis-Ord  $G_i^*$ -Werte für häufige Hashtags mit unklarem geographischen Bezug.**

### 4.3. Stellenwert des Hashtag-Standort-Typus

Eine Tendenz zur räumlichen Clusterbildung von Hashtags zeigt sich in erster Linie bei Toponymen oder anderen Namen, die mit bestimmten Orten assoziiert werden können, und weniger bei anderen Arten von Hashtags. Dabei kann jedoch auch die Art der Metadaten, mit denen Tweets zu einem Ort zugeordnet werden, eine Rolle spielen. Die Verteilungen von Morans  $I$ -Werte für Hashtags mit einem Mindestaufkommen von 100 im Korpus zeigen je nach Standortmetadaten-Typus leichte Verschiebungen im Verteilungsprofil (Abb. 10). Hashtags in Tweets mit *user-location*-Angaben haben im Durchschnitt den niedrigsten Morans- $I$ -Wert – diese Tweets sind entweder von Nutzern auf der Twitter-Webseite veröffentlicht oder von Nutzern, die keine genaueren *coordinates*- oder *place*-Angaben gewählt haben bei der Tweetveröffentlichung. Hashtags mit *place*-Information haben einen leicht erhöhten Durchschnittswert, und Hashtags mit genauen GPS-Koordinaten (*coordinates*-Metadaten) haben den höchsten Durchschnittswert. Es besteht die Möglichkeit, dass Benutzer, die Tweets von zu Hause aus schreiben, nicht-toponymische Hashtags verwenden, während Mobilgerätebenutzer möglicherweise unterwegs sind und die Orte dann per Hashtag mitteilen oder kommentieren.



**Abbildung 10: Verteilung von Morans-I-Absolutwerten für Hashtags mit Mindestaufkommen von 100 nach Standortmetadaten-Typus.**

## 5. Diskussion: Vorbehalte, Schlussfolgerungen und Ausblick

In dieser Studie wurde die raumgeographische Verteilung von Hashtags in einem auf den deutschsprachigen Raum orientierten Korpus von über 39 Millionen Tweets und 24 Millionen Hashtags skizziert, das eine Stichprobe der Aktivität des mehrsprachigen, grenzüberschreitenden Kommunikationsnetzwerkes Twitter repräsentiert.

Einige Vorbehalte bezüglich der in der Studie verwendeten Datenerhebungsmethode müssen erwähnt werden. Erstens liegt die Wahrscheinlichkeit hoch, dass die Daten geographische Fehlzuordnungen für einzelne Tweets oder auch für Nutzer enthalten. Aus methodischer Sicht kann es aus verschiedenen Gründen zu Fehlallokationen kommen: Nutzer, die keine Tweets mit *place*- oder *coordinates*-Metadaten veröffentlichen und im Feld *user:location* einen anderen Ort als den tatsächlichen Wohn- oder Aufenthaltsort angegeben haben, werden von der in dieser Studie verwendeten Datenerhebungsmethode falschen Orten zugeteilt. Der in Abschnitt 3 erwähnte Vergleich der *user:location*-Tweets mit aggregierten *coordinates*-Tweets pro Nutzer zeigt allerdings, dass *user:location*-Angaben relativ zuverlässig sind. Eine andere potentielle Fehlerquelle ist das Standortverzeichnis von Twitter selbst: Das Verzeichnis, das von FourSquare, einem auf Standortinformationen spezialisierten amerikanischen Sozialmediendienst entwickelt wurde, enthält hauptsächlich Ortsbezeichnungen in englischer Sprache. Die Suchfunktion, mit der ein Nutzer Standorte finden kann, um sie als *place*-Metadatei zu einem Tweet hinzuzufügen, enthält nur teilweise die deutschsprachigen Ortsnamen europäischer Standorte: Es fehlen z. B. Städtenamen wie Reval, Lemberg, Kolmar, oder Agram.<sup>9</sup> Im Juni 2020 enthält das Standortverzeichnis zudem noch Fehlzuordnungen (Košice, Slowakei, wird als „Košice, Česká republika“ angegeben), und es fehlen überhaupt alle österreichischen Bundesländer als *place*-Möglichkeiten: Eine Suche nach „Tirol“ gibt „Tirol, Trentino-South Tyrol, Italy“ zurück, und „Burgenland“ ergibt keine Treffer. Die möglichen Auswirkungen dieser Fehlerquelle sind allerdings relativ gering, da es in dieser Studie hauptsächlich um Standorte im deutschsprachigen Raum geht, und Tweets nicht zu Bundesländern zugewiesen wurden.

