

BACHELOR THESIS  
COMPUTING SCIENCE



RADBOD UNIVERSITY

---

**The effects of search languages on  
multilingual online news  
consumption**

---

*Author:*

Matti Eisenlohr

S1000794

m.eisenlohr@student.ru.nl

*First supervisor/assessor:*

dr. ir., Eelco Herder

e.herder@cs.ru.nl

*Second assessor:*

prof. dr. ir., Djoerd Hiemstra

djoerd.hiemstra@ru.nl

June 28, 2020

## **Abstract**

In 2020, a global pandemic hit the world in the form of COVID-19, better known as the coronavirus. In times like these following the latest news is very important to keep up with what's happening in the world and how one's own country reacts to the pandemic, which could impose severe restrictions on one's personal everyday life. In recent years news consumption has increasingly shifted towards the internet and websites like Google News, where an algorithm determines which articles one sees.

As many Europeans are multilingual and the coronavirus was dominating headlines all over the world, we want to determine if the usage of different search languages (the language of the query and language setting on Google News) yields a wider variety of content and opinions compared to the usage of just one single search language.

In a data study of 11,264 headlines in German, English and Dutch, collected from eight predefined search terms during a period of about three weeks, we determined, that, while all three search languages mainly feature mainstream news outlets and, with a few exceptions, popular opinions on news topics, which were often similar between languages, one does get a wider variety of news stories when using multiple search languages, mostly topics that are relevant to specific regions like the lockdown measures introduced by the government in a certain country. This provides a user with more context on certain news topics, however the content is still largely determined by mainstream news outlets (large and popular regional and national newspapers and broadcasters).

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Preliminaries</b>	<b>5</b>
2.1	Filter bubble . . . . .	5
2.2	German media . . . . .	6
2.3	Dutch media . . . . .	7
2.4	English media . . . . .	8
2.4.1	UK media . . . . .	9
2.4.2	US media . . . . .	10
2.5	Popular opinions on the examined topics . . . . .	11
2.5.1	Measures against the coronavirus - Relaxation or further lockdown? . . . . .	11
2.5.2	Coronavirus Tracking Apps . . . . .	12
2.5.3	Donald Trump . . . . .	12
<b>3</b>	<b>Related Work</b>	<b>13</b>
3.1	Filter bubble . . . . .	13
3.2	Search languages . . . . .	15
<b>4</b>	<b>Research</b>	<b>16</b>
4.1	Data collection . . . . .	16
4.2	Data analysis methods . . . . .	17
4.3	Observations . . . . .	18
4.3.1	Relaxations of measures . . . . .	19
4.3.2	Measures . . . . .	22
4.3.3	Tracking App . . . . .	24
4.3.4	Face masks . . . . .	26
4.3.5	Donald Trump . . . . .	28
4.3.6	EU . . . . .	30
4.3.7	Italy . . . . .	32
4.3.8	WHO . . . . .	34

<b>5</b>	<b>Discussion and Conclusions</b>	<b>38</b>
5.1	Summary and Discussion of Observations . . . . .	38
5.2	Limitations . . . . .	39
5.3	Conclusions . . . . .	40
5.4	Further discussion . . . . .	42
5.5	Acknowledgements . . . . .	43
<b>A</b>	<b>Python Script for Data Collection</b>	<b>49</b>
<b>B</b>	<b>Python Script for analyzing the headline files</b>	<b>52</b>
<b>C</b>	<b>Python Script for analyzing the topic codes</b>	<b>62</b>

# Chapter 1

## Introduction

In the last decades news consumption has shifted away from traditional mediums like newspapers, radio and television towards online news consumption, with studies showing that among Europeans under the age of 50 most get their news online [24].

By far the most popular way to consume news online is by directly accessing the websites of news outlets, with search engines being the second most popular way [16], with by far the most popular search engine being Google [4]. As opposed to the websites of news outlets, search engines use personalized algorithms to determine what news articles are being shown to the individual, which according to researchers like Eli Pariser leads to the creation of so-called "filter bubbles" (more on that in the next chapter) [28].

Most people consume news primarily in their first language, but many multilingual users also tend to use multiple languages for news consumption, especially, when they are using search engines [21]. With more than 65 percent of EU citizens now being able to speak at least one language besides their own mother tongue [7], multilinguals make up a significant part of the demographic within the European Union. During this study, we want to answer the (research) question:

Do multilinguals get a wider variety of content and perspectives if they use multiple search languages while browsing for news online on Google News?

Especially in the context of the global coronavirus pandemic, the information we receive is essential to our daily lives, since that information influences how serious one takes the pandemic, how one handles the crisis and how one views the measures taken by their government to slow down the spread of the coronavirus. So access to more diverse perspectives can help form a more well-informed opinion on current news topics.

It is reasonable to expect, that each language would feature news stories that are relevant to the country where the corresponding article was published, however we are also interested in seeing if the search language acts

as a filter for perspectives on certain news topics, if the perspectives found are different between search language and if they are, if it is just a representation of what opinions are popular in a certain country and which one are not. The effect of search languages in this context has not been scientifically examined yet (see also chapter 3).

To gain an insight into how the output of Google News differs per language, we are going to collect headlines for predefined search terms in German, English and Dutch over a period of about three weeks and examine the sources, the news stories presented and if they present a perspective, either positive or negative, on the subject. The languages are limited to the languages I am able to speak and read fluently, as it does not make sense to do qualitative analysis of data I am unable to apprehend.

Before we delve into the study, we are going to examine the concept of the filter bubble and also take a look at popular news sources online in Germany, the Netherlands, the United Kingdom and the United States and the public opinion in these countries on our predefined topics in chapter 2 to get the necessary context for this research.

After that, we are going to look at previous scientific work on the topics of the filter bubble and the topic of search languages in chapter 3 to establish further scientific context for the study.

Then, in chapter 4 the methods of data collection and data analysis will be explained and the observations made during the analysis of the data will be explored.

Finally, chapter 5 will discuss the observations from chapter 4 and draw conclusions from them to answer the research question.

## Chapter 2

# Preliminaries

Before diving into the previous research conducted on the topics of the filter bubble and search languages, it is important to explain some concepts and necessary knowledge needed for this study.

### 2.1 Filter bubble

A crucial concept for this research is the filter bubble, a term that was originally coined by internet activist Eli Pariser [6]. He used the term to describe the intellectual isolation of users, which can result from personalized searches and algorithms [6], that present users with content that matches their own views, creating an "echo chamber" [20].

There are four basic mechanisms that Eli Pariser's concept is built on:

1. The personalization of the content selection presented to a user. This personalization and its degree is arguably the most important mechanism, since political filter bubbles depend on a high degree of personalization [20].
2. There is low to no overlap with the search results of other groups and their filter bubbles.
3. The type of content being displayed to the user, since only controversial topics are actually problematic for society when it comes to filter bubbles.
4. The individual's isolation from different sources of information, that might have a different perspective on controversial political topics [20]

## 2.2 German media

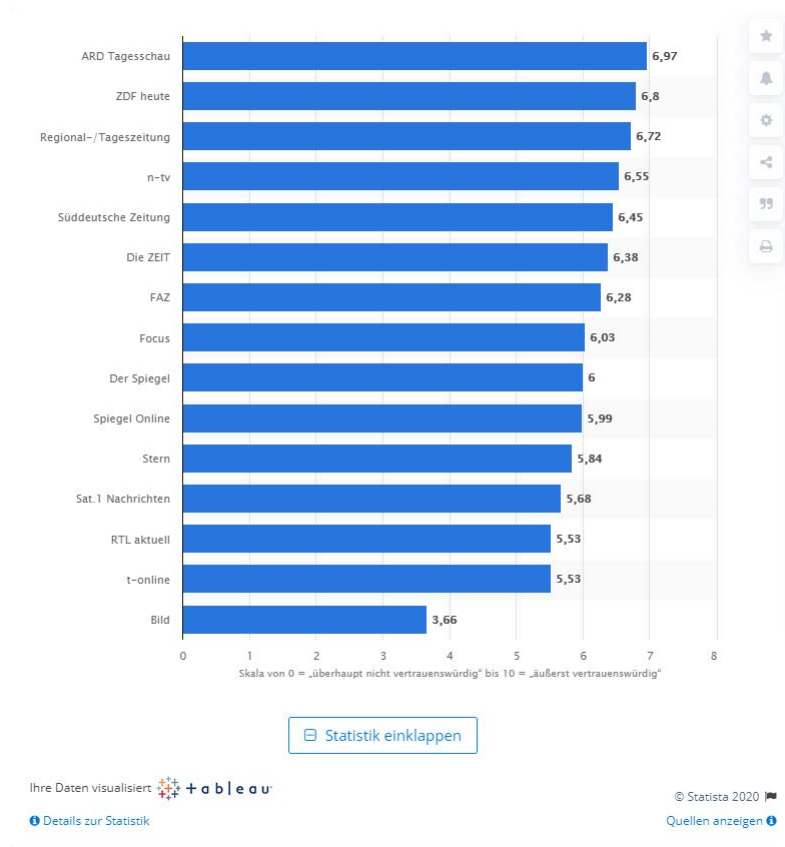


Figure 2.1: The most trusted news sources in Germany 2019 on a scale from 0 = 'completely untrustworthy' to 10 = '100% trustworthy' (Source: Statista [42])

From figure 2.1, we can see that the German population generally trusts its media outlets with moderate scepticism and has the highest amounts of trust in the publicly funded media (ARD and ZDF) and regional news sources. Daily or weekly mainstream newspapers like Süddeutsche Zeitung, Die ZEIT, Der Spiegel or FAZ are also generally trusted, while tabloids like BILD or private television stations like RTL and Sat.1 are less trusted, though BILD still receives a lot of traffic on its website [33]. A list of the 50 most popular German news websites in 2018 can also be found under citation [33] and in figure 2.2 the 15 most popular news websites in March 2020 can be found.



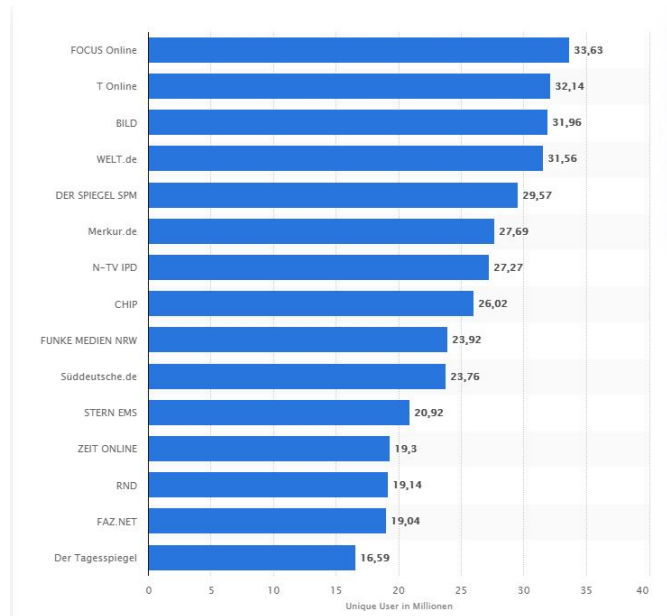


Figure 2.2: The 15 most popular news websites in Germany in March 2020 (Source: Statista [11])

General trust in the media is relatively high, although there is a visible polarization on the topics the AfD (right-wing populist party in Germany) and Islam leading to about a quarter of the population being generally distrustful of the media as opposed to about half of the population generally trusting the media [32].

## 2.3 Dutch media

From figure 2.3 we can see, that non-newspaper based sources like Tweakers and Dumpert are much more popular than in Germany but overall mainstream newspapers and broadcasting companies are still the most popular news sources [2], with the most trusted sources being NPO (publicly funded broadcast) and NU.nl [24].

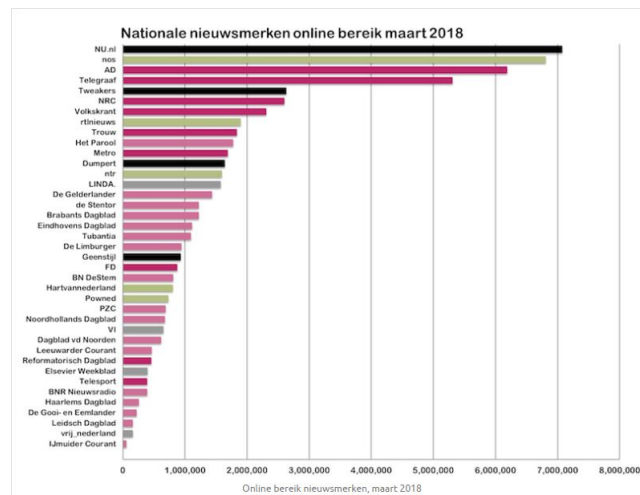


Figure 2.3: The most popular news sources in the Netherlands in March 2018 (Source: Stimuleringsfonds voor de Journalistiek [2])

Even though the Dutch media have been criticized for sensationalizing and exaggerating conflicts in order to get more attention, with even the mainstream 'quality media' like NRC, the NOS and the Volkskrant having been accused of such practices [39], trust in the media generally remains high in the Netherlands as they are still seen as largely unpartisan [29] (see also figure 2.4)



Enkele details over Nederland uit het Digital News Report 2019 van het Reuters Institute.

Figure 2.4: Trust in the Dutch media (Source: Villamedia [29])

## 2.4 English media

English language media is published all over the world but the two countries with the most popular media in the English language are the United States and the United Kingdom.

### 2.4.1 UK media

In the UK, the BBC is by far the most popular news outlet on TV as well as online [27]. As can be seen in figure 2.5, just as in Germany and the Netherlands, the websites of mainstream newspapers are among the most popular sources of news as well, but a noticeable difference lies in the number of tabloids or tabloid-esque websites such as the Daily Mail, The Sun, BuzzFeed, LADBible and Metro, which appear to be far more popular than in those countries [27], with many tabloids being among the most popular newspapers in the UK in general, seemingly enjoying a similar popularity as regular quality papers such as The Guardian, The Times or Financial Times [12].

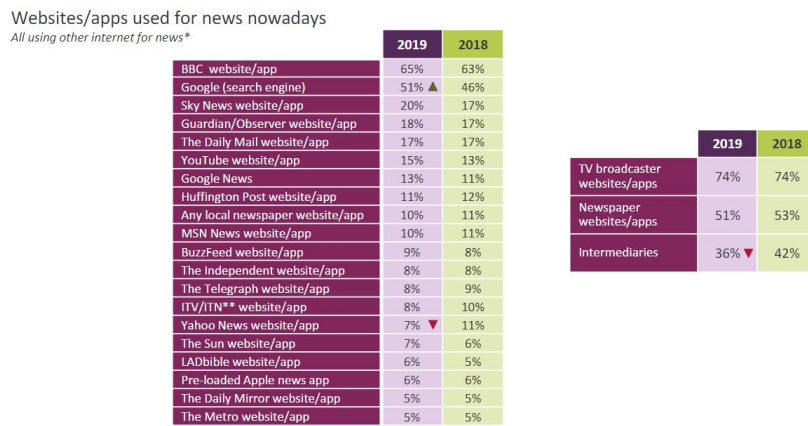


Figure 2.5: Most popular news sources online in the UK as of 2019 (Source: Office of Communication [27])

The trust in news outlets is also more divisive than in Germany and the Netherlands. While sources like the BBC, ITV seem to be well trusted, sources like The Guardian appear more controversial and especially tabloids such as the Daily Mail are deemed untrustworthy by many despite their popularity (see also figure 2.6) [43]. There seems to also be a higher political divide when it comes to the trust in certain news outlets, e.g. the Daily Mail is trusted more among people on the right side of the political spectrum, while The Guardian is trusted more on the left side of the political spectrum [43].

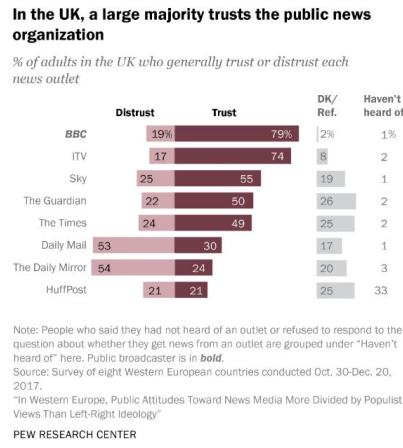


Figure 2.6: Trust in news outlets in the UK in 2017 (Source: Bond/Pew Research Center [43])

## 2.4.2 US media

In the United States, trust in the media seems to be lower than in the other examined countries, with just 41% of Americans stating in a Gallup survey in 2019, that they have a good amount of trust in the media [3]. As can be seen in 2.7, there also seems to be a large partisan divide, with Democrats trusting the media far more than republicans (69% as opposed to 15%).

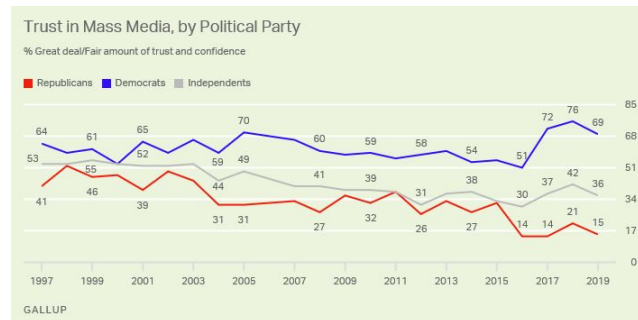


Figure 2.7: Partisan divide in the trust in mass media in the United States (Source: Gallup [3])

When it comes to the most popular news outlets, the websites of mainstream newspapers such as The New York Times, The Washington Post and The Guardian, like in the other examined countries, are very popular, but cable news and their websites like CNN, FOX News and NBC News appear to be much more relevant than in other countries, as well as search engines like Google News and Yahoo News (see figure 2.8)[41].