---

<sup>9</sup> Dies gilt auch für Ortsnamen in anderen Sprachen, z. B. Ungarisch, (Kassa = Kaschau = Košice, Bécs = Wien).

Eine andere Überlegung betrifft das Signal von auf Mobilgeräten veröffentlichten Tweets. Tweets von im Untersuchungsgebiet ansässigen Nutzern können sowohl *user:location*- als auch *place*- und *coordinates*-Metadaten haben, aber Tweets von vorübergehenden Besuchern mit Wohnsitz außerhalb des Untersuchungsgebiets können nur die zwei letzteren Kategorien von Standortmetadaten enthalten. Diese Tatsache könnte auch zu den Unterschieden bei den Verteilungen der Morans-*I*-Werte für die Standortmetadaten-Typen beigetragen haben.

Schließlich ist zu beachten, dass für einige Kartenflächen ein großer Teil der Token auf die Twitter-Aktivität einiger weniger Nutzer zurückzuführen sein kann, mit der Folge, dass bei häufiger Verwendung eines bestimmten Hashtags das Signal in den entsprechenden Flächen verstärkt ist im Vergleich zu Signalen aus Flächen mit vielen Nutzern. Eine repräsentative Stichprobe zu erstellen, die sowohl die Bevölkerungszahl als auch die Anzahl von Twiternutzern in den Landkreisen berücksichtigt, obwohl eine zeit- und arbeitsintensive Unternehmung, könnte diese potenzielle Fehlerquelle beseitigen.

Was die zwei Forschungsfragen betrifft, lässt sich wie folgt zusammenfassen. Erstens hat die Tendenz, überhaupt Hashtags in Tweets zu verwenden, kein raumgeographisches Muster in dem untersuchten Gebiet: Laut einer globalen Autokorrelationsanalyse mit Morans *I*-Statistik ist die Hashtagdichte über die 730 Flächen, die das Gebiet umfassen, ungefähr zufällig verteilt. Zweitens haben die meisten laut Morans-*I*-Werte geclusterten Hashtags durch ihre Semantik eine Verbindung zu einem spezifischen Ort, z. B. als Toponym, Eigenname, oder Appellativum eines regional begrenzten Phänomens.

Es lässt sich also feststellen, dass sich Hashtags in Bezug auf die Art und Weise, wie sie den geografischen Raum markieren, ähnlich verhalten wie andere lexikalische Elemente: Nicht nur Elemente, die direkt bestimmte Orte bezeichnen, wie z. B. die Namen von Städten oder Ländern, sondern auch Elemente, die in bestimmten lokalen Kontexten verwendet werden, können Bedeutungen annehmen, die die Lokalität anzeigen. Auf diese Weise werden bestimmte Hashtags (solche, die lokal geclustert sind, aber keine eindeutigen geografischen oder anderen toponomastischen Merkmale bezeichnen) mit Orten assoziiert. Diese Indexikalität und Markierung wird für lokale Gemein-

schaften konstitutiv, wenn sie in globalen Netzwerken wie dem Internet vertreten sind, und kann in gewisser Weise der Art und Weise entsprechen, in der bestimmte lexikalische Elemente mit bestimmten lokalen Dialekten in der traditionellen Dialektologie assoziiert werden (Purschke & Hovy 2019).

Für die Zukunft bieten sich für die korpusbasierte Analyse von Hashtagraumverteilungen mehrere mögliche Forschungsrichtungen an. Weil der thematische Inhalt von Tweets, die Toponym-Hashtags wie z. B. Ländernamen enthalten, stark variieren kann, von Reiseberichten und anderen persönlichen Berichterstattungen zu Kontextualisierung von lokalen Nachrichten oder Bekanntmachung von kommerziellen Inhalten (Raamukar et al. 2016), wäre eine Auswertung der Semantik von raumspezifischen Hashtags aufgrund unterschiedlicher Tweetkategorien wünschenswert, möglichst auf der Grundlage größerer Korpora aus Twitter und anderen Sozialmediendiensten und anhand gängiger Techniken wie distributiver Semantik und Vektorraumrepräsentation (Mikolov et al. 2013; vgl. Ferragina et al. 2015 und Zhang 2019). Somit könnten neue Perspektiven eröffnet werden auf linguistische Fragen wie z. B. das Verfahren, wonach Hashtags lexikalisiert werden, den Status von Hashtags als Eigennamen oder die Korrelation zwischen semantischem und raumgeographischem Abstand (vgl. Eisenstein et al. 2014, Zhang 2019).

Eine weitere Perspektive, die dieser Forschungsansatz unterstützt, ist die Möglichkeit, bei ausreichend großen Korpora eine korpusbasierte Methode zur Analyse von linguistischen Elementen mit begrenzter geographischer Verteilung anzuwenden, wobei die typische Untersuchungsvorgehensweise umgekehrt wird. Anstatt Hashtags auf der Grundlage vordefinierter Kriterien auszuwählen, um sie dann in Bezug auf ihre etymologischen, semantischen, pragmatischen oder sonstigen linguistischen Eigenschaften zu untersuchen, kann ein auf Autokorrelation basierender Ansatz linguistische Formen mit raumgeographischer Spezifität im Voraus identifizieren. Besonders für Hashtags, neuartige Kontextualisierungsmittel mit semantischen, indexikalischen, und onomastischen Eigenschaften, könnte eine solche korpusorientierte Vorgehensweise neue Impulse für die Erforschung von Toponymen und anderen Eigennamen liefern (vgl. Motschenbacher 2020).

Auch wenn es noch zu früh ist, um von „fossilisierten Hashtags“ zu sprechen, deutet die semantische Entwicklung einiger Hashtags, vor allem derjenigen, die nur indirekt einen Ortsbezug aufweisen, darauf hin, dass diese vielseitigen Elemente für bestimmte Orte repräsentativ werden können. [N](#)