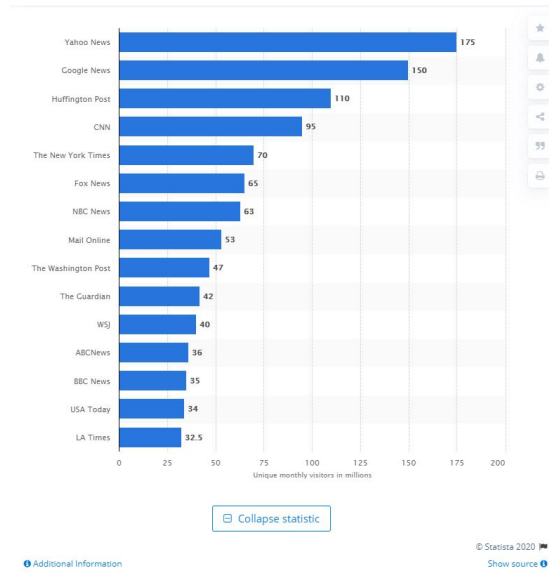


Figure 2.8: Most popular news websites in the US in February 2020 (Source: Statista [41])

## 2.5 Popular opinions on the examined topics

In order to make reasonable conclusions on the results of the research, it is helpful to know the popular opinions in late April/early May 2020 on the examined topics in Germany, the Netherlands, the United States and the United Kingdom. These topics are the lockdown measures against the coronavirus and the relaxation of those measures, tracking apps, Donald Trump (more on the selection of topics in chapter 4). I was unfortunately unable to find helpful polls for the World Health Organization and the European Union apart from one German poll showing that a majority of Germans (48 percent) were not satisfied with the EU’s response to the coronavirus pandemic [14] and one American article that showed that around 55 percent US voters approved the WHO’s handling of the pandemic with 30 percent disapproving [40].

### 2.5.1 Measures against the coronavirus - Relaxation or further lockdown?

In Germany, the people generally approved of the lockdown measures that their government has taken to prevent the spread of the coronavirus, specifically of the social distancing rules [14] and the face mask requirement [30]. In the weeks, where the headlines were collected, in late April/early May of 2020, Germany started relaxing some of those measures. These relaxations proved to be somewhat controversial. Even though the majority of

the population approved of the relaxations the government implemented, a significant portion of almost 40 percent still thought that this came too early [13].

Similarly, in the UK, people generally approved of the governments handling of the coronavirus situation and subsequent lockdown [25]. However, public opinion was more opposed to the relaxation of those measures than in Germany, since many people felt that reasonable conditions for a relaxation have not yet been met [25]. Face mask requirements were also supported by a majority of the population [23].

In the United States, a majority also supported lockdowns and social distancing and generally opposed relaxations of such as they felt the situation was not sufficiently under control [5]. There seems to be a partisan divide here however, with significantly more Republicans being in favor of relaxation than Democrats [5]. Restrictions such as face mask requirements were generally supported [34].

The people of the Netherlands on the other hand seemed to be generally in favor of a relaxation of lockdown measures [38], though it seemed like there were not as many professional polls as for the other countries. Just as in the other countries, face mask requirements were supported by the majority of the population [36].

### **2.5.2 Coronavirus Tracking Apps**

Apps, that could track infected people and notify other people that got into contact so that they can quarantine themselves, seemed to be rather controversial in Germany, with in late April 2020 47 percent of Germans in a poll by the ZDF/Forschungsgruppe Wahlen claiming they would use such an app and 42 percent claiming they would not use it [14]. These apps are similarly controversial in the UK with just about 50 percent willing to use such an app [31].

In the US about 60 percent of people were opposed to the idea of getting a Coronavirus tracking/tracing app [37]. I did not find a similar poll for the Netherlands, however there seemed to be a big concern about privacy in the context of such apps [22].

### **2.5.3 Donald Trump**

Donald Trump was very unpopular in Germany [10] and the UK [15]. In the US a summary of a variety of polls ranks his approval rating as of May 2020 was at about 43 percent, with a disapproval rating of about 53 percent, showing a disapproving tendency [35], but nevertheless he is much more controversial and popular in the United States than elsewhere .

## Chapter 3

# Related Work

Before the start of our study, it is important to look at past research concerning the topics of the filter bubble and the use of language on search engines, particularly when searching for news to get an overview of existing knowledge on these fields that can be built on and possibly expanded.

### 3.1 Filter bubble

Several studies on the filter bubbles have been conducted in recent years. One example is a study by Flaxman, Goel and Rao from 2016, which examined the online news consumption of 50,000 US users, that regularly read news online [16]. The study examined two opposing hypotheses on the effects of recent technological developments on the news consumption of these users.

The first hypothesis assumes that, given more options than before, users will just read content that "accords with their previously held beliefs" (Flaxman, Goel and Rao 2016, p.299) [16]. This assumption is supported by other studies such as Iyengar and Hahn (2009), in whose controlled experiment conservative subjects tended to choose conservative leaning news outlets such as FOX News while avoiding liberal leaning outlets like CNN and NPR while liberal subjects chose liberal leaning outlets while avoiding conservative outlets, not only for controversial news topics, but also for other news categories such as crime and travel [19].

The other hypothesis suggests that an increased number of options will result in increased exposure to diverse and different ideas [16], a hypothesis supported by studies by Goel, Mason and Watts (2010), according to which a lot of connections on social media are between individuals with opposing political beliefs and opinions, allowing for differing views to be seen by users [17], and Messing and Westwood (2012), who concluded that people often view content that has been endorsed by their social circles, leading to a less one-sided exposure [26].

During the study a moderate degree of ideological segregation was observed, with the effect being stronger for opinion pieces than for descriptive news and higher for social media users than for people accessing their news directly from the websites of news outlets (for both descriptive news and opinion pieces) [16]. The reason for this overall moderate degree of ideological segregation found was, that opinion pieces, which tended to have a higher degree, made up only six percent of the overall consumption (even when not taking apolitical topics into account) and subjects mainly accessed their news via the news outlets directly, mostly mainstream news outlets, which tended to be less divided in their political opinions compared to social media and search engines [16]. Search engines in particular were found to have the highest degree of ideological segregation (especially for opinion pieces), with the authors speculating that it could be either due to the personalization of web searches (as suggested by Eli Pariser in 2011 [28]), query formulation having a direct influence on the leanings of the results or users tending to choose outlets aligning with their views over other sources [16]. This establishes search engines as a potential source where news are filtered to a specific user, making it the subject of our study (particularly Google as the by far most popular search engine with a market share of 87.35% as of January 2020 [4]).

A study, that specifically focused on the filter bubble effect on Google, was conducted by Krafft, Gamer and Zweig in 2016, which analyzes the search results of over 1,500 participants from Germany, who searched for German political parties and politicians every four hours [20]. The results of participants on Google were mostly similar, however participants that used a different search language got almost completely different results, with participants mostly getting results in their native language and search results being influenced by search language and location [20]. The paper shows that the search language can be a factor which influences the search results, the effects of which might be interesting to explore, which our research wants to do, with the paper even naming the search language explicitly as an important filter to potential search results [20], though that by itself is not enough to conclude that the search language does create its own filter bubble.

The filter bubble effect on Google News was examined by German researchers Haim, Graefe and Brosius in 2017 [18]. They found that the hypothesis, that Google News creates a personalized filter bubble, which blocks out certain information and perspective, could not be supported [18]. However they did find an overrepresentation of certain news outlets on Google News over other popular news outlets [18].



## 3.2 Search languages

Past research on search engines and languages has mostly focused which languages are being used while searching and why. It has been shown, that, although multilingual users do make use of more than one language online, their native language remains the preferred choice for most users when it comes to consuming news, especially for native speakers of English, who used different languages than their first language significantly less than people, whose first language is not English [21].

Users primarily used a non-native language, when there was a lack of available information in their first language, they perceived the quality of information to be higher in different language or they wanted to check the correctness of phrases in a foreign language, with the search language only being switched during a search if the desired information has not been found in the search language originally used [1]. Most participants in those studies based their language choice on comfort, regional relevance, language preference for specific topics (e.g. for topics concerning technology most participants preferred to use English) and availability of information [21].

No research however has been done yet which examined the search languages of users in the context of filter bubbles, examining if the content in different languages convey a different picture of controversial news topics, thus creating precisely the knowledge gap, that we want to examine during this study.

## Chapter 4

# Research

The aim of this study is to compare headlines from Google News on current news topics in three different languages (German, English and Dutch) in order to measure differences and overlap of topics and perspectives in news stories presented and perspectives on the different topics. I conducted this study by myself, so I restricted it to precisely the three languages I am able to speak fluently.

### 4.1 Data collection

During the data collection phase, I collected the headlines from `news.google.de` for German news, `news.google.co.uk` for English news and `news.google.nl` for Dutch news using previously defined search terms. The search terms were chosen on what I personally perceived as current news topics that were relevant internationally at the time, so a certain bias is given as my own perception of what is a current news topic played the decisive role in choosing the search terms.

My friend Egidius Mysliwietz helped me on 24 April 2020 with a script that automatically collects headlines from Google News and stores them in `.txt` files for each search term and language. The most recent version of this script can be found in the appendix A and on Github <sup>1</sup>. This script takes the first 25 headlines with source and timestamp of the publishing time of the article for each search term and adds all new articles (meaning articles that have not occurred previously) to the corresponding `.txt` file. However some articles still appeared multiple times when they were updated and their timestamp thus changed. I restricted it to the first 25 headlines because the headlines became increasingly less relevant and I assumed that a regular user would not scroll much further.

The data was collected from 24 April 2020 to 12 May 2020. The search terms used were (in English) 'corona mask requirement' (German: 'corona

---

<sup>1</sup><https://github.com/emysliwietz/newsscraeper>

maskenpflicht'; Dutch: 'corona masker' (later I noticed that mondkapje would have been the better translation, but I kept it for the sake of consistency)), 'corona tracking app' (German: 'corona tracking app'; Dutch: 'corona tracking app'), 'corona measures' (German: 'corona maßnahmen'; Dutch: 'corona maatregelen'), 'who' (German: 'who'; Dutch: 'who'), 'donald trump' (German: 'donald trump'; Dutch: 'donald trump'), 'eu' (German: 'eu'; Dutch: 'eu'), 'italy' (German: 'italien'; Dutch: 'italie') and 'corona relaxation' (German: 'corona lockierung'; Dutch: 'corona versoepeling'). The script was run every two hours during each day of the data collection phase. The collected headlines can be found on Github.<sup>2</sup>

The search terms all revolve around the topic of the coronavirus. For 'corona mask requirement', 'corona tracking app', 'corona measures' and 'corona relaxation' the connection is obvious. The topic Donald Trump was chosen, as the United States were hit especially hard and as president of the United States, Donald Trump was one of the most influential figures of the time and also one of the most polarizing ones. He was constantly in the headlines due to his statements or his fight with China and the WHO. The WHO (World Health Organization) was also chosen, since they were also constantly in the headlines as they were coordinating the fight against the virus and their relationship with China and Donald Trump was pretty controversial. Italy was chosen, because it was one of the first countries to be hit very hard by the pandemic. Lastly, the EU was trying to achieve more cooperation between its member states in the fight against the virus and their actions are often controversial.

## 4.2 Data analysis methods

I used a self-written script for quantitative analysis of the data (see appendix B). This script checks, which sources, dates and words occur in a file and how often they occur in that file and what percentage of headlines have this source or date (in the case of words: percentage of total number of words in the whole file), orders them by descending number of occurrences and writes this data to a .txt file. It also counts the sources per language. A special case for both qualitative and quantitative analysis are the results for "who" in English where I filtered out all headlines that were in completely unrelated to the World Health Organization

Additionally I used Atlas.ti 8, a qualitative text analysis software, for qualitative analysis of the data. In Atlas.ti I went through each file and assigned so called 'codes' (which are basically just markers) to headlines.

There were two categories of codes I used: one for topics, which I only assigned once per file, if the topic has not occurred before in the specific file,

---

<sup>2</sup><https://github.com/emysliwietz/newsscrafer> in the folder "news\_old" and "news\_15-05-2020.19-50.zip" or at [matti.sermak.xyz](mailto:matti.sermak.xyz)

due to time constraints, and one to mark the tone of the headline, i.e. if this portrays a person, organization or political measure in a positive or negative light, which I assigned to every headline, that could be interpreted as such. Neutral articles were not marked, unless they mention a new topic, in which case they were marked with a topic code.

Since the tone of some headlines can be subtle, the way I coded articles with the second type of codes is subject to my own perception of the tone of a respective headline, which not every person, that has read the headline, might agree with. I tried to avoid marking headlines with those codes where the tone was too ambiguous, but some of the headlines I marked could possibly be interpreted differently by another person, who might perceive the tone as neutral.

After that I exported a table generated by Atlas.ti that displays how often each code has been used in each file and for the codes that marked criticism or support I counted the number of occurrences per language to get a general tendency of how a topic is portrayed in each language. For the topic codes I wrote another python script to list all the codes used in each individual file and grouped them by search term to make the comparison easier, e.g. I put the results for all three languages of the search term "donald trump" in one file, which contain three list of topic codes in total, one for each language (see appendix C). The generated files were then used to examine the difference in topics for a search term between the three languages. The Atlas.ti project and the exported table can also be found in the Github repository<sup>3</sup>

### 4.3 Observations

After examining the sources, dates and codes for each file, it is possible to make some observations and comparisons for each topic across the three languages. In total, I examined 4811 German headlines, 4298 English headlines and 2155 Dutch headlines, combining for a total of 11,264 headlines and used a total of 296 codes in Atlas.ti to mark those headlines<sup>4</sup>. To determine positive or negative tendencies on a topic, I counted together the number of headlines marked with the corresponding codes in all files for each language and compared the number of positive headlines with the number of negative headlines.

---

<sup>3</sup>[https://github.com/emysliwietz/newsscraeper/blob/master/Bachelor%20thesis%20\(Megaport%202020-05-17\).atlproj](https://github.com/emysliwietz/newsscraeper/blob/master/Bachelor%20thesis%20(Megaport%202020-05-17).atlproj) and [https://github.com/emysliwietz/newsscraeper/blob/master/Bachelor\\_thesis\\_CD.xlsx](https://github.com/emysliwietz/newsscraeper/blob/master/Bachelor_thesis_CD.xlsx)

<sup>4</sup>A full list of codes and in which files they were applied and how often they were applied for each file can be found at [https://github.com/emysliwietz/newsscraeper/blob/master/Bachelor\\_thesis\\_CD.xlsx](https://github.com/emysliwietz/newsscraeper/blob/master/Bachelor_thesis_CD.xlsx)

### 4.3.1 Relaxations of measures

	German			English			Dutch		
	Source	No.	%	Source	No.	%	Source	No.	%
3	<a href="#">merkur.de</a>	111	12.75	Business Today	46	12.71	NOS	40	15.5
4	<a href="#">NDR.de</a>	88	10.1	ABC News	38	10.5	RTL Nieuws	34	13.18
5	Süddeutsche Zeitung	42	4.82	Financial Times	33	9.12	<a href="#">AD.nl</a>	18	6.98
6	<a href="#">tagesschau.de</a>	35	4.018	BBC News	23	6.35	<a href="#">NU.nl</a>	16	6.2
7	DER SPIEGEL	27	3.1	The Guardian	17	4.7	<a href="#">Telegraaf.nl</a>	14	5.43
8	<a href="#">Nordbayern.de</a>	27	3.1	India Today	15	4.14	<a href="#">Hartvannederland.nl</a>	10	3.88
9	<a href="#">t-online.de</a>	26	2.99	Times of India	14	3.87	Volkskrant	8	3.1
10	WELT	24	2.76	The New York Times	13	3.59	NRC	7	2.71
11	MDR	24	2.53	DW (English)	11	3.038	<a href="#">Rijksoverheid.nl</a>	6	2.33
12	WDR Nachrichten	21	2.41	The Local Germany	10	2.76	<a href="#">Parool.nl</a>	6	2.33
13	RP ONLINE	21	2.41	CNN	9	2.48	Omroep Brabant	6	2.33
14	SWR	18	2.066	The Hindu	8	2.21	Trouw	5	1.94
15	ZEIT ONLINE	16	1.84	The Indian Express	8	2.21	Leeuwarder Courant	4	1.55
16	buten un binnen	16	1.84	CNBC	5	1.38	NH Nieuws	4	1.55
17	ka-news.de	16	1.84	CNN International	5	1.38	<a href="#">DeOndernemer.nl</a>	4	1.55
18	Berliner Morgenpost	15	1.72	Sky News	4	1.1	<a href="#">BD.nl</a>	3	1.16
19	rbb 24	14	1.61	dinешr	4	1.1	Accountancy Vanmorgen	3	1.16
20	RTL Online	14	1.61	Economic Times	4	1.1	RTV Utrecht	3	1.16
21	<a href="#">hessenschau.de</a>	14	1.61	ABP Live	4	1.1	Binnenlandsbestuur	3	1.16
22	Tagesspiegel	13	1.5	News18	4	1.1	Dagblad van het Noorden	3	1.16
23	FAZ - Frankfurter Allgemeine Zeitung	13	1.5	Al Jazeera English	4	1.1	De Gelderlander	2	0.78
24	Deutschlandfunk	12	1.38	The Statesman	4	1.1	Boerderij	2	0.78
25	<a href="#">fnp.de</a>	12	1.38	Daily Mail	4	1.1	CMweb - Controllers Magazine	2	0.78
26	<a href="#">hna.de</a>	10	1.15	<a href="#">india.com</a>	3	0.83	TaxLive	2	0.78
27	Passauer Neue Presse	10	1.15	Hindustan Times	3	0.83	<a href="#">Nieuws.nl</a>	2	0.78

Figure 4.1: The top 25 sources in each language with number of occurrences and percentage in its language

For the German headlines, there were 117 different sources in total, with most of them having less than five headlines. Figure 4.1 shows all sources with more than ten headlines, from which we can see, that most of those sources are either big mainstream papers like DER SPIEGEL and WELT, publicly funded broadcasters like the Tagesschau, NDR and WDR or regional papers like buten en binnen and hna. These are all among the most popular and most trusted outlets in Germany as seen in Chapter 2, showing a dominance of the mainstream with almost no outliers. The same phenomenon can be seen with the English sources (total: 77 sources) with mainstream news outlets from India (e.g. India Today, Times of India), the US (e.g. ABC, The New York Times) and the UK (e.g. BBC, The Guardian) dominating, as well as the Dutch sources (total: 68 sources) where popular mainstream sources like NOS, RTL Nieuws, AD.nl and NU.nl are also the most prevalent sources.