STEVEN COATS  
UNIVERSITÄT OULU

## Bibliographie

- Amador Diaz Lopez, Julio et al. 2017. Predicting the Brexit vote by tracking and classifying public opinion using Twitter data, statistics. *Politics and Policy* 8(1): 85–104. <https://doi.org/10.1515/spp-2017-0006>
- Andéhn, Mikael et al. 2014. User-generated place brand equity on Twitter: The dynamics of brand associations in social media. *Place Branding and Public Diplomacy* 10: 132–144. <https://doi.org/10.1057/pb.2014.8>
- Bastos, Marco/Mercea, Dan 2018. Parametrizing Brexit: mapping Twitter political space to parliamentary constituencies. *Information, Communication & Society* 21(7): 921–939. <https://doi.org/10.1080/1369118X.2018.1433224>
- Bastos, Marco/Recuero, Raquel/Zago, Gabriela 2014. Taking tweets to the streets: A spatial analysis of the Vinegar Protests in Brazil. *First Monday* 19(3). <https://doi.org/10.5210/fm.v19i3.5227>
- Bauer, Matthias Johannes/Goetz, Miriam (Hrsg.) 2021. *Der Hashtag als interdisziplinäres Phänomen in Marketing und Kommunikation: Sprache, Kultur, Betriebswirtschaft und Recht*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-32085-0>
- Bernard, Andreas 2021. Theorie des Hashtags. *Der Hashtag als interdisziplinäres Phänomen in Marketing und Kommunikation: Sprache, Kultur, Betriebswirtschaft und Recht*, hrsg. von Matthias Johannes Bauer/Miriam Goetz. 1–28. Wiesbaden: Springer. [https://doi.org/10.1007/978-3-658-32085-0\\_1](https://doi.org/10.1007/978-3-658-32085-0_1)
- Cliff, Andrew D./Ord, J. Keith 1973. *Spatial autocorrelation*. London: Pion.
- Cliff, Andrew D./Ord, J. Keith 1981. *Spatial processes: Models and applications*. London: Pion.
- Coats, Steven 2019. Online language ecology: Twitter in Europe. *Building computer-mediated communication corpora for sociolinguistic analysis*, hrsg. von Egon Stemle/Ciara Wigham. 73–96. Clermont-Ferrand: Presses Universitaires Blaise Pascal.
- Coats, Steven 2020. Articulation rate in American English in a corpus of YouTube videos. *Language and Speech* 63(4): 799–831. <https://doi.org/10.1177/0023830919894720>
- Dang-Anh, Mark/Einspänner, Jessica/Thimm, Caja 2013. Kontextualisierung durch Hashtags: die Mediatisierung des politischen Sprachgebrauchs im Internet. *Öffentliche Wörter: Analysen zum öffentlich-medialen Sprachgebrauch*, hrsg. von Haj Diekmannshenke/Thomas Niehr. 137–159. Stuttgart: Ibidem.
- De Cock, Barbara/Pedraza, Andrea Pizarro 2018. From expressing solidarity to mocking on Twitter: Pragmatic functions of hashtags starting with #jesuis across languages. *Language in Society* 47(2): 197–217. <https://doi.org/10.1017/S0047404518000052>
- Eisenstein, Jacob et al. 2014. Diffusion of lexical change in social media. *PLOS ONE* 9(11), e113114. <https://doi.org/10.1371/journal.pone.0113114>
- Esmukov, Kostya et al. 2020. *Geopy*. <https://github.com/geopy/geopy>
- Ferragina, Paolo/Piccinno, Francesco/Santoro, Roberto 2015. On analyzing hashtags in Twitter. *Proceedings of the 2015 International Conference on Weblogs and Social Media (ICWSM)*.

- 110–119. Palo Alto: AAAI. <https://doi.org/10.1609/icwsm.v9i1.14584>
- Georgiadou, Elena/Angelopoulos, Spyros/Drake, Helen 2020. Big data analytics and international negotiations: Sentiment analysis of Brexit negotiating outcomes. *International Journal of Information Management* 51. <https://doi.org/10.1016/j.ijinfomgt.2019.102048>
- Getis, Arthur 2010. Spatial autocorrelation. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, hrsg. von Manfred M. Fischer/Arthur Getis. 255–278. Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-03647-7>
- Grčar, Miha et al. 2017. Stance and influence of Twitter users regarding the Brexit referendum. *Computational Social Networks* 4(6). <https://doi.org/10.1186/s40649-017-0042-6>
- Goebel, Hans 1982. *Dialektometrie: Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*. Vienna: Verlag der Österreichischen Akademie der Wissenschaft.
- Grieve, Jack 2016. *Regional variation in written American English*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9781139506137>
- Grieve, Jack et al. 2019. Mapping lexical dialect variation in British English using Twitter. *Frontiers in Artificial Intelligence*. <https://doi.org/10.3389/frai.2019.00011>
- Grieve, Jack/Speelman, Dirk/Geeraerts, Dirk 2011. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change* 23: 193–221. <https://doi.org/10.1017/S095439451100007X>
- Han, Bo/Cook, Paul/Baldwin, Timothy 2014. Text-based Twitter user geolocation prediction. *Journal of Artificial Intelligence Research* 49: 451–500. <https://doi.org/10.1613/jair.4200>
- Hecht, Brent et al. 2011. Tweets from Justin Bieber’s heart: The dynamics of the location field in user profiles. In: *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, Association for Computing Machinery: 237–246. <https://doi.org/10.1145/1978942.1978976>
- Hochmair, Hartwig H./Juhász, Levente/Cvetojevic, Sreten 2018. Data quality of points of interest in selected mapping and social media platforms. *Progress in Location Based Services 2018*, eds. In: Peter Kiefer, Haosheng Huang, Nico Van de Weghe & Martin Raubal. 293–313. Cham: Springer. [https://doi.org/10.1007/978-3-319-71470-7\\_15](https://doi.org/10.1007/978-3-319-71470-7_15)
- Hübl, Franziska et al. 2017. Analyzing refugee migration patterns using geo-tagged Tweets. *International Journal of Geo-Information* 6(10). <https://doi.org/10.3390/ijgi6100302>
- Hürlimann, Manuela et al. 2016. A Twitter sentiment gold standard for the Brexit referendum. *SEMANTiCS 2016: Proceedings of the 12th International Conference on Semantic Systems September 2016*. 193–196. <https://doi.org/10.1145/2993318.2993350>
- Johnson, Isaac L. et al. 2016. The geography and importance of localness in geotagged social media. In: *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 515–526. New York: Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858122>

- Kamath, Krishna Y. et al. 2013. Spatio-temporal dynamics of online memes: A study of geo-tagged Tweets. In *WWW, 13: Proceedings of the 22nd international conference on World Wide Web*. New York: Association for Computing Machinery. <https://doi.org/10.1145/2488388.2488447>
- Lameli, Alfred 2013. *Strukturen im Sprachraum: Analysen zur arealtypologischen Komplexität der Dialekte in Deutschland*. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783110331394>
- Laylavi, Farhad/Rajabifard, Abbas/Kalantari, Moshen 2016. A multi-element approach to location inference of Twitter: A case for emergency response. *International Journal of Geo-Information* 5(56). <https://doi.org/10.3390/ijgi5050056>
- Lazer, David et al. 2009. Computational social science. *Science* 323(5915): 721–723. <https://doi.org/10.1126/science.1167742>
- Leetaru, Kalev H. et al. 2013. Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday* 18(5–6). <https://doi.org/10.5210/fm.v18i5.4366>
- Mikolov, Tomas/Yih, Wen-tau/Zweig, Geoffrey 2013. Linguistic regularities in continuous space word representations. In *Proceedings of HLT-NAACL 13*: 746–751.
- Moran, Patrick A. P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17–23.
- Motschenbacher, Heiko 2020. Corpus linguistic onomastics: A plea for a corpus-based investigation of names. *Names* 68(2): 88–103. <https://doi.org/10.1080/00277738.2020.1731240>
- Murzintcev, Nikita/Cheng, Changxiu 2017. Disaster hashtags in social media. *International Journal of Geo-Information* 6(7). <https://doi.org/10.3390/ijgi6070204>
- Nerbonne, John/Kleiweg, Peter 2007. Toward a dialectological yardstick. *Journal of Quantitative Linguistics* 14(2–3): 148–166. <https://doi.org/10.1080/09296170701379260>
- Nguyen, Dong et al. 2016. Computational sociolinguistics: A survey. *Computational Linguistics* 42(3): 537–593. [https://doi.org/10.1162/COLI\\_a\\_00258](https://doi.org/10.1162/COLI_a_00258)
- Page, Ruth 2012. The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication* 6(2): 181–201. <https://doi.org/10.1177/1750481312437441>
- Pickl, Simon/Pröll, Simon 2019. Ergebnisse geostatistischer Analysen arealsprachlicher Variation im Deutschen. *Deutsch: Sprache und Raum - Ein internationales Handbuch der Sprachvariation* (= HSK 30.4), hrsg. von Joachim Herrgen/Jürgen Erich Schmidt. 861–878. Berlin/Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110261295-032>
- Purschke, Christoph/Hovy, Dirk. (2019). Lörres, Möppes, and the Swiss: (Re) Discovering regional patterns in anonymous social media data. *Journal of Linguistic Geography* 7(2): 113–134. <https://doi.org/10.1017/jlg.2019.10>
- Raamukar, Aravind/Pang, Natalie/Foo, Schubert 2016. When countries become the talking point in microblogs: Study on country hashtags in Twitter. *First Monday* 21(1–4). <https://doi.org/10.5210/fm.v21i1.6101>
- Rey, Sergio J./Anselin, Luc 2007. PySAL: A Python library of spatial analytical methods. *Review of Regional Studies* 37: 5–27. <https://doi.org/10.52324/001c.8285>