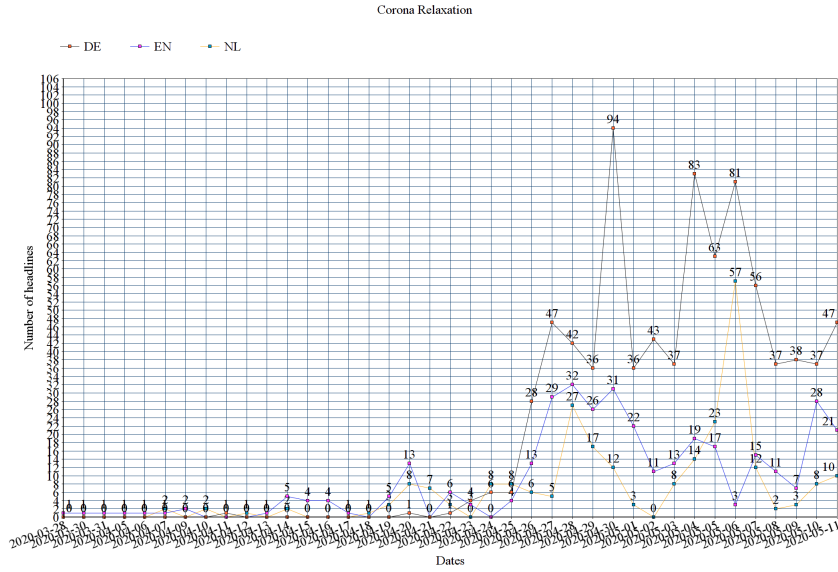


Figure 4.2: The number of headlines over time for each language

From figure 4.2 we can observe that the topic of relaxation of lockdown measures was most discussed in Germany and that the curves all go differently from each other. This has most likely to do with the fact that the coronavirus situation varies between different countries and thus, the governments are discussing the topic at different times. For example it can clearly be seen that new relaxation were announced or about to be announced in Germany on April 30 and May 4, while in the Netherlands the big peak is on May 5 when relaxations were announced by the Dutch government, thus the curves look very different for each language.

For German and English, there is a clear tendency against relaxations (German: 64 Anti vs. 33 Pro; English: 33 Anti vs. 7 Pro), while the Dutch headlines are almost evenly split (15 Anti vs. 16 Pro). For English this seems to coincide with public opinion in the respective countries, for German however this goes somewhat against the opinion of the majority of the German people, who generally support the relaxations of the German government (see Chapter 2), though there are around 40% who felt these relaxations came to early [13], so this is still a popular opinion. Similarly, the public opinion seems to be more favorable towards relaxation than those results would suggest [38]. The numbers also show that all three languages generally show both sides of the debate, but in the case of German and English one side is showcased significantly more.

1	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL
2	Anti border closures	X			Mask requirement in Spain		X	
3	Anti Lockdown	X			Relaxation measures in Australia		X	
4	Anti-relaxation	X	X	X	Relaxation measures in the US		X	
5	Border closures	X			Relaxation measures UK		X	
6	Corona Tracking App in Germany	X			Relaxation of UK abortion rules		X	
7	Coronavirus in Africa	X	X		Reports of China pressuring WHO to suppress information about virus		X	
8	Coronavirus in Europe	X	X	X	Rising number of deaths after relaxation in Italy		X	
9	Coronavirus in Portugal	X	X		Coronavirus in Japan			X
10	Coronavirus number DE	X	X	X	Coronavirus numbers NL			X
11	Crisissituation in Italy	X			Information on Dutch relaxation measures			X
12	Criticism of government lockdown measures	X			Measures taken NL			X
13	Criticism of Italian government	X			More domestic violence during lockdown			X
14	Criticism of relaxation measures	X	X	X	Privacy concerns over digital exams			X
15	Criticism of RKI	X			Relaxation of measures Belgium			X
16	Economical impact of coronavirus	X	X	X	Support for (DIY) masks			X
17	Favorable towards corona app	X	X		Support for using masks			X
18	Football/soccer in times of the coronavirus	X		X	Trump replacing corona taskforce			X
19	German regions reentering lockdown due to rising numbers	X			Virus originated from Wuhan lab' theory			X
20	Mask requirements in Austria	X						
21	People's behaviour in public	X	X	X				
22	Pro-Relaxation	X	X	X				
23	Protests against lockdown	X						
24	Relaxation measures in Austria	X		X				
25	Relaxation measures in Germany	X	X	X				
26	Relaxation measures in Italy	X	X	X				
27	Relaxation measures in Spain	X		X				
28	Tourism and Travel	X		X				
29	Vaccine against the coronavirus	X						
30	Corona app in the Netherlands		X					
31	Coronavirus deaths in Italy		X					
32	Coronavirus in India		X					
33	Coronavirus in Mexico		X					
34	Coronavirus numbers US		X					
35	Coronavirus numbers India		X					
36	Lockdown measures in France		X	X				

Figure 4.3: Topics marked by me in Atlas.ti in the files for relaxation

There is a visible difference in the news stories reported in each language (see figure 4.3). This mostly comes down to regional relevance, since a story like the relaxations in Belgium is not really interesting for people outside of Belgium and the Netherlands, whereas the relaxations in Australia are not that interesting to Dutch or German people. Relaxations in Germany and Italy on the other hand are reported in all languages, in the case of Italy, because it has been one of the first and hardest hit countries in Europe, and Germany, because it is one of the first big countries to relax its measures <sup>5</sup>.

<sup>5</sup>Google also knows I am German, so that might have an influence on that, but that is just speculation

### 4.3.2 Measures

	German			English			Dutch		
	Source	No.	%	Source	No.	%	Source	No.	%
3	<a href="#">merkur.de</a>	192	20.45	Financial Times	172	27.52	NOS	65	18.21
4	<a href="#">tagesschau.de</a>	47	5	BBC News	78	12.48	RTL Nieuws	49	13.73
5	Berliner Morgenpost	36	3.83	The Guardian	30	4.8	<a href="#">Hartvannederland.nl</a>	20	5.6
6	Tagesspiegel	32	3.41	The New York Times	28	4.48	<a href="#">Telegraaf.nl</a>	20	5.6
7	<a href="#">NDR.de</a>	30	3.19	Al Jazeera English	28	4.48	<a href="#">NU.nl</a>	18	5.042
8	RP ONLINE	28	2.98	<a href="#">Metro.co.uk</a>	26	4.16	VRT NWS	16	4.48
9	Deutschlandfunk	26	2.77	CNN	25	4	<a href="#">AD.nl</a>	14	3.92
10	WELT	24	2.56	<a href="#">Telegraph.co.uk</a>	22	3.52	Dagblad van het Noorden	9	2.52
11	rbbj24	23	2.45	BBC South East Wales	21	3.36	Omroep Brabant	9	2.52
12	<a href="#">t-online.de</a>	22	2.34	ABC News	16	2.56	ANWB	9	2.52
13	BR24	18	1.92	The Sun	14	2.24	<a href="#">Rijksoverheid.nl</a>	8	2.24
14	DER SPIEGEL	18	1.92	The Washington Post	14	2.24	Volkskrant	6	1.68
15	RTL Online	15	1.6	CNN International	13	02.08	Omroep West	6	1.68
16	ZDFheute	14	1.49	Sky News	11	1.76	NRC	5	1.4
17	Nordkurier	14	1.49	Daily Mail	9	1.44	RTV Utrecht	5	1.4
18	<a href="#">fr.de</a>	14	1.49	The Telegraph	9	1.44	<a href="#">Amstelveenz.nl</a>	4	1.12
19	BNN - Badische Neueste Nachrichten	13	1.38	The Local Germany	9	1.44	Plusonline	4	1.12
20	Süddeutsche Zeitung	13	1.38	The Berlin Spectator	7	1.12	NH Nieuws	4	1.12
21	RND	13	1.38	CNBC	7	1.12	Blik op nieuws	4	1.12
22	buten un binnen	13	1.38	NPR	6	0.96	RTV Oost	3	0.84
23	SWR	12	1.28	Times of India	4	0.64	<a href="#">Parool.nl</a>	3	0.84
24	ZEIT ONLINE	12	1.28	The Local Sweden	3	0.48	Binnenlandsbestuur	3	0.84
25	<a href="#">Nordbayern.de</a>	12	1.28	ITV News	3	0.48	RTV Noord	3	0.84
26	idowa	11	1.17	AS.com Móvil	3	0.48	RTV Rijnmond	3	0.84
27	FAZ - Frankfurter Allgemeine Zeitung	9	0.96	The Lippy	3	0.48	De Utrechtse Internet Courant	2	0.56

Figure 4.4: The top 25 sources in each language with number of occurrences and percentage in its language

Similarly to before, in all three languages mainstream news outlets and regional papers are dominating the list of sources (see also figure 4.4). This time however, there are a few more less common news sources among the German sources with one or two headlines each, e.g. skispringen.com or FIT FOR FUN. In total there are 133 German sources, with 70 of them having just one or two headlines each, 71 English sources, with 43 of them having less than 3 headlines each, and 78 Dutch sources, with 64 of them having less than 3 headlines each.

Figure 4.5 shows that again, the curves vary a lot between languages, that often have different peaks and lowpoints. Just like in the previous section, this can probably be attributed to regional differences in measures and different stories being reported.

When looking at the different topics and news stories covered in the headlines (see figure 4.6), it becomes apparent, that there are a lot of topics/stories that are only covered in English and most topics that are covered in German or Dutch are also covered in at least one other language as well. Those stories that exclusively appear in German or Dutch are mostly regional topics in the respective countries, whereas stories that are exclusively in English contains a lot of international news on countries all over the world.



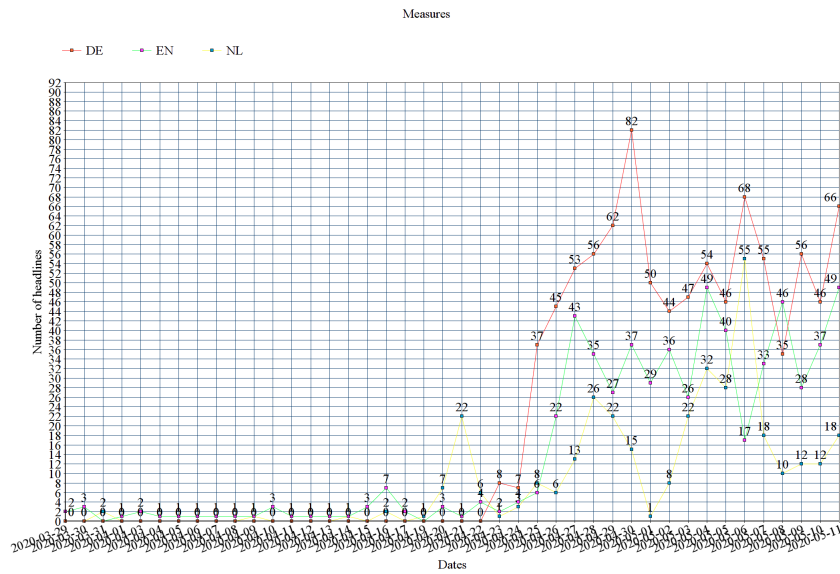


Figure 4.5: The number of headlines over time for each language

1	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL
2	Anti-Lockdown	X	X	X	Coronavirus numbers US	X	X	X	Praise for Sweden's measures		X	
3	Anti-relaxation	X	X	X	Coronavirus testing	X	X	X	Relaxation measures in Australia		X	
4	Articles about Fake news about the coronavirus			X	Crisissituation in Italy	X	X	X	Relaxation measures in Italy		X	
5	Corona measures Belgium			X	Critical of lockdown protests		X	X	Relaxation measures in the US		X	
6	Coronavirus in China		X	X	Criticism of relaxation measures	X	X	X	Remdesivir medicine		X	
7	Coronavirus in Europe	X	X	X	Criticism of RKI	X	X	X	Social Distancing		X	
8	Coronavirus in Japan		X	X	Elon Musk	X	X	X	Support/Praise for German government		X	
9	Coronavirus NL			X	German regions reentering lockdown due to rising numbers	X	X	X	Trump criticizing China		X	
10	Coronavirus numbers Belgium			X	Mask requirement in Germany	X	X	X	US election		X	
11	Coronavirus numbers NL			X	Measures China	X	X	X	Vaccines		X	
12	Coronavirus worldwide	X	X	X	Measures taken Germany	X	X	X	Virus originated from Wuhan lab' theory		X	
13	Critical of Trump	X	X	X	Measures Turkey	X	X	X				
14	Criticism of (DIY) masks	X	X	X	No proof of immunity after recovering from infection	X	X	X				
15	Criticism of government lockdown measures	X	X	X	Relaxation measures in Austria	X	X	X				
16	Economical impact of coronavirus	X	X	X	Relaxation measures in Germany	X	X	X				
17	Football/soccer in times of the coronavirus	X	X	X	Relaxation measures in Spain	X	X	X				
18	Information on Dutch relaxation measures	X	X	X	Relaxation Switzerland	X	X	X				
19	Lockdown measures in Spain	X	X	X	Vaccine against the coronavirus	X	X	X				
20	Measures China	X	X	X	WHO praising Sweden	X	X	X				
21	Measures taken NL			X	WHO warning pandemic nowhere near over	X	X	X				
22	People's behaviour in public	X	X	X	Corona app international		X	X				
23	Personal stories		X	X	Corona tracking app UK		X	X				
24	Pro-Relaxation	X	X	X	Coronavirus in Brazil		X	X				
25	Protests against lockdown	X	X	X	Coronavirus in Hong Kong		X	X				
26	Relaxation measures UK			X	Coronavirus in India		X	X				
27	Relaxation of measures Belgium	X	X	X	Coronavirus in Ireland		X	X				
28	Support of government lockdown measures	X	X	X	Coronavirus in Mexico		X	X				
29	Tourism and Travel	X	X	X	Coronavirus numbers UK		X	X				
30	Trump replacing corona taskforce			X	Coronavirusnumbers Russia		X	X				
31	Call for excluding Boris Palmer from Green Party	X			Criticism of Corona App		X	X				
32	China spreading fake news			X	Criticism of EU measures/actions		X	X				
33	Corona measures influencing influenza	X			Hong Kong Protests		X	X				
34	Coronavirus in Africa	X	X		Mask requirement in Spain		X	X				
35	Coronavirus in Portugal	X			Measures Singapore		X	X				
36	Coronavirus number DE	X			Measures taken UK		X	X				

Figure 4.6: Topics marked by me in Atlas.ti in the files for measures

By examining the opinions about lockdown measures, we can observe that in German there is a strong tendency towards criticism of those measures (35 Criticism vs. 9 Supportive). The same holds for Dutch to a lesser

extend (5 Criticism vs. 2 Supportive), whereas in English the headlines seem to be more supportive of lockdown measures (4 Criticism vs. 11 Supportive). While English and Dutch mostly reflect the majority opinion in the Netherlands/the UK and the US, for German this goes very much against the opinion of a vast majority of the population [14] and seems very contradictory to what we saw in the last section. It should be noted however, that most of the articles criticising the lockdown measures of the German government were written in April and thus before some measures were relaxed.

### 4.3.3 Tracking App

1	German			English			Dutch		
2	Source	No.	%	Source	No.	%	Source	No.	%
3	Deutschlandfunk	14	7.53	BBC News	71	20.82	ICT&health.nl	19	12.41
4	<a href="#">connect.de</a>	9	4.84	The Guardian	49	14.37	Telecompaper	17	11.11
5	Berliner Morgenpost	8	4.3	Daily Mail	11	3.23	Datanews	16	10.46
6	WELT	7	3.76	The New York Times	10	2.93	<a href="#">Security.nl</a>	11	7.19
7	Tagesspiegel	7	3.76	9News	9	2.64	<a href="#">Bright.nl</a>	9	5.88
8	<a href="#">laz.de</a>	6	3.23	The Sun	9	2.64	RTL Nieuws	8	5.23
9	<a href="#">tagesschau.de</a>	5	2.69	The Washington Post	9	2.64	AG Connect	6	3.92
10	Süddeutsche Zeitung	5	2.69	ABC News	9	2.64	Het Financieele Dagblad	6	3.92
11	Die Presse	5	2.69	CNN	7	2.052	Tweakers	5	3.27
12	Handelsblatt	5	2.69	BBC South East Wales	7	2.052	Emerce	4	2.61
13	heise online	4	2.15	The Times	5	1.47	Androidworld	4	2.61
14	BR24	4	2.15	Mirror Online	5	1.47	iCulture	4	2.61
15	rbb24	4	2.15	Global News	5	1.47	Volkskrant	3	1.96
16	CHIP Online	4	2.15	Sky News	4	1.17	<a href="#">AD.nl</a>	3	1.96
17	COMPUTER BILD	4	2.15	CNBC	4	1.17	Dutch IT-channel	2	1.31
18	InfoDigital / INFOSAT	4	2.15	The Verge	4	1.17	One More Thing	2	1.31
19	<a href="#">inFranken.de</a>	3	1.61	The Next Web	4	1.17	TechPulse	2	1.31
20	<a href="#">mannheim24.de</a>	3	1.61	CNET	4	1.17	Marketingfacts	2	1.31
21	ZEIT ONLINE	3	1.61	<a href="#">Metro.co.uk</a>	4	1.17	<a href="#">Parool.nl</a>	2	1.31
22	n-tv NACHRICHTEN	3	1.61	The Wall Street Journal	4	1.17	Binnenlandsbestuur	2	1.31
23	<a href="#">Netzpolitik.org</a>	3	1.61	Inkstone	4	1.17	Medisch Contact	2	1.31
24	Mac & i	3	1.61	<a href="#">7NEWS.com.au</a>	3	0.88	Plusonline	1	0.65
25	SWR	3	1.61	<a href="#">Wired.co.uk</a>	3	0.88	<a href="#">XGN.nl</a>	1	0.65
26	DER SPIEGEL	3	1.61	Reuters	3	0.88	Want	1	0.65
27	<a href="#">oe24.at</a>	3	1.61	The Indian Express	3	0.88	<a href="#">Techzine.nl</a>	1	0.65

Figure 4.7: The top 25 sources in each language with number of occurrences and percentage in its language

As opposed to the previous two topics, this time there are less mainstream news outlets and more "obscure" sources (though English still seems to be dominated by mainstream news outlets). Most of these "obscure" sources are websites and magazines focusing on electronics and technology, like e.g. CHIP Online, COMPUTER BILD and InfoDigital in German and ICT&health.nl, Datanews and Security.nl for Dutch (see also figure 4.7). This makes sense in the context of the topic of a Contact Tracking/Tracing App, since this topic does revolve around technology. In total, there are 72 German sources (45 with less than 3 headlines), 97 English sources (75 with less than 3 headlines) and 43 Dutch sources (29 with less than 3 headlines).