- Roesslein, Joshua. 2015. *Tweepy* (Python library). <https://github.com/tweepy/tweepy>
- Schepanski, Kim/Seppelfricke, Thomas 2021. Twitter-Hashtags am Beispiel Klimawandel: Eine technologiegestützte psychologische Analyse. *Der Hashtag als interdisziplinäres Phänomen in Marketing und Kommunikation: Sprache, Kultur, Betriebswirtschaft und Recht*, hrsg. von Matthias Johannes Bauer/Miriam Goetz. 93–112. Wiesbaden: Springer. [https://doi.org/10.1007/978-3-658-32085-0\\_6](https://doi.org/10.1007/978-3-658-32085-0_6)
- Schudy, Fabian 2021. Hashtags im Sprachgebrauch: Linguistische Betrachtungen zur Funktion. *Der Hashtag als interdisziplinäres Phänomen in Marketing und Kommunikation: Sprache, Kultur, Betriebswirtschaft und Recht*, hrsg. von Matthias Johannes Bauer/Miriam Goetz. 29–48. Wiesbaden: Springer. [https://doi.org/10.1007/978-3-658-32085-0\\_2](https://doi.org/10.1007/978-3-658-32085-0_2)
- Séguy, Jean 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35: 335–357.
- Shibuya, Yuya 2017. Mining social media for disaster management: Leveraging social media data for community recovery. In: *Proceedings of the 2017 IEEE International Conference on Big Data*. 3111–3118. New York: IEEE. <https://doi.org/10.1109/Big-Data.2017.8258286>
- Sibler, Pius 2011. *Visualisierung und geostatistische Analyse mit Daten des Syntaktischen Atlas der Deutschen Schweiz (SADS)*. Zürich: Geographisches Institut der Universität Zürich, Masterarbeit.
- Sloan, Luke/Morgan, Jeffrey 2015. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. *PLOS ONE* 10/11: e0142209. <https://doi.org/10.1371/journal.pone.0142209>
- Squires, Lauren 2015. Twitter: Design, discourse, and implications of public text. *The Routledge Handbook of Language and Digital Communication*, hrsg. von Alexandra Georgakopoulou/Tereza Spilioti. 239–256. London/New York: Routledge. <https://doi.org/10.4324/9781315694344>
- Stoeckle, Philipp 2016. Horizontal and vertical variation in Swiss German morphosyntax. *The future of dialects*, hrsg. von Marie-Hélène Côté/Remco Knooihuizen/John Nerbonne. 195–215. Berlin: Language Science Press. <https://doi.org/10.17169/langsci.b81.150>
- Thimm, Caja/Einspänner, Jessica/Dang-Anh, Mark 2012. Twitter als Wahlkampfmedium. *Publizistik* 57(3): 293–313. <https://doi.org/10.1007/s11616-012-0156-7>
- Wikström, Peter 2014. #srynotfunny: Communicative functions of hashtags on Twitter. *SKY Journal of Linguistics* 27: 127–152.
- Zappavigna, Michele 2012. *Discourse of Twitter and social media: How we use language to create affiliation on the web*. London, New York: Continuum.
- Zappavigna, Michele 2018. *Searchable talk: Hashtags and social media meta-discourse*. London: Bloomsbury.
- Zhang, Yang 2019. Language in our time: An empirical analysis of hashtags. In *Proceedings of the 2019 World Wide Web Conference (WWW '19)*. 2378–2389. New York, Association for Computing Machinery. <https://doi.org/10.1145/3308558.3313480>