By looking at the graph in figure 4.8, we make two observations. The first

observation is that there are a lot more articles in English on this topic than in German or Dutch. The second possible observation is that this time the progressions of the curves appear to look more similar than for the two previous search terms, though there are still visible differences. A possible explanation for this could be that the topic of tracking apps allows for more international stories that are of interest for all of the relevant countries and that the differences come due to the debate about apps evolving differently among the relevant countries.



Figure 4.8: The number of headlines over time for each language

1	Code/Language	DE	EN	NL
2	Centralized/de-centralized data storage for corona app	X	X	
3	Corona Tracking App in Germany	X		X
4	Corona tracking app UK	X	X	
5	Criticism of Corona App	X	X	X
6	Data protection/Privacy	X	X	X
7	Explanation of how corona app works	X	X	X
8	Favorable towards corona app	X	X	X
9	Google's and Apple's progress on corona app	X	X	X
10	SAP and Telekom developing corona tracking app	X		
11	Tracking app in Asia	X		X
12	App shows actual infection numbers are higher than official numbers		X	
13	Corona app in the Netherlands		X	X
14	Corona app international		X	
15	Corona tracking app in Australia		X	
16	Coronavirus deaths in Italy		X	
17	Coronavirus worldwide		X	
18	European Initiative for corona app			X
19	How the Dutch follow coronavirus news			X

Figure 4.9: Topics marked by me in Atlas.ti in the files for tracking app

This hypothesis, that the topic of tracking app allows for more internationally relevant news stories, can be supported when looking at the covered topics and news stories in figure 4.9, since this time most topics are featured in at least two languages with less topics that are exclusive to one language previously. Again, stories that are exclusive in a language mostly revolve around a country where the respective language is spoken (in this case the Netherlands, the UK and Australia).

In English (22 Contra vs. 12 Pro) and Dutch (28 Contra vs. 6 Pro) headlines are more critical of a tracking app, while in German the headlines seem more split with a light advantage for the Pro side (10 Contra vs. 12 Pro), which seems to be in line with the respective opinions in the respective countries as mentioned in chapter 2.

#### 4.3.4 Face masks

On the topic of face masks and face masks requirements, we can again observe that most of the sources are mainstream news outlets and regional papers for all three languages. There are however more regional papers among the English sources than for the previous search terms as can be seen in figure 4.10. In total there are 89 German sources (50 with less than 3 headlines), 84 English sources (59 with less than 3 headlines) and 56 Dutch sources (40 with less than 3 headlines).

	German			English			Dutch		
	Source	No.	%	Source	No.	%	Source	No.	%
1	<a href="#">merkur.de</a>	92	19.45	The New York Times	29	13.68	VRT NWS	22	11.06
2	<a href="#">hna.de</a>	34	9.090	Financial Times	17	8.018	<a href="#">AD.nl</a>	21	10.55
3	<a href="#">NDR.de</a>	33	6.98	The Guardian	9	4.25	RTL Nieuws	15	7.54
4	Berliner Morgenpost	20	4.22	USA TODAY	7	3.3	NOS	12	6.030
5	SWP	19	04.01	The Washington Post	6	2.83	De Gelderlander	10	5.025
6	WELT	16	3.38	CNBC	5	2.35	Het Nieuwsblad	10	5.025
7	DER SPIEGEL	15	3.17	KGO-TV	5	2.35	Het Laatste Nieuws	9	4.52
8	<a href="#">tz.de</a>	14	2.96	MarketWatch	4	1.89	NRC	9	4.52
9	RTL Online	12	2.54	Los Angeles Times	4	1.89	Volkskrant	7	3.52
10	RP ONLINE	10	2.11	The Local Germany	4	1.89	<a href="#">Parool.nl</a>	5	2.51
11	<a href="#">tagesschau.de</a>	8	1.69	NBC10 Boston	4	1.89	Datanews	5	2.51
12	<a href="#">t-online.de</a>	8	1.69	Al Jazeera English	4	1.89	Trouw	4	2.010
13	rbb24	8	1.69	The Philadelphia Inquirer	3	1.42	<a href="#">Bright.nl</a>	4	2.010
14	<a href="#">ruhr24.de</a>	8	1.69	CNN International	3	1.42	<a href="#">NU.nl</a>	4	2.010
15	BILD	7	1.48	The Sun	3	1.42	Dagblad van het Noorden	3	1.51
16	<a href="#">kreiszeitung.de</a>	7	1.48	WLS-TV	3	1.42	Omroep Brabant	3	1.51
17	buten un binnen	7	1.48	Austin American-Statesman	3	1.42	Eos Wetenschap	2	1.005
18	RND	6	1.27	<a href="#">nbc.com</a>	3	1.42	BN DeStem	2	1.005
19	watson	6	1.27	NBC Chicago	3	1.42	Nieuws en sport uit Limburg	2	1.005
20	<a href="#">inFranken.de</a>	6	1.27	Press-Enterprise	3	1.42	RTV Rijnmond	2	1.005
21	Tagesspiegel	6	1.27	WBUR	3	1.42	<a href="#">Telegraaf.nl</a>	2	1.005
22	<a href="#">heidelberg24.de</a>	5	1.057	KOMO News	3	1.42	De Tijd	2	1.005
23	<a href="#">fnp.de</a>	5	1.057	Las Cruces Sun-News	3	1.42	Het Financieele Dagblad	2	1.005
24	Westfälischer Anzeiger	5	1.057	Chicago Sun-Times	3	1.42	Tubantia	2	1.005
25	<a href="#">STERN.de</a>	5	1.057	<a href="#">Telegraph.co.uk</a>	3	1.42	Ditjes en Datjes	2	1.005

Figure 4.10: The top 25 sources in each language with number of occurrences and percentage in its language

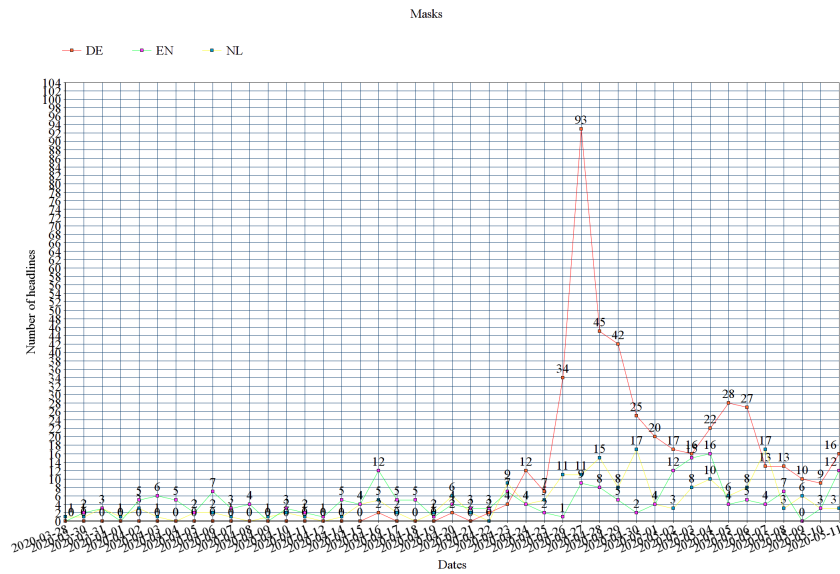


Figure 4.11: The number of headlines over time for each language

From figure 4.11 we can see that the topic of face masks was much more relevant in Germany than in other countries and that most headlines are from the last days of April when face masks requirements were introduced all over Germany. In contrast, in the Netherlands, where masks are not mandatory in supermarkets and many other places, there are far fewer articles about face masks due to less relevance, and thus there are less Dutch headlines.

1	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL
2	Coronavirus deaths in Italy	X			Coronavirus number DE	X		
3	Coronavirus numbers US	X			Information on German measures	X		
4	Critical of lockdown protests	X			Relaxation measures in Austria	X		
5	Criticism of (DIY) masks	X	X	X	Relaxation measures in Germany	X		
6	Criticism of relaxation measures	X	X	X	Trump and Pence not wearing masks	X	X	
7	DIY masks	X	X	X	Articles about Fake news about the coronavirus			X
8	Economical impact of coronavirus	X			Corona outbreak on US navy ship			X
9	Face mask requirement policies worldwide	X			Coronavirus numbers NL			X
10	Information about masks	X	X	X	Critical of Trump			X
11	Lockdown measures in Spain	X			Dogs detecting corona			X
12	Lockdown measures in France	X			Dutch companies producing masks			X
13	Lockdown measures in the US	X			Hospital entrance checks			X
14	Mask requirement in Germany	X	X	X	Information on Dutch relaxation measures			X
15	Mask requirement in Singapore	X			Less deaths in Italy			X
16	Mask requirement in Spain	X			Mask shortage in hospitals			X
17	Mask requirements in Austria	X	X	X	Masks as fashion items			X
18	Mask requirements in the US	X	X	X	Masks in Belgium			X
19	Mask rules in the UK	X			More domestic violence during lockdown			X
20	Oil prices	X			New lockdown in South Korea			X
21	Parallels to 1918 Spanish flu pandemic	X	X	X	Personal stories			X
22	People's behaviour in public	X	X	X	Praise for Taiwan			X
23	Protests against lockdown	X			Reduced numbers of corona patients			X
24	Relaxation measures in the US	X			Reinforcement of distance rules			X
25	Support for (DIY) masks	X	X	X	Relaxation measures in Italy			X
26	Support for using masks	X	X	X	Remdesivir medicine			X
27	Anti Lockdown	X			Virus originated from Wuhan lab' theory			X

Figure 4.12: Topics marked by me in Atlas.ti in the files for masks



For this search term, there is a comparatively big differences in stories and topics covered by the headlines in each language, with most topics being exclusive to one language (see figure 4.12). Especially in English and Dutch, there is a wide range of topics not covered in another language, and this is mostly but not only due to regional news stories. Especially in Dutch there are a lot of topics covered that are more indirectly (if at all) related to face masks such as a rise in domestic violence during lockdown or the Remdesivir medicine.

While the English (1 Contra vs. 10 Pro) and Dutch headlines (1 Contra vs. 5 Pro) are mostly in line with the popular opinion in the Netherlands, the UK and the US of supporting face mask requirements, the German headlines go completely against the opinion of the majority of the German public (see also chapter 2) and feature a lot of criticism of face mask requirements (31 Contra vs. 5 Pro).

### 4.3.5 Donald Trump

Figure 4.13 shows that - again - most sources appear to be mainstream news outlets and magazines. It should be noted however, that, in English, news outlets that are often perceived to be biased towards the Democratic party, like CNN, are far more prevalent than those that are perceived as conservative, like FOX News. In total there are 65 German sources (37 with less than 3 headlines), 155 English sources (100 with less than 3 headlines) and 35 Dutch sources (16 with less than 3 headlines).

	German		English		Dutch	
	Source	No. %	Source	No. %	Source	No. %
1	<a href="#">fr.de</a>	101 21.49	CNN	207 21.68	<a href="#">AD.nl</a>	33 16.5
2	DER SPIEGEL	48 10.21	The Guardian	144 15.08	VRT NWS	21 10.5
3	<a href="#">t-online.de</a>	48 10.21	Daily Mail	56 5.86	De Standaard	12 6
4	<a href="#">merkur.de</a>	30 6.38	POLITICO	50 5.24	Volkskrant	11 5.5
5	<a href="#">Nau.ch</a>	23 4.89	CNBC	29 3.036	Het Nieuwsblad	11 5.5
6	<a href="#">STERN.de</a>	22 4.68	Deadline	26 2.72	Trouw	11 5.5
7	RND	17 3.62	The Atlantic	26 2.72	De Tijd	10 5
8	BILD	14 2.98	<a href="#">Metro.co.uk</a>	22 2.3	<a href="#">Telegraaf.nl</a>	9 4.5
9	ZEIT ONLINE	13 2.77	USA TODAY	19 1.99	RTL Nieuws	8 4
10	WELT	12 2.55	ABC News	16 1.67	Het Laatste Nieuws	8 4
11	FAZ - Frankfurter Allgemeine Zeitung	11 2.34	Bloomberg	15 1.57	NRC	7 3.5
12	Tagesspiegel	9 1.91	BBC News	13 1.36	Business Insider	6 3
13	Berliner Morgenpost	8 1.7	<a href="#">Express.co.uk</a>	12 1.26	<a href="#">NU.nl</a>	6 3
14	Tages-Anzeiger	8 1.7	<a href="#">Telegraph.co.uk</a>	12 1.26	Nederlands Dagblad	5 2.5
15	RP ONLINE	6 1.28	The New York Times	11 1.15	<a href="#">Gids.tv</a>	4 2
16	<a href="#">tz.de</a>	6 1.28	Vanity Fair	11 1.15	Joop	3 1.5
17	WEB.DE News	6 1.28	Business Insider	11 1.15	Wellington Kringen	3 1.5
18	Frankfurter Rundschau	6 1.28	Financial Times	8 0.84	Gazet van Antwerpen	3 1.5
19	FOCUS Online	5 1.063	The Washington Post	8 0.84	BNR Nieuwsradio	3 1.5
20	watson	4 0.85	Mirror Online	7 0.73	<a href="#">Bladna.nl</a>	2 1
21	<a href="#">Merkur.de</a>	4 0.85	<a href="#">RTE.ie</a>	6 0.63	De Dagelijkse Standaard	2 1
22	<a href="#">tagesschau.de</a>	4 0.85	<a href="#">Vox.com</a>	6 0.63	Want	2 1
23	Süddeutsche Zeitung	4 0.85	<a href="#">NBCNews.com</a>	6 0.63	<a href="#">Autoblog.nl</a>	2 1
24	COMPUTER BILD	4 0.85	<a href="#">Whitehouse.gov</a>	6 0.63	FilmTotaal	2 1
25	Cash	3 0.63	Hindustan Times	6 0.63	RTL Boulevard	2 1

Figure 4.13: The top 25 sources in each language with number of occurrences and percentage in its language



Figure 4.14: The number of headlines over time for each language

This time the curves of headlines per date are very similar for all languages, as can be seen in figure 4.14, with similar peaks and trajectories. The reason behind this is probably that all three languages report very similar topic, as can also be seen when examining the covered topics and news stories (see figure 4.15), since apart from English, there is not much room for regional topics, which is also the reason there are far more exclusive stories/topics in English than in the other languages. These exclusive topics mostly revolve around very US-specific topics like white nationalism, Trump having a fight with FOX News or the Republican party trying to repeal Obamacare.

Unsurprisingly, the headlines in all three languages paint a very negative picture of Trump. In German there are 72 negative and 2 positive articles about Trump, in English 123 negative and 3 positive articles and in Dutch 29 negative and 3 positive articles about Trump. This coincides with the majority opinions on Donald Trump's presidency in all examined countries, though I expected there to be more positive articles in English since he still has a job approval rating of over 40 percent in the US [35].

	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL
1	"Obamagate"	X	X		Trump vs his critics	X			Oil prices			X
2	Anti-relaxation	X	X		Trump vs the media	X	X	X	Pro Trump in election			X
3	Celebrities criticizing Trump	X	X	X	Trump's bleach/disinfectant "advice"	X	X	X	Resource shortage in hospitals			X
4	Continuation American sports competitions	X	X	X	Trump's new spokesperson	X	X	X				
5	Coronavirus numbers US	X	X	X	Trump-WHO conflict	X	X	X				
6	Coronavirus testing	X	X	X	US election	X	X	X				
7	Crisis checks for US citizens	X	X	X	US immigration ban	X	X	X				
8	Critical of Trump	X	X	X	US-Iran conflict	X		X				
9	Defense of/Positive towards Trump	X	X		Vaccine against the coronavirus	X	X	X				
10	Economical impact of coronavirus	X	X		Virus originated from Wuhan lab' theory	X	X	X				
11	Former Trump advisor found not guilty	X	X	X	Conservatism in America	X						
12	Gretchen Whitmer	X			Coronavirus in Mexico	X						
13	Jared Kushner's controversial statements	X	X	X	Debate about state bailouts USA	X						
14	Joe Exotic pardoning	X	X	X	Donald Trump's family	X						
15	Kushner providing Trump with inaccurate numbers	X			Donald Trump's personal life	X						
16	Mask requirements in the US	X			GOP fight against Obamacare	X						
17	Parody Song of Trump	X			Melania Trump	X						
18	Protests against lockdown	X			Rumors about food shortage US	X						
19	Relaxation measures in the US	X	X		Trump and Fox News fighting	X						
20	Remdesivir medicine	X	X		Trump demanding relaxation	X						
21	Restrictions to freedom of the press	X			Trump giving Saudi-Arabia ultimatum	X						
22	Rumors about Trump firing Secretary of Health	X	X		Trump giving \$69 million to ventilator entrepreneur	X						
23	Speculation about Kim Jong-Un's health	X	X	X	Trump influencing courts	X						
24	Support/Praise for German government	X			Trump not taking responsibility for bleach incidents	X	X					
25	Travel expenditures of the Trump family	X			Trump valet testing positive	X	X					
26	Trump advisor on tape: "Number of deaths will rise"	X			Trump's business	X						
27	Trump and Pence not wearing masks	X	X	X	US Postal Service	X						
28	Trump bodyguards arrested in Venezuela	X			US stopping WHO financing	X						
29	Trump criticizing China	X	X	X	White nationalism in the US	X						
30	Trump praising Helene Fischer	X			Anthony Fauci			X				
31	Trump praising himself	X	X	X	Coronavirus numbers Belgium			X				
32	Trump praising Ivanka	X			Elon Musk			X				
33	Trump press briefings	X	X	X	Governor Cuomo			X				
34	Trump replacing corona taskforce	X	X	X	Impeachment			X				
35	Trump tweets	X	X	X	Lockdown measures in the US			X				

Figure 4.15: Topics marked by me in Atlas.ti in the files for Donald Trump

#### 4.3.6 EU

	German	No.	%	English	No.	%	Dutch	No.	%
1	Source			Source			Source		
2	Süddeutsche Zeitung	47	6.33	<a href="https://www.express.co.uk">Express.co.uk</a>	74	9.71	Trouw	23	6.99
3	Deutschlandfunk	38	5.12	EU News	60	7.87	<a href="https://www.telegraaf.nl">Telegraaf.nl</a>	21	9.38
4	Handelsblatt	30	4.042	Financial Times	55	7.35	Leeuwarder Courant	18	5.47
5	DER SPIEGEL	25	3.37	The Guardian	37	4.86	Volkskrant	17	5.16
6	FAZ - Frankfurter Allgemeine Zeitung	23	3.1	The New York Times	35	4.59	Boerderij	15	4.56
7	<a href="https://www.tagesschau.de">tagesschau.de</a>	23	3.1	BBC News	33	4.33	Het Financieele Dagblad	14	4.26
8	WELT	22	2.96	EURACTIV	31	4.068	NRC	13	3.95
9	Tagesspiegel	18	2.43	EUobserver	22	2.89	Nederlands Dagblad	12	3.65
10	KURIER	18	2.43	Bloomberg	21	2.76	NOS	10	3.039
11	<a href="https://www.orf.at">ORF.at</a>	17	2.29	Reuters UK	20	2.62	Elsevier Weekblad	10	3.039
12	DW (Deutsch)	15	2.021	Reuters	19	2.49	<a href="https://www.nu.nl">NU.nl</a>	9	2.74
13	Die Presse	15	2.021	The Irish Times	18	2.36	RTL Nieuws	9	2.74
14	RND	14	1.89	DW (English)	15	1.97	Nieuwe Oogst	8	2.43
15	EURACTIV Germany	13	1.75	<a href="https://www.telegraph.co.uk">Telegraph.co.uk</a>	13	1.7	<a href="https://www.knack.be">Knack.be</a>	8	2.43
16	ZEIT ONLINE	13	1.75	Euronews	13	1.7	<a href="https://www.parool.nl">Parool.nl</a>	7	2.13
17	Euronews	11	1.48	Anadolu Agency	13	1.7	RTL Z	7	2.13
18	<a href="https://www.taz.de">taz.de</a>	10	1.35	Daily Mail	12	1.57	Reformatorisch Dagblad	6	1.82
19	Oldenburger Onlinezeitung	10	1.35	<a href="https://www.schengenvisainfo.com">SchengenVisaInfo.com</a>	9	1.18	VRT NWS	5	1.52
20	onvista	10	1.35	The Sun	9	1.18	BNR Nieuwsradio	5	1.52
21	Cash	10	1.35	CNBC	8	1.049	De Tijd	5	1.52
22	<a href="https://www.derstandard.de">derStandard.de</a>	10	1.35	The New European	7	0.92	De Standaard	5	1.52
23	<a href="https://www.airliners.de">airliners.de</a>	9	1.21	EU Reporter	7	0.92	<a href="https://www.ad.nl">AD.nl</a>	4	1.22
24	Reuters Deutschland	9	1.21	The Brussels Times	6	0.79	<a href="https://www.wnl.tv">wnl.tv</a>	4	1.22
25	top agrar online	9	1.21	The Independent	6	0.79	Dagblad van het Noorden	4	1.22
26	<a href="https://www.finanzen.net">finanzen.net</a>	9	1.21	European Western Balkans	6	0.79	<a href="https://www.boerenbusiness.nl">Boerenbusiness.nl</a>	4	1.22
27									

Figure 4.16: The top 25 sources in each language with number of occurrences and percentage in its language

Apart from the repeated dominance of mainstream news outlets, figure 4.16 shows that this time there are also a lot of economic and financial magazines,



newspapers and websites such as Handelsblatt, finanzen.net, Het Financieele Dagblad and Financial Times among the list of sources, most likely due to the importance of the EU and its actions for the economy for most countries in Europe.

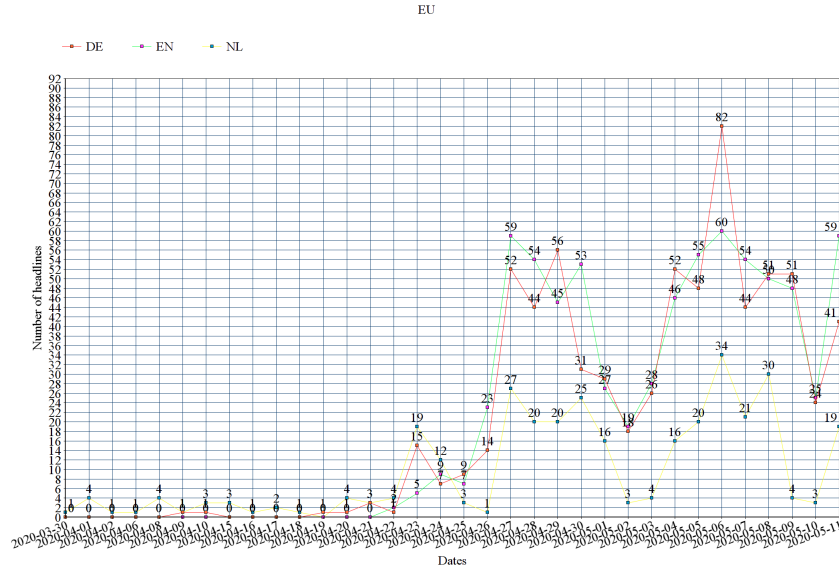


Figure 4.17: The number of headlines over time for each language

Figure 4.17 documents relatively similar progression of the daily number of headlines in all languages, with different intensities in each language (for example there are less articles in Dutch than in German or English). The most likely reason for this that topics concerning the European Union are usually of similar importance in Germany, the Netherlands and the UK, especially when it concerns the handling of a pandemic like the coronavirus, and thus many news get reported at the same time in all three countries. This assumption is somewhat supported when looking at figure 4.18, since it can be seen from the topics I marked, most topics are reported in multiple languages and a lot of those even in all three languages. However there are still a lot of topics that are exclusive to one language. These are, again, mostly centered around the countries where the respective language is spoken. For example, exclusive topics in English often relate to Brexit, e.g. 'Northern Ireland issue in Brexit talks' or 'UK-EU relations post-Brexit', or the United States such as 'US medical academy supporting EU'. However there are also topics that are only covered in one language that do not directly relate to the corresponding countries. Some examples for this are 'EU-China relations' being exclusive to English, 'Search for missing peo-

ple from Kosovo war' being exclusive to English or 'Lockdown measures in France' being exclusive to German. The most reasonable explanation is that the public interest in certain stories and topics varies between countries and can thus sometimes be found in one language but not in another.

1	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL
2	5G	X			Summer vacation rules	X			Dutch people buying from EU webshops			X
3	Articles about Fake news about the coronavirus	X	X	X	Vaccine against the coronavirus	X	X		EU agriculture			X
4	Asylum and migration in EU	X	X	X	Anti border closures	X			EU negotiations with Albania			X
5	Border closures	X	X	X	Anti Euro currency	X	X		Exit strategies in EU			X
6	Brexit	X	X	X	Anti Lockdown	X			Homophobic EU attorney fired			X
7	China censors newspaper piece by EU diplomats	X	X	X	Coronavirus deaths in Italy	X			Relaxation measures in Austria			X
8	Climate protection	X	X	X	Coronavirus numbers in Italy	X			Remdesivir medicine			X
9	Conflict between EU and Germany over travel reimbursements	X	X	X	Critical of Trump	X			The role of the Netherlands in the EU			X
10	Coronavirus in Europe	X	X		Cybercrime	X						
11	Criticism of Corona App	X			Data protection/Privacy	X	X					
12	Criticism of EU measures/actions	X	X	X	EU audit into Big Tech	X						
13	Criticism of usage of EU money for energy saving	X			EU strategy in Syria conflict	X						
14	Division among EU member states	X	X		EU support in Ireland sinking	X						
15	ECB verdict of Karlsruhe	X	X	X	EU travel ban	X						
16	Economical impact of coronavirus	X	X	X	EU-China relations	X						
17	Emergency laws in Hungary	X	X		European Initiative for corona app	X	X					
18	EU data protection laws	X			Fight over EU office in Belfast	X						
19	EU fight against moneylaundering	X	X		Internal conflicts in the EU	X						
20	EU mission "Irini"	X			Italy's conflict with Germany and the Netherlands over eurobonds	X						
21	EU report about Chinese propaganda	X	X		Northern Ireland issue in Brexit talks	X						
22	EU Solidarity funds	X	X	X	Protests against lockdown	X						
23	EU trade agreement with Mexico	X	X	X	Search for missing people from Kosovo war	X						
24	EU West Balkan Summit	X	X		Support for WHO	X						
25	EU-China conflict	X	X	X	Tourism and Travel	X						
26	Eurobonds	X	X		UK-EU relations post-Brexit	X						
27	Financial support for Italy	X			Ukraine sovereignty	X						
28	Germany's role in the EU	X	X		US medical academy supporting EU	X						
29	Hisbollah ban	X			VW Diesel scandal	X						
30	Libya conflict	X	X	X	WHO investigation into virus origin	X						
31	Lockdown measures in France	X			African swine flu in EU		X					
32	Medienstaatsvertrag	X			Call for solidarity by EU leaders		X					
33	Positive article about the role of the EU	X	X	X	Conflict over support for Palestinian groups		X					
34	Process against Polish justice reform	X	X	X	Conflict Weststellingwerf vs Vermilion over gas extraction		X					
35	Restrictions to freedom of the press	X			Copa-Cogeba buying potatoes in EU		X					
36	State help for Condor	X			Coronavirus in Africa	X						

Figure 4.18: Topics marked by me in Atlas.ti in the files for EU

When looking at how the actions of the EU are portrayed in the headlines, we can find that all three languages convey a rather negative image of the measures and actions of the EU, especially in English (50 negative vs. 11 positive), but also in German (20 negative vs. 8 positive) and Dutch (26 negative vs. 12 positive). As already mentioned in chapter 2, I was unfortunately only able to find a proper opinion survey/poll for Germany about how the EU actions, especially in relation to the coronavirus pandemic, have been perceived, which showed that a majority of Germans was not satisfied with the way the EU was handling the pandemic [14], which seems to be in line with the impression gained from examining the headlines. The high number of English headlines that portray the EU in a negative light is largely related to Brexit and points of conflict between the EU and the UK during the negotiations, since most of these headlines are from British tabloids that take the side of their government in those conflicts and are thus less related to the coronavirus pandemic specifically.

### 4.3.7 Italy

Just as in many of the previous sections, we can see from figure 4.19 that mainstream news outlets are dominating again, especially in German and

English. However an interesting observation we can make, is that there are multiple football (soccer) magazines like VTBL and Ajax Showtime among the list of Dutch sources. These sources are reporting on the Serie A and Italian football in general, with the season having been on hold in late March, April and May, with its fate having been uncertain during the time where the data was collected.

	German		English		Dutch	
	Source	No. %	Source	No. %	Source	No. %
3	<a href="#">merkur.de</a>	37 7.019	The New York Times	61 7.64	<a href="#">Telegraaf.nl</a>	39 13.64
4	Suddeutsche Zeitung	36 6.26	BBC News	58 7.27	<a href="#">AD.nl</a>	21 7.34
5	<a href="#">Stol.it</a>	31 5.88	Bloomberg	49 6.14	Noordhollands Dagblad	19 6.64
6	<a href="#">ORF.at</a>	26 4.93	The Guardian	42 5.26	NOS	18 6.29
7	Suedtirol News	20 3.8	Reuters	38 4.76	Wellingelichte Kringen	14 4.9
8	FOCUS Online	19 3.61	Daily Mail	26 3.26	Nederlands Dagblad	13 4.55
9	FAZ - Frankfurter Allgemeine Zeitung	18 3.42	Euronews	25 3.13	RTL Nieuws	12 4.2
10	Frankfurter Rundschau	16 3.036	Financial Times	24 3	Zeelandnet Nieuws	10 3.5
11	onvista	16 3.036	Al Jazeera English	19 2.38	Transport Online	10 3.5
12	WELT	15 2.85	The Local Italy	18 2.26	NRC	9 3.15
13	DER SPIEGEL	12 2.28	<a href="#">ABC News</a>	18 2.26	<a href="#">NU.nl</a>	8 2.8
14	n-tv NACHRICHTEN	12 2.28	Anadolu Agency	18 2.26	Volkskrant	7 2.45
15	RTL Online	11 2.087	<a href="#">Telegraph.co.uk</a>	17 2.13	<a href="#">Parool.nl</a>	7 2.45
16	Vatican News	11 2.087	CNBC	15 1.88	<a href="#">Dolcevia.com</a>	6 2.1
17	Deutschlandfunk	10 1.9	Reuters UK	14 1.75	De Stentor	6 2.1
18	<a href="#">tagesschau.de</a>	9 1.71	Forbes	12 1.5	Trouw	5 1.75
19	<a href="#">STERN.de</a>	9 1.71	Sky News	11 1.38	Het Laatste Nieuws	5 1.75
20	KURIER	9 1.71	DW (English)	9 1.13	Het Financieele Dagblad	5 1.75
21	Handelsblatt	9 1.71	TRT World	8 1	VTBL	4 1.4
22	<a href="#">domradio.de</a>	8 1.52	<a href="#">Express.co.uk</a>	8 1	Ajax Showtime	4 1.4
23	Neue Zürcher Zeitung	7 1.33	The Economist	8 1	<a href="#">Knack.be</a>	4 1.4
24	Neues Volksblatt	7 1.33	NPR	7 0.88	FCUpdate mobiel	4 1.4
25	Tagesspiegel	6 1.14	The Wall Street Journal	6 0.75	De Limburger	3 1.048
26	ZEIT ONLINE	6 1.14	CNN	6 0.75	<a href="#">Bladna.nl</a>	3 1.048
27	<a href="#">t-online.de</a>	6 1.14	FRANCE 24 English	6 0.75	De Morgen	3 1.048

Figure 4.19: The top 25 sources in each language with number of occurrences and percentage in its language

Interestingly enough, figure 4.20 shows rather different progressions in the number of headlines per day between the languages. However there is a common peak in headlines on 4 May 2020 in all three languages, which has to do with Italy announcing, that they are gonna relax some of their strict lockdown measures at that time. Those differences are most likely the result of different stories being reported in each language.

This hypothesis can be supported when examining the topic codes as seen in figure 4.21. Figure 4.21 shows that there are a lot of topics and stories that are featured in multiple languages or even all three languages such as the 'Mafia', 'Coronavirus deaths in Italy' or 'Relaxation measures in Italy', but also a lot of stories that are only featured in one language. English does feature especially many exclusive topics and stories, such as 'Conflict between Italian federal government and regional governments' or the 'ECB verdict of Karlsruhe'. The reason for this is probably that there are vastly more news outlets publishing in English than news outlets publishing in German or Dutch, therefore this is not very surprising.

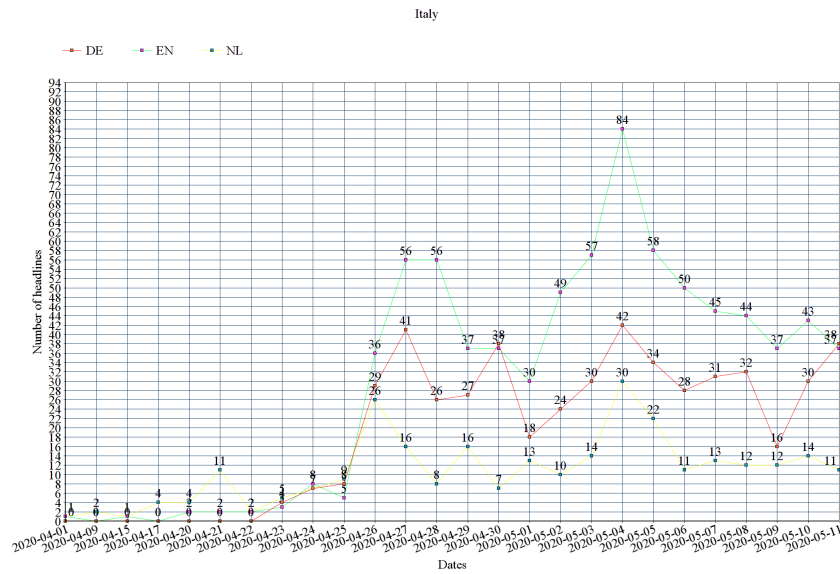


Figure 4.20: The number of headlines over time for each language

1	Code/language	DE	EN	NL	Code/language	DE	EN	NL
2	Against imports from China			X	Lockdown measures in Italy	X		
3	Coronavirus deaths in Italy	X	X	X	Personal stories	X	X	
4	Coronavirus in Lombardy			X	Pro-Relaxation	X	X	
5	Coronavirus numbers in Italy	X	X	X	Rising number of deaths after relaxation in Italy		X	
6	Coronavirus numbers UK	X	X	X	Sinking crime numbers in Italy		X	
7	Crisissituation in Italy	X	X	X	Tourism and Travel		X	X
8	Criticism of Italian government	X	X	X	Anti Euro currency		X	
9	Economical impact of coronavirus	X	X	X	Anti Lockdown		X	
10	EU Solidarity funds	X	X	X	Avalanche in Northern Italy		X	
11	Football/soccer in times of the coronavirus	X	X	X	China spreading fake news		X	
12	Italian health care workers infected	X	X	X	Conflict between Italian federal government and regional governments		X	
13	Italy-EU relations	X	X	X	Coronavirus in Brazil		X	
14	Italy's conflict with Germany and the Netherlands over eurobonds		X	X	Coronavirus worldwide		X	
15	Mafia	X	X	X	Critical of Trump		X	
16	Masks in Italy			X	Criticism of government lockdown measures		X	
17	People's behaviour in public			X	Criticism of relaxation measures		X	
18	Relaxation measures in Italy	X	X	X	DIY masks		X	
19	Sinking corona numbers in Italy		X	X	ECB verdict of Karlsruhe		X	
20	World War 2 Allies Victory Anniversary	X	X	X	EU-China conflict		X	
21	Asylum and migration in EU	X	X		Lockdown measures in France		X	
22	Corona app international	X			No proof of immunity after recovering from infection		X	
23	Coronavirus testing	X			Positive article about the role of the EU		X	
24	Criticism of EU measures/actions	X			Praise for Italian mayors		X	
25	Eurobonds	X	X		Relaxation measures in Spain		X	
26	Genoa bridge rebuilt	X	X		Vaccine against the coronavirus		X	
27	Less deaths in Italy	X	X					

Figure 4.21: Topics marked by me in Atlas.ti in the files for Italy

### 4.3.8 WHO

At this point, it should come as no surprise that most sources are mainstream news outlets and regional papers, as figure 4.22 shows. The only really noteworthy observation here is, that the website of the World Health Organization itself is the third most common source in English.

1	German			English			Dutch		
2	Source	No.	%	Source	No.	%	Source	No.	%
3	<a href="#">merkur.de</a>	35	5.81	CNBC	21	8.68	<a href="#">NU.nl</a>	36	9.65
4	<a href="#">tagesschau.de</a>	31	5.15	The Guardian	19	7.85	<a href="#">Telegraaf.nl</a>	30	8.042
5	Süddeutsche Zeitung	22	3.65	World Health Organization	14	5.79	<a href="#">AD.nl</a>	27	7.24
6	DW (Deutsch)	21	3.49	Reuters	13	5.37	NOS	23	6.17
7	<a href="#">t-online.de</a>	20	3.32	Reuters UK	10	4.13	Het Laatste Nieuws	20	5.36
8	WELT	17	2.82	BBC News	9	3.72	Dagblad van het Noorden	19	5.093
9	Deutschlandfunk	16	2.66	Daily Mail	9	3.72	Wellingelichte Kringen	15	4.021
10	<a href="#">Nau.ch</a>	15	2.49	World Economic Forum	8	3.31	Nederlands Dagblad	14	3.75
11	FOCUS Online	15	2.49	New York Post	6	2.48	Volkskrant	12	3.22
12	Tagesspiegel	15	2.49	CNA	6	2.48	RTL Nieuws	12	3.22
13	DER SPIEGEL	14	2.33	CNN	5	2.066	NRC	12	3.22
14	n-tv NACHRICHTEN	13	2.16	Al Jazeera English	5	2.066	<a href="#">Parool.nl</a>	11	2.95
15	onvista	12	1.99	Fox News	5	2.066	Noordhollands Dagblad	11	2.95
16	FAZ - Frankfurter Allgemeine Zeitung	11	1.83	NBC News	4	1.65	Leeuwarder Courant	10	2.68
17	RND	11	1.83	Bloomberg	4	1.65	Trouw	9	2.41
18	Telepolis	10	1.66	Anadolu Agency	4	1.65	De Dagelijkse Standaard	7	1.88
19	Business Insider Deutschland	10	1.66	<a href="#">Express.co.uk</a>	3	1.24	De Limburger	7	1.88
20	ZDFheute	10	1.66	NPR	3	1.24	VRT NWS	4	1.072
21	UNRIC	10	1.66	Financial Times	3	1.24	Transport Online	4	1.072
22	Die Presse	9	1.5	The New York Times	3	1.24	ThePostOnline	4	1.072
23	MDR	8	1.33	The Globe and Mail	3	1.24	De Morgen	4	1.072
24	<a href="#">ORF.at</a>	8	1.33	Focus Taiwan News Channel	3	1.24	Reformatisch Dagblad	4	1.072
25	<a href="#">STERN.de</a>	7	1.16	Jakarta Post	3	1.24	<a href="#">BD.nl</a>	3	0.8
26	Reuters Deutschland	7	1.16	Times Now	3	1.24	LINDA	3	0.8
27	Deutschlandfunk Kultur	7	1.16	CBBC Newsround	2	0.83	BN DeStem	3	0.8

Figure 4.22: The top 25 sources in each language with number of occurrences and percentage in its language

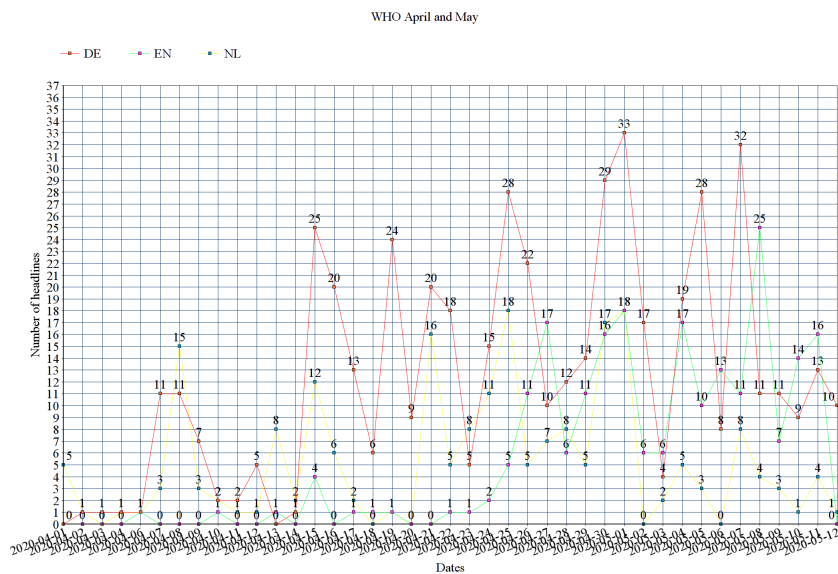


Figure 4.23: The number of headlines over time for each language (shortened version with only April and May)

One thing that is interesting to note is that German and Dutch have fairly similar progressions of the number of headlines per day, but English



has a fairly different progression, as seen in 4.23. There are two visible common peak in the number of headlines on 15 April 2020 and 1 May 2020 but otherwise the curves are pretty different from one another. The reason for this lies in the formulation of the search term.

When using 'who' as a search term, almost all headlines in German or Dutch will be about the World Health Organization (those that were not were about the band The Who). However in English 'who' is also a common word within the language, so there were a lot of headlines which were in no way related to the World Health Organization. I had to filter out those headlines. However, this still means that, among the headlines my script examined, there were less about the WHO than in German or Dutch, so some topics might be missing due to that. In hindsight it would have been better to use 'world health organization' as the English search term to avoid this.

Another thing I have to note here is that figure 4.23 is an abridged version of the progression chart for the number of headlines for the search term 'WHO'. The full chart goes back to the summer of 2019, but I used the abridged version since the before April the curves are not very interesting and the full chart is almost unreadable.

1	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL	Code/Language	DE	EN	NL
2	Anti-relaxation	X		X	Virus origin in animals	X	X	X	Coronavirus number US			X
3	Articles about Fake news about the coronavirus	X		X	Virus originated from Wuhan lab' theory	X	X	X	Coronavirus numbers UK			X
4	Bolsonaro: WHO encourages masturbation and homosexuality	X			WHO chief Tedros	X			Cyberattacks on WHO			X
5	China blocking the WHO	X		X	WHO criticizing countries	X	X	X	Remdesivir medicine			X
6	China censors newspaper piece by EU diplomats	X			WHO defending/praising China	X	X	X	Vaccines			X
7	China-Taiwan conflict	X	X		WHO employee killed in Myanmar	X			WHO approving US Federal coronavirus plan			X
8	Coronavirus declared pandemic	X	X	X	WHO investigation into virus origin	X						
9	Coronavirus in Africa	X	X	X	WHO praising Sweden	X	X					
10	Coronavirus in Europe	X	X	X	WHO recommends playing videogames	X						
11	Coronavirus worldwide	X	X	X	WHO refusing to recognize Taiwan	X	X	X				
12	Critical of Trump	X	X	X	WHO warning pandemic nowhere near over	X	X					
13	Criticism of (DIY) masks	X			Anti border closures			X				
14	Criticism of Corona App	X			Anti Lockdown			X				
15	Criticism of government lockdown measures	X			Border closures			X				
16	Criticism of relaxation measures	X	X	X	Burn-out officially declared illness			X				
17	Criticism of the WHO	X	X	X	Call for independant investigation of WHO's approach		X	X				
18	Criticism of Tedros	X		X	Call for lockdown measures			X				
19	Drosten contradicts WHO: "Healed patients immune"	X			Coronavirus in South America		X	X				
20	Economical impact of coronavirus	X		X	Coronavirus number DE			X				
21	Explanation of the WHO	X	X		Coronavirus testing		X	X				
22	Immunity passports	X	X	X	Covid spread through air			X				
23	More domestic violence during lockdown	X		X	Criticism of governments handling in Spain			X				
24	No proof of immunity after recovering from infection	X	X	X	Ebola in the Congo			X				
25	Pandemic plan of East Germany (DDR)	X			Football/soccer in times of the coronavirus			X				
26	Praise for Sweden's measures	X			Measle epidemic in the Congo			X				
27	Praise for Taiwan	X	X	X	Most infections at home			X				
28	Reports of China pressuring WHO to suppress information about virus	X	X		Olympic games			X				
29	Support for Trump in conflict with WHO	X		X	Personal stories			X				
30	Support for WHO	X	X	X	Plastic in the drinking water			X				
31	Support/Praise for German government	X			Smoking			X				
32	Trump-WHO conflict	X	X	X	Social Distancing			X				
33	Unrelated to WHO	X		X	Trump criticizing China			X				
34	US stopping WHO financing	X		X	Trump press briefings			X				
35	USA conflict with Germany over WHO	X			WHO data leak			X				
36	Vaccine against the coronavirus	X	X	X	Coronavirus in Brazil	X						

Figure 4.24: Topics marked by me in Atlas.ti in the files for WHO

The previous observations in this section are also reflected in the topics covered in each language, which is visualized in figure 4.24. Here we can see that a significant amount of topics are covered in both Dutch and German and that English has by far the least amount of covered topics. English

also has few exclusive topics, most of which are about the US or the UK. Dutch has the largest amount of exclusive topics, many of which however are unrelated to the coronavirus pandemic, such as 'Burn-out officially declared illness', 'Ebola in the Congo' or 'Plastic in the drinking water'. The headlines of these stories are often from 2019.

In chapter 2 I mentioned that I was not able to find any polls/surveys on about how people felt about the actions of the WHO apart from one survey that said that 55 percent of US voters approved the WHO's actions and 30 percent did not [40]. If one were to base a conclusion on how the people of Germany, the Netherlands, the UK and the USA feel about the actions of the WHO just from looking at the headlines collected and the codes I assigned to them, one would get the impression that that the WHO and their actions were seen negatively in most of these countries, since in English (8 critical vs. 4 supportive) and Dutch (22 critical vs. 7 supportive) the number of critical headlines is much higher than the number of supportive headlines. German is a noticeable exception, since the the numbers of critical and supportive headlines are very similar (20 critical vs. 18 supportive).

## Chapter 5

# Discussion and Conclusions

### 5.1 Summary and Discussion of Observations

In chapter 4 we examined the results for each individual search term. From those observation we can derive some more general observations.

The first general observation we can derive is that the vast majority of sources with very few exceptions found in all three languages are among the most popular news websites in their respective languages. However in some cases there were also more obscure sources. For example for the search term "corona tracking app" the sources in all three languages included technology magazines and websites or for the search term "eu" there were economic and financial websites among the list of sources. Thus, while sources in all languages mostly reflect what news outlets are popular in the respective countries, outlets that specialize on a very specific area related to the topic such as technology for the tracking app are featured as well.

The results of our study also support the observation made by Haim et al. (2017), that some news outlets are featured much more frequently than others [18]. For example we can see that in German, news outlets like merkur.de, Süddeutsche Zeitung or DER SPIEGEL make up a much bigger percentage of all headlines than most other news outlets like e.g. BILD.de or FAZ. The same goes e.g. for The Guardian, the BBC, Financial Times and New York Times in English and AD.nl, NOS, NU.nl and Telegraaf.nl in Dutch.

The second possible observation is that the news stories featured for each search term can differ vastly between languages. For some search terms the featured news stories were more similar among the three languages, e.g. for the search terms "donald trump", "eu" and "corona tracking app", there were a lot of news stories that were featured in at least two languages, while for search terms like "corona relaxation" or "corona mask requirement" a lot of news stories were only covered in one language. As discussed in the previous chapter, this often came down to regional relevance. For example the actions of the Dutch or Belgian government are often only reported in



Dutch, since outside of Belgium or the Netherlands, there is little public interest for this topic, while many Brexit specific topics for example are only really interesting for people in the UK and were therefore only reported in English. On the other hand, stories like the conflict between Donald Trump and the WHO or the economic consequences of COVID-19 were just as relevant for the German and the Dutch public as they were for American and British public and were thus found in all of the three languages. English covered the most amount of news stories due to English being the most widely spoken language worldwide, which means that news in the English language are published all over the world, while German and Dutch are only spoken in a small amount of countries and thus news in those languages is only published in very few countries.

When it comes to headlines that highlight certain opinions on the overall news topics, tendencies were often very similar in all three languages. For example all three languages featured mostly negative reports about Donald Trump and were critical of relaxing the measures taken against the coronavirus. However on some topics there was a visible difference in the opinion ratio, e.g. in German there were more negative headlines on face masks than positive ones, while in English and Dutch the opposite was the case. And the opinions don't always reflect the public opinion of their country, since for example, the use of face masks was generally seen positively in Germany, which is the opposite of what the headlines would suggest. Overall, while there was a visible tendency towards one side of the discussion on the topics for most of the topics, with the exception of Donald Trump, both sides of the discussion were usually featured, just that one side was featured more prominently than the other side. Hence, there seems to be no real visible "echo chamber" in any language, with the exception maybe for the topic of Donald Trump in all three languages, where the reporting was very one-sided and negative, however in that case this mostly reflects the majority opinion on Donald Trump in the respective countries.

## 5.2 Limitations

Before we make our final conclusions and answer the research question, it should be noted that all results from this study are limited to Google News. As already mentioned in earlier chapters, Google News is not the primary way people consume news online, but instead around 75 percent of the time news is accessed directly from the websites of news outlets [16].

In May 2020 Google News was visited 570.11 million times worldwide [8], making it only the 12th most accessed news website behind sites like yahoo.com, cnn.com, dailymail.co.uk, news.yahoo.co.jp and bbc.co.uk [9]. However as an international search engines for news, which does not include the Japanese version of Yahoo News, it is still the most widely used of its kind,

so it makes sense to conduct such a study on this website rather than a different website.

However that still means that the results of the study are specifically limited to Google News and can not be directly related to a similar website like e.g. Yahoo News or online news consumption overall, since there are many different options for the consumption of news online that were not examined (these, for example, include social media or direct access from the websites of news outlets).

## 5.3 Conclusions

The aim of this research was to find out, if there is a visible difference in the results when searching for current political topics on Google news in different languages. After examining the results, we can conclude that the results are often indeed somewhat different in each language. More specifically, there are often different news stories covered in each language and sometimes the overall tendency towards positive or negative articles can be different for different languages. However we also saw that the headlines found on Google News mostly come from mainstream news outlets of the countries where the news in a specific languages are published. So it seems that, what results one can find in a language, depends on what mainstream news outlets report. For example if the mainstream news outlets in one's search languages are all mainly critical of Donald Trump, it is unlikely that one might find many positive articles on Donald Trump in any of those search languages.

In general, it can be concluded that a multilingual user on Google News does only get a more diverse perspective on a topic in the sense, that a multilingual user is able to find a bigger amount of news stories for the same search term than a monolingual user, which is not very surprising. For example for the search term "corona relaxation" a monolingual user that only speaks German might find a lot of news about the state of the coronavirus in Germany and Austria, which measures were introduced by the government to stop its spread and which measures are about to be relaxed, and see, what politicians and virologists think of those relaxations. This user would not really get informed about the situations and opinions in other countries like the Netherlands, the UK or the United States or stories about a rise in domestic violence during the lockdown. A multilingual user would be able to see articles on these stories in English or Dutch in this case.

As already stated previously, most of the time, the Google News results did feature articles with different perspectives in each individual language. However, it should be noted, that, as discussed in section 5.1 or seen in chapter 4, one perspective often had more headlines leaning towards it than

other perspectives, which could be interpreted as a moderate echo chamber effect. This seems to be partially supporting both hypotheses mentioned by Flaxman, Goel and Rao in the beginning of their 2016 study (see also chapter 3)[16].

- Their first hypothesis, that users do get to view a lot of the same content (often content they are likely to agree with) [16], is supported in the sense, that, often, one perspective was featured much more frequently than the other. The best example for this is the search term "donald trump", which featured a lot of headlines that portrayed Donald Trump in a negative way (German: 72, English: 123, Dutch: 29), but only very few headlines that were somewhat positive towards him (German: 2, English: 3, Dutch: 3). Also the perspective, that was featured more frequently, was, as already mentioned earlier, often (but not always) the most popular opinion for the corresponding country, which would also support this hypothesis.
- Their second hypothesis, that the large number of options when consuming news online will lead to greater exposure to different perspectives [16], is also supported, since as already mentioned, other perspectives were almost always also featured, sometimes the number of headlines was even equally distributed, e.g. for Dutch on the topic of relaxation (15 negative vs. 16 positive), for the WHO in German (20 negative headlines vs. 18 positive headlines) or in German for Corona Tracking Apps (10 negative headlines vs. 12 positive headlines).

However, the results of this study indicate that the chosen search language does act as a kind of filter bubble, not in the classical sense of an echo chamber (even though one could argue that there is a moderate echo chamber effect, as mentioned in the previous paragraph), but rather a filter, that determines which news stories you see and which news stories you do not see, based on what is reported by popular mainstream national and regional news outlets. This search language-based filter bubble can be expanded through the use of multiple search languages, which gives access to more news stories, often stories concerning the countries of the publishing news outlets and sometimes different perspectives, that are often - but not always - based on what perspectives are popular among the general population and mainstream media of the particular country.

In the concrete context of the coronavirus pandemic, one gets to see how the situation develops in different countries, which are maybe less prominently featured in the native search language or with less detail or nuance, how different governments deal with the situation and how action by international organizations are perceived in different countries. This provides more context on which one could base their personal opinions on these issues and their perception of the overall situation. An example for this would be the

search term "corona measures". If a Dutch user were to also search for news in German or English, he would find more concrete information on the measures taken in Germany or the UK and how the people in these countries feel about these measures, which he would not see in Dutch news. The user would also find information on a lot of other countries, especially in English, where there are news stories from all over the world, e.g. from Hong Kong, Mexico, Australia and Singapore. Based on this information, he could compare it to his home country and then in this context evaluate how his own country is handling the pandemic in comparison to other countries. If one were to stick to just one search language, presumably their native language, the given context would be smaller, since the amount of news stories they are aware of would be smaller.

## 5.4 Further discussion

While the research was able to answer the original research question, the headlines collected could be examined even further. Since I had to do the analysis by hand and went through each file one by one, I could not examine every aspect due to time constraints. Most notably, I did not examine how often specific news stories were reported and instead just marked them once per file they occurred in. If one were to analyze this, especially for news stories that were found in more than one language, one could potentially answer the research question in more depth. Due to time constraints I also was not able to analyze all the data I collected. For example the script also collected all the headlines on the front page of Google News, when one does not enter any search term. The three files created from this contain the same amount of data as all the other created files combined, which meant that I unfortunately did not have the time to analyze them. However analyzing these files might be interesting in their own right, since it gives an insight into what news stories were seen as important in each language.

If someone were to do further research with the collected data, one could, for example, train a machine-learning algorithm to analyze headlines and categorize them by news story or reporting style (i.e. negative and positive articles). As I also mentioned, analyzing the files with the headlines from the Google News front page could also be part of future research to gain insight into the interest in different news stories in different countries in the period of 24 April 2020 to 12 May 2020 and beyond (on Github the data goes up to 21 May 2020 and at <https://matti.sermak.xyz/> the data has been collected on a server from 24 April 2020 to 23 June 2020 every two hours). All scripts (see Appendix A, B and C), collected data and analysis results can be found in the Github repository.<sup>1</sup>

---

<sup>1</sup><https://github.com/emysliwietz/newsscraaper>

## 5.5 Acknowledgements

Finally, I want to thank my friend Egidius Mysliwietz, who wrote most the original python script, that collected the headlines, and created the Github repository, my supervisor Eelco Herder, who provided me with very helpful feedback, helped me come up with the next step and provided me with some helpful sources, and my second reader Djoerd Hiemstra, who also provided me with helpful feedback.

# Bibliography

*Non-scientific sources are marked in blue.*

- [1] Anne Aula and Melanie Kellar. Multilingual search strategies. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA 09*, page 3875–3870, Apr 2009. <https://dl.acm.org/doi/10.1145/1520340.1520585>.
- [2] Piet Bakker. Dit zijn de 39 nederlandse online nieuwsmerken met het grootste bereik, May 2018. <https://www.svdj.nl/de-stand-van-de-nieuwsmedia/39-online-nieuwsmerken-bereik/>.
- [3] Megan Brennan. Americans’ trust in mass media edges down to 41%, Sep 2019. <https://news.gallup.com/poll/267047/americans-trust-mass-media-edges-down.aspx>.
- [4] J. Clement. Search engine market share worldwide 2019, Mar 2020. <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>.
- [5] AP News Agency. Poll: Most americans favour lockdowns to curb coronavirus, 22 Apr 2020. <https://www.aljazeera.com/news/2020/04/poll-americans-favour-lockdowns-curb-coronavirus-200422183545091.html>.
- [6] Cambridge Dictionary. Filter bubble: meaning in the cambridge english dictionary. <https://dictionary.cambridge.org/dictionary/english/filter-bubble>.
- [7] Eurostat. 65% know at least one foreign language in the eu, Sep 2018. <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20180926-1>.
- [8] SimilarWeb. news.google.com traffic statistics may 2020, Jun 2020. <https://www.similarweb.com/website/news.google.com>.

- [9] SimilarWeb. Top sites ranking for news and media in the world, Jun 2020. <https://www.similarweb.com/top-websites/category/news-and-media>.
- [10] Statista Research Department. Sorgen wegen der politik von donald trump 2019, Feb 2019. <https://de.statista.com/statistik/daten/studie/664841/umfrage/umfrage-zu-sorgen-wegen-der-politik-von-donald-trump/>.
- [11] Statista Research Department. Reichweite von nachrichtenwebsites in deutschland 2020, Apr 2020. <https://de.statista.com/statistik/daten/studie/165258/umfrage/reichweite-der-meistbesuchten-nachrichtenwebsites/>.
- [12] YouGov. The most popular newspapers in the uk: Media: Yougov ratings, 2020. <https://yougov.co.uk/ratings/media/popularity/newspaper/all>.
- [13] ZDF. Politbarometer extra: Mehrheit: Lockerungen so richtig, 08 May 2020. <https://www.zdf.de/nachrichten/politik/politbarometer-extra-coronavirus-lockerungen-gastronomie-bundesliga-100.html?slide=1587653772463>.
- [14] ZDF. Politbarometer: Mehrheit befürwortet coronavirus-lockerungen, 24 Apr 2020. <https://www.zdf.de/nachrichten/politik/politbarometer-mehrheit-befuerwortet-lockerungen-coronavirus-100.html>.
- [15] Benjamin Fearnow. Donald trump’s popularity in the u.k. remains far, far lower than barack obama’s, Jun 2019. <https://www.newsweek.com/how-popular-donald-trump-uk-presidents-favorability-rating-far-lower-barack-oba>
- [16] Seth Flaxman, Sharad Goel, and Justin M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1):298–320, 2016. <https://5harad.com/papers/bubbles.pdf>.
- [17] Sharad Goel, Winter Mason, and Duncan J. Watts. Real and perceived attitude agreement in social networks. *Journal of Personality and Social Psychology*, 99(4):611–621, 2010. <https://5harad.com/papers/friendsense.pdf>.
- [18] Mario Haim, Andreas Graefe, and Hans-Bernd Brosius. Burst of the filter bubble? effects of personalization on the diversity of google news. *Digital Journalism*, 6(3):330–343, Jul 2017. [https://www.researchgate.net/publication/318256136\\_Burst\\_of\\_the\\_Filter\\_Bubble\\_Effects\\_of\\_personalization\\_on\\_the\\_diversity\\_of\\_Google\\_News](https://www.researchgate.net/publication/318256136_Burst_of_the_Filter_Bubble_Effects_of_personalization_on_the_diversity_of_Google_News).

- [19] Shanto Iyengar and Kyu S Hahn. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1):19–39, 2009. <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1460-2466.2008.01402.x>.
- [20] Tobias D. Krafft, Michael Gamer, and Katharina A. Zweig. What did you see? personalization, regionalization and the question of the filterbubble in google’s search engine, 2016. <https://arxiv.org/pdf/1812.10943.pdf>.
- [21] Chenjun Ling, Ben Steichen, and Silvia Figueira. Multilingual news – an investigation of consumption, querying, and search result selection behaviors. *International Journal of Human–Computer Interaction*, 36(6):516–535, Sep 2019. <https://www.tandfonline.com/doi/full/10.1080/10447318.2019.1662636>.
- [22] Kim Loohuos. Coronavirus: Dutch covid-19 tracking app stirs national debate, 24 Apr 2020. <https://www.computerweekly.com/news/252482131/Coronavirus-Dutch-Covid-19-tracking-app-stirs-national-debate>.
- [23] Edward Malnick. Support for wearing face masks in public outweighs opposition, 18 Apr 2020. <https://www.telegraph.co.uk/news/2020/04/18/public-support-wearing-face-masks-public-outweighs-opposition/>.
- [24] Katerina Eva Matsa, Laura Silver, Elisa Shearer, and Mason Walker. Western europeans under 30 view news media less positively, rely more on digital platforms than older adults, Oct 2018. <https://www.journalism.org/2018/10/30/western-europeans-under-30-view-news-media-less-positively-rely-more-on-digital>.
- [25] Robin McKie, Lin Jenkins, and Toby Helm. Fearful britons remain strongly opposed to lifting coronavirus lockdown, 02 May 2020. <https://www.theguardian.com/world/2020/may/02/fearful-britons-oppose-lifting-lockdown-schools-pubs-restaurants-opinion-poll>.
- [26] Solomon Messing and Sean J. Westwood. Selective exposure in the age of social media. *Communication Research*, 41(8):1042–1063, Dec 2012. <https://journals.sagepub.com/doi/full/10.1177/0093650212466406>.
- [27] Office of Communication. News consumption in the uk: 2019 report - ofcom, Jul 2019. [https://www.ofcom.org.uk/\\_\\_data/assets/pdf\\_file/0027/157914/uk-news-consumption-2019-report.pdf](https://www.ofcom.org.uk/__data/assets/pdf_file/0027/157914/uk-news-consumption-2019-report.pdf).
- [28] Eli Pariser. *The Filter Bubble: What the Internet is Hiding from You*. Penguin Books, 2011.



- [29] Lars Pasveer. Vertrouwen in nederlandse media hoog, lage betalingsbereidheid voor nieuws / villamedia, Jun 2019. <https://www.villamedia.nl/artikel/vertrouwen-in-nederlandse-media-hoog-lage-betalingsbereidheid-voor-nieuws>.
- [30] David Ruch. So gut wird die mundschutzpflicht umgesetzt, 03 May 2020. [https://www.t-online.de/nachrichten/deutschland/id\\_87808644/umfrage-zur-maskenpflicht-so-gut-wird-die-mundschutzpflicht-umgesetzt.html](https://www.t-online.de/nachrichten/deutschland/id_87808644/umfrage-zur-maskenpflicht-so-gut-wird-die-mundschutzpflicht-umgesetzt.html).
- [31] Michael Savage. Only 50% of britons would download nhs tracing app – poll, 10 May 2020. <https://www.theguardian.com/world/2020/may/10/only-50-of-britons-would-download-nhs-tracing-app-poll>.
- [32] Christian Schemer, Nikolaus Jakob, Oliver Quiring, Tanjev Schultz, Marc Ziegele, and Ilka Jakobs. Forschungsergebnisse der welle 2019, Feb 2020. <https://medienvertrauen.uni-mainz.de/forschungsergebnisse-der-welle-2019/>.
- [33] Jens Schröder. Ivw-news-top-50: Deutsche online-nachrichtenangebote so populär wie nie zuvor – visits-rekorde für bild, upday, welt, zeit, stern und f.a.z. ' meedia, Feb 2018. <https://meedia.de/2018/02/08/ivw-news-top-50-deutsche-online-nachrichtenangebote-so-populaer-wie-nie-zuvor-v>
- [34] Gabriela Schulte. Poll: Majority of voters support mandatory face masks in some public settings, 25 May 2020. <https://thehill.com/hilltv/what-americas-thinking/499223-poll-majority-of-voters-support-mandatory-face-masks-in-some>.
- [35] Nate Silver. How popular is donald trump?, May 2020. <https://projects.fivethirtyeight.com/trump-approval-ratings/>.
- [36] Loren Snel. Ov-gebruiker pakt liever fiets dan mondkapje, 10 May 2020. <https://www.metronieuws.nl/in-het-nieuws/2020/05/ov-gebruiker-pakt-liever-fiets-dan-mondkapje/>.
- [37] Craig Timberg, Drew Harwell, and Alauna Safarpour. Most americans are not willing or able to use an app tracking coronavirus infections. that's a problem for big tech's plan to slow the pandemic., 29 Apr 2020. <https://www.washingtonpost.com/technology/2020/04/29/most-americans-are-not-willing-or-able-use-an-app-tracking-coronavirus-infections/>.
- [38] Caspar van Oirschot. Peiling — nederland smacht naar versoepeling corona-regime ..., 20 Apr 2020. <https://www.destentor.nl/zwolle/peiling-nederland-smacht-naar-versoepeling-corona-regime-maar-slechts-helft-ver>

- [39] Simon van Teutem. Nederlandse media overdrijven tegenstellingen enorm, Aug 2018. <https://www.quotenet.nl/nieuws/a216876/nederlandse-media-overdrijven-tegenstellingen-enorm-216876/>.
- [40] James Walker. World health organization approval rating tanks by 29 points in less than a month: Poll, 21 Apr 2020. <https://www.newsweek.com/world-health-organization-approval-rating-tanks-29-points-poll-1499163>.
- [41] Amy Watson. Leading u.s. news websites by unique monthly visitors 2020, Feb 2020. <https://www.statista.com/statistics/381569/leading-news-and-media-sites-usa-by-share-of-visits/>.
- [42] Bernhard Weidenbach. Nachrichtenquellen mit der größten vertrauenswürdigkeit 2019, Oct 2019. <https://de.statista.com/statistik/daten/studie/877238/umfrage/ranking-der-vertrauenswuerdigsten-nachrichtenquellen-in-deutschland/>.
- [43] Thomas Worley. How much does the uk public trust the media?, May 2019. <https://www.bond.org.uk/news/2019/05/how-much-does-the-uk-public-trust-the-media>.

## Appendix A

# Python Script for Data Collection

This is the script used to extract the headlines from Google news:

```
#!/bin/env python3
```

```
import requests
from bs4 import BeautifulSoup
import os
```

```
# Name of file, Url to get data from
```

```
urls = [
    ('corona_maßnahmen_de.txt',
     'https://news.google.com/search?q=corona%20ma%C3%9Fnahmen&hl=de&gl=DE&ceid=DE%3Aen'),
    ('corona_measures_en.txt',
     'https://news.google.com/search?q=corona%20measures&hl=en-GB&gl=GB&ceid=GB%3Aen'),
    ('corona_maatregelen_nl.txt',
     'https://news.google.com/search?q=corona%20maatregelen&hl=nl&gl=NL&ceid=NL%3Aen'),
    ('who_de.txt',
     'https://news.google.com/search?q=who&hl=de&gl=DE&ceid=DE%3Ade'),
    ('who_en.txt',
     'https://news.google.com/search?q=who&hl=en-GB&gl=GB&ceid=GB%3Aen'),
    ('who_nl.txt',
     'https://news.google.com/search?q=who&hl=nl&gl=NL&ceid=NL%3Anl'),
    ('donald_trump_de.txt',
     'https://news.google.com/search?q=donald%20trump&hl=de&gl=DE&ceid=DE%3Ade'),
    ('donald_trump_en.txt',
     'https://news.google.com/search?q=donald%20trump&hl=en-GB&gl=GB&ceid=GB%3Aen'),
    ('donald_trump_nl.txt',
     'https://news.google.com/search?q=donald%20trump&hl=nl&gl=NL&ceid=NL%3Anl'),
    ('eu_de.txt',
```

```

        'https://news.google.com/search?q=eu&hl=de&gl=DE&ceid=DE%3Ade'),
    ('eu_en.txt',
     'https://news.google.com/search?q=eu&hl=en-GB&gl=GB&ceid=GB%3Aen'),
    ('eu_nl.txt',
     'https://news.google.com/search?q=eu&hl=nl&gl=NL&ceid=NL%3Anl'),
    ('italien_de.txt',
     'https://news.google.com/search?q=italien&hl=de&gl=DE&ceid=DE%3Ade'),
    ('italy_en.txt',
     'https://news.google.com/search?q=italy&hl=en-GB&gl=GB&ceid=GB%3Aen'),
    ('italie_nl.txt',
     'https://news.google.com/search?q=italie&hl=nl&gl=NL&ceid=NL%3Anl'),
    ('corona_maskenpflicht_de.txt',
     'https://news.google.com/search?q=corona%20maskenpflicht&hl=de&gl=DE&ceid=DE%3Ade'),
    ('corona_mask_requirement.txt',
     'https://news.google.com/search?q=corona%20mask%20requirement&hl=en-GB&gl=GB&ceid=GB%3Aen'),
    ('corona_masker_nl.txt',
     'https://news.google.com/search?q=corona%20masker&hl=nl&gl=NL&ceid=NL%3Anl'),
    ('corona_tracking_app_de.txt',
     'https://news.google.com/search?q=corona%20tracking%20app&hl=de&gl=DE&ceid=DE%3Ade'),
    ('corona_tracking_app_en.txt',
     'https://news.google.com/search?q=corona%20tracking%20app&hl=en-GB&gl=GB&ceid=GB%3Aen'),
    ('corona_tracking_app_nl.txt',
     'https://news.google.com/search?q=corona%20tracking%20app&hl=nl&gl=NL&ceid=NL%3Anl'),
    ('corona_lockerung_de.txt',
     'https://news.google.com/search?q=corona%20lockerung&hl=de&gl=DE&ceid=DE%3Ade'),
    ('corona_relaxation_en.txt',
     'https://news.google.com/search?q=corona%20relaxation&hl=en-GB&gl=GB&ceid=GB%3Aen'),
    ('corona_versoepeling_nl.txt',
     'https://news.google.com/search?q=corona%20versoepeling&hl=nl&gl=NL&ceid=NL%3Anl'),
    ('schlagzeilen_de.txt',
     'https://news.google.com/topstories?hl=de&gl=DE&ceid=DE:de'),
    ('headlines_en.txt',
     'https://news.google.com/topstories?hl=en-GB&gl=GB&ceid=GB:en'),
    ('artikelkoppen_nl.txt',
     'https://news.google.com/topstories?hl=nl&gl=NL&ceid=NL:nl')
]

if not os.path.exists("news"):
    os.mkdir("news")

for (filename, url) in urls:
    filename = os.path.join("news", filename)
    if os.path.exists(filename):
        with open(filename, "r", encoding='utf-8') as f:

```

```

        lines_already_in_file = f.read()
        write_mode = "a" # Append if file exists
    else:
        lines_already_in_file = ""
        write_mode = "w" # And create file (overwrite) if it does not

    lines_to_append = ""
    page = requests.get(url)
    soup = BeautifulSoup(page.content, 'html.parser')
    titles = soup.findAll("a", {"class": "DY5T1d"})
    sources = soup.findAll("a", {"class": "wEwYrc AVN2gc uQIVzc Sksgp"})
    timestamps = soup.findAll("time", {"class": "WW6dff uQIVzc Sksgp"})

    articles = zip(titles, sources, timestamps)
    i = 0
    for (title, source, timestamp) in articles:
        i += 1
        line = "{} | {} | {}\n".format(title.text, source.text,
                                       timestamp['datetime'])

        if line not in lines_already_in_file:
            lines_to_append += line
        if i > 24:
            break

    with open(filename, write_mode, encoding='utf-8') as f:
        f.write(lines_to_append)
    print("Finished updating {}".format(filename))

```

Acknowledgement go out to Egidius Mysliwietz for helping me with the script. The script and all files created by the script can be found in the Github repository (<https://github.com/emysliwietz/newsscrafer>)

## Appendix B

# Python Script for analyzing the headline files

This is the script used to analyze the files created by the data collection script:

```
#!/bin/env python3

import os

# Names of files
files_de = [
    ('corona_maßnahmen_de.txt'),
    ('who_de.txt'),
    ('donald_trump_de.txt'),
    ('eu_de.txt'),
    ('italien_de.txt'),
    ('corona_maskenpflicht_de.txt'),
    ('corona_tracking_app_de.txt'),
    ('corona_lockerung_de.txt'),
    ('schlagzeilen_de.txt')
]

files_en = [
    ('corona_measures_en.txt'),
    ('who_en.txt'),
    ('donald_trump_en.txt'),
    ('eu_en.txt'),
    ('italy_en.txt'),
    ('corona_mask_requirement.txt'),
    ('corona_tracking_app_en.txt'),
    ('corona_relaxation_en.txt'),
```

```

        ('headlines_en.txt')
    ]

files_nl = [
    ('corona_maatregelen_nl.txt'),
    ('who_nl.txt'),
    ('donald_trump_nl.txt'),
    ('eu_nl.txt'),
    ('italie_nl.txt'),
    ('corona_masker_nl.txt'),
    ('corona_tracking_app_nl.txt'),
    ('corona_versoepeling_nl.txt'),
    ('artikelkoppen_nl.txt')
]

files = [
    ('corona_maßnahmen_de.txt'),
    ('who_de.txt'),
    ('donald_trump_de.txt'),
    ('eu_de.txt'),
    ('italien_de.txt'),
    ('corona_maskenpflicht_de.txt'),
    ('corona_tracking_app_de.txt'),
    ('corona_lockerung_de.txt'),
    ('schlagzeilen_de.txt'),
    ('corona_measures_en.txt'),
    ('who_en.txt'),
    ('who_en_with_lines_removed.txt'),
    ('donald_trump_en.txt'),
    ('eu_en.txt'),
    ('italy_en.txt'),
    ('corona_mask_requirement.txt'),
    ('corona_tracking_app_en.txt'),
    ('corona_relaxation_en.txt'),
    ('headlines_en.txt'),
    ('corona_maatregelen_nl.txt'),
    ('who_nl.txt'),
    ('donald_trump_nl.txt'),
    ('eu_nl.txt'),
    ('italie_nl.txt'),
    ('corona_masker_nl.txt'),
    ('corona_tracking_app_nl.txt'),
    ('corona_versoepeling_nl.txt'),
    ('artikelkoppen_nl.txt')
]

```

```

]

if not os.path.exists("analysis"):
    os.mkdir("analysis")

if not os.path.exists("analysis/source_count"):
    os.mkdir("analysis/source_count")

if not os.path.exists("analysis/date_count"):
    os.mkdir("analysis/date_count")

if not os.path.exists("analysis/word_count"):
    os.mkdir("analysis/word_count")

def sort_lines(lines):
    lines_split = lines.split("\n")
    lines_list = []
    lines_for_file = ""
    for line in lines_split:
        if line == '':
            break

        numbers = line.split(": ")
        lines_list.append((numbers[0], int(numbers[1])))

    lines_list.sort(key=lambda x:x[1], reverse=True)
    length = len(lines_list) - 1
    for i in range(length):
        line = "{}: {}\n".format(lines_list[i][0], lines_list[i][1])
        lines_for_file += line

    return lines_for_file

def calculate_percentages(lines, total_headlines):
    lines_split = lines.split("\n")
    percentages = []
    lines_for_file = ""
    for line in lines_split:
        if line == '':
            break

        index = lines.index(line)

```



```

        numbers = line.split(": ")
        num = int(numbers[1])
        percentages.append(num/total_headlines * 100)

length = len(lines_split) - 1
for i in range(length):
    line = "{} ({}%)\n".format(lines_split[i], str(percentages[i]))
    lines_for_file += line

return lines_for_file

def get_sources_and_occurences(lines, sources, occurences):
    for line in lines:
        if line == '':
            break

        line_elements = line.split(' | ')
        source_index = len(line_elements) - 2
        source = line_elements[source_index]
        if source not in sources:
            if ": " in source:
                source = source.replace(": ", " - ")
                sources.append(source)
                occurences.append(1)
            else:
                index = sources.index(source)
                occurences[index] += 1

    return sources, occurences

def perform_source_analysis(file, sources, occurences, total_headlines, unrelated_lines):
    original_path = os.path.join("news", file)
    with open(original_path, "r", encoding='utf-8') as f:
        lines_in_file = f.read()

    lines = lines_in_file.split("\n")
    #if file == "who_en.txt":
    #     for line in lines:
    #         if not ("WHO" in line or "World Health Organization" in line):
    #             unrelated_lines += 1
    #             lines.remove(line)

```

```

total_headlines += len(lines)
sources, occurrences = get_sources_and_occurences(lines, sources, occurrences)
return sources, occurrences, total_headlines, lines, unrelated_lines

def write_to_file(sources, occurrences, filepath, write_mode, lines):
    num_sources = len(sources) - 1
    lines_to_sort = ""
    for i in range(num_sources):
        line = "{}: {}\n".format(sources[i], occurrences[i])
        lines_to_sort += line

    lines_to_append = sort_lines(lines_to_sort)
    lines_to_append = calculate_percentages(lines_to_append, len(lines) - 1)
    with open(filepath, write_mode, encoding='utf-8') as f:
        f.write(lines_to_append)

def write_to_file_words(words, occurrences, filepath, write_mode, lines):
    num_words = len(words) - 1
    lines_to_sort = ""
    for i in range(num_words):
        line = "{}: {}\n".format(words[i], occurrences[i])
        lines_to_sort += line

    lines_to_append = sort_lines(lines_to_sort)
    total_words = 0
    for line in lines:
        if line == '':
            break

        line_elements = line.split(' | ')
        for line_element in line_elements:
            words_in_line = line_element.split(' ')
            total_words += len(words_in_line)

    lines_to_append = calculate_percentages(lines_to_append, total_words)
    with open(filepath, write_mode, encoding='utf-8') as f:
        f.write(lines_to_append)

def source_count_per_file():
    for file in files:
        filename = file.replace('.txt', '')

```

```

filepath = os.path.join("analysis/source_count", filename + "_source_count.txt")
if os.path.exists(filepath):
    os.remove(filepath)

write_mode = "w" # Append to file
original_path = os.path.join("news", file)
with open(original_path, "r", encoding='utf-8') as f:
    lines_in_file = f.read()

lines = lines_in_file.split("\n")
sources = []
occurences = []
unrelated_lines = 0
#if file == "who_en.txt":
#     for line in lines:
#         if not ("WHO" in line or "World Health Organization" in line):
#             unrelated_lines += 1
#             lines.remove(line)

print("Total number of headlines: " + str(len(lines) - 1))
sources, occurences = get_sources_and_occurences(lines, sources, occurences)
write_to_file(sources, occurences, filepath, write_mode, lines)
#if file == "who_en.txt":
#     with open(filepath, "a", encoding='utf-8') as f:
#         f.write("Removed " + str(unrelated_lines) + " lines unrelated to t

print("Finished analysing " + file)

def source_count_de():
    sources = []
    occurences = []
    total_headlines = 0
    filepath = os.path.join("analysis/source_count", "source_count_de.txt")
    if os.path.exists(filepath):
        os.remove(filepath)

    write_mode = "w" # Append to file
    for file in files_de:
        sources, occurences, total_headlines, lines, x = perform_source_analysis(fi

    print("Total number of headlines: " + str(total_headlines - 1))
    write_to_file(sources, occurences, filepath, write_mode, lines)

```

```

print("Finished analysing German files")

def source_count_en():
    sources = []
    occurrences = []
    total_headlines = 0
    unrelated_lines = 0
    filepath = os.path.join("analysis/source_count", "source_count_en.txt")
    if os.path.exists(filepath):
        os.remove(filepath)

    write_mode = "w" # Append to file
    for file in files_en:
        sources, occurrences, total_headlines, lines, unrelated_lines = perform_source_analysis(file)

    print("Total number of headlines: " + str(total_headlines - 1))
    write_to_file(sources, occurrences, filepath, write_mode, lines)
    with open(filepath, "a", encoding='utf-8') as f:
        f.write("Removed " + str(unrelated_lines) + " lines unrelated to the WHO in")

    print("Finished analysing English files")

def source_count_nl():
    sources = []
    occurrences = []
    total_headlines = 0
    filepath = os.path.join("analysis/source_count", "source_count_nl.txt")
    if os.path.exists(filepath):
        os.remove(filepath)

    write_mode = "w" # Append to file
    for file in files_nl:
        sources, occurrences, total_headlines, lines, x = perform_source_analysis(file)

    print("Total number of headlines: " + str(total_headlines - 1))
    write_to_file(sources, occurrences, filepath, write_mode, lines)

    print("Finished analysing Dutch files")

def source_count():
    source_count_per_file()

```

```

source_count_de()
source_count_en()
source_count_nl()

def date_count():
    for file in files:
        filename = file.replace('.txt', '')
        filepath = os.path.join("analysis/date_count", filename + "_date_count.txt")
        if os.path.exists(filepath):
            os.remove(filepath)

        write_mode = "w" # Append to file
        original_path = os.path.join("news", file)
        with open(original_path, "r", encoding='utf-8') as f:
            lines_in_file = f.read()

        lines = lines_in_file.split("\n")
        dates = []
        occurrences = []
        unrelated_lines = 0
        #if file == "who_en.txt":
        #     for line in lines:
        #         if not ("WHO" in line or "World Health Organization" in line):
        #             unrelated_lines += 1
        #             lines.remove(line)

        for line in lines:
            if line == '':
                break

            line_elements = line.split(' | ')
            time_index = len(line_elements) - 1
            time_elements = line_elements[time_index].split('T')
            date = time_elements[0]
            if date not in dates:
                dates.append(date)
                occurrences.append(1)
            else:
                index = dates.index(date)
                occurrences[index] += 1

        write_to_file(dates, occurrences, filepath, write_mode, lines)
        #if file == "who_en.txt":

```

```

#     with open(filepath, "a", encoding='utf-8') as f:
#         f.write("Removed " + str(unrelated_lines) + " lines unrelated to t

print("Finished analysing dates in " + file)

def word_count():
    for file in files:
        filename = file.replace('.txt', '')
        filepath = os.path.join("analysis/word_count", filename + "_word_count.txt")
        if os.path.exists(filepath):
            os.remove(filepath)

        write_mode = "w" # Append to file
        original_path = os.path.join("news", file)
        with open(original_path, "r", encoding='utf-8') as f:
            lines_in_file = f.read()

        lines = lines_in_file.split("\n")
        words = []
        occurrences = []
        unrelated_lines = 0
        if file == "who_en.txt":
            for line in lines:
                if not ("WHO" in line or "World Health Organization" in line):
                    unrelated_lines += 1
                    lines.remove(line)

        for line in lines:
            if line == '':
                break

            line_elements = line.split(' | ')
            words_used = line_elements[0].split(' ')
            for word in words_used:
                if ":" in word:
                    word.replace(':', '')
                elif "," in word:
                    word.replace(',', '')
                elif "\"\" in word:
                    word.replace('\\"', '')
                elif "\"' in word:
                    word.replace('\\"', '')
                elif "?" in word:

```

```

        word.replace('?', '')
    elif "!" in word:
        word.replace('!', '')

    if word not in words:
        words.append(word)
        occurrences.append(1)
    else:
        index = words.index(word)
        occurrences[index] += 1

write_to_file_words(words, occurrences, filepath, write_mode, lines)
#if file == "who_en.txt":
#    with open(filepath, "a", encoding='utf-8') as f:
#        f.write("Removed " + str(unrelated_lines) + " lines unrelated to t

print("Finished analysing words in " + file)

```

```

source_count()
date_count()
word_count()

```

The script and all files (in the directory "analysis") created by the script can be found in the Github repository (<https://github.com/emysliwietz/newsscrafer>).

## Appendix C

# Python Script for analyzing the topic codes

This is the script used to analyze which topic codes were used for each file:

```
import os
import string
import xlrd
import openpyxl

files_relaxation = [
    'corona_lockerung_de',
    'corona_relaxation_en',
    'corona_versoepeling_nl'
]

files_measures = [
    'corona_maßnahmen_de',
    'corona_measures_en',
    'corona_maatregelen_nl'
]

files_tracking_app = [
    'corona_tracking_app_de',
    'corona_tracking_app_en',
    'corona_tracking_app_nl'
]

files_mask_requirement = [
    'corona_maskenpflicht_de',
    'corona_mask_requirement_en',
    'corona_masker_nl'
]
```



```

]

files_donald_trump = [
    'donald_trump_de',
    'donald_trump_en',
    'donald_trump_nl'
]

files_eu = [
    'eu_de',
    'eu_en',
    'eu_nl'
]

files_italy = [
    'italien_de',
    'italy_en',
    'italie_nl'
]

files_who = [
    'who_de',
    'who_en_with_lines_removed',
    'who_nl'
]

#alphabet = list(string.ascii_uppercase)

if not os.path.exists("analysis"):
    os.mkdir("analysis")

if not os.path.exists("analysis/atlasti_codes"):
    os.mkdir("analysis/atlasti_codes")

def code_scan():
    workbook = xlrd.open_workbook('Bachelor_thesis_CD.xlsx')
    worksheet = workbook.sheet_by_name('CodeDocumentTable')
    for column in range(1, 25):
        file = worksheet.cell(0, column).value.split('\n')[0]
        print(file)
        if file == "who_en_with_lines_removed":
            language = "EN"
        else:

```

```

        language = file.split('_')[len(file.split('_')) - 1].upper()
print(language)
topic = ""
string_to_append = ""
if file in files_relaxation:
    topic = "relaxation"
elif file in files_measures:
    topic = "measures"
elif file in files_tracking_app:
    topic = "tracking_app"
elif file in files_mask_requirement:
    topic = "mask_requirement"
elif file in files_donald_trump:
    topic = "donald_trump"
elif file in files_eu:
    topic = "eu"
elif file in files_italy:
    topic = "italy"
elif file in files_who:
    topic = "who"

filepath = os.path.join("analysis/atlasti_codes", topic + "_codes.txt")
line = "{}: \n".format(language)
string_to_append += line
write_mode = "a" # Append to file
char = ""
codes_to_append = ""
for row in range(1, 297):
    code = worksheet.cell(row, 0).value.split("\n")[0]
    char = code[0:2]
    if int(worksheet.cell(row, column).value) > 0:
        codes_to_append += code

code_list = codes_to_append.split(char)
for code in code_list:
    line = "{}\n".format(str(code))
    string_to_append += line
string_to_append += "\n"
with open(filepath, write_mode, encoding='utf-8') as f:
    f.write(string_to_append)

print("Finished checking codes for " + file)

```

`code_scan()`

The script and all files (in the directory "analysis/atlasti\_codes") created by the script can be found in the Github repository (<https://github.com/emysliwietz/newsscrafer>).