Comparing Belgian and Dutch crisis communication by government officials in the COVID-19 era.

An extensive keyword analysis along the countries' epidemiological curves.

10/01/2021 - Maarten De Winter

Abstract

While governments both in Belgium and the Netherlands have received a variety of critical reviews on its COVID-19 crisis communication, less has been said about the differences between both communication approaches. This paper aims at finding dissimilarities in the topics discussed at certain points in time by means of an extensive keyword analysis along the countries' epidemiological curves.

Table of Contents

Introduction

The purpose of this comparative paper between the Netherlands and Belgium is twofold. The first goal is to identify potential differences in terms of the general crisis communication by government officials about the coronavirus. The second aim is to find the top keywords associated with the different behaviour of the epidemiological curves. More concretely, we would like to determine what words are significantly more used by government officials when COVID-19 cases are for instance rising or slowing down for each country.

Through this analysis, we want to make a clear comparative overview of which country mostly focused on what and additionally at which point in the span of the epidemic. While some ad-hoc corpus analysis for government officials' communication in the COVID-19 era already exists for other countries like Britain and Malaysia[1] or media coverage analyses[2,3], it is now the goal to investigate the government's communication in Belgium and the Netherlands specifically.

Corpus construction

In order to obtain a representative corpus for both countries, the literal texts of the national COVID-19 press conferences and speeches of the relevant ministers were aggregated from their start in early March '20 up until the end of December '20.

Dutch corpus

Given that the literal texts of national press conferences in the Netherlands can be found on the government's online archive, it was rather easy to scrape all the necessary data.[4] As a result, 32 speeches were obtained, mostly by the prime minister Mark Rutte and the minister of health Hugo De Jonge. The names of the people from whom the speeches originate are included in the file names together with the speech date.

While responses to questions by journalists were also available in the archive, they were redacted from the corpus for two reasons. First, this otherwise might have given considerable weight to some words that are not entirely relevant in the context of the epidemic (e.g. 'volgende vraag', or 'laatste vraag'). Second, due to the fact that for the Belgian press conferences such replies were not

[1] Tan, K. H., Woods, P., Azman, H., Abdullah, I. H., Hashim, R. S., Rahim, H. A., Idrus, M. M., Said, N. E. M., Lew, R., & Kosem, I. (2020). COVID-19 insights and linguistic methods. 3L: Language, Linguistics, Literature, 26(2), 1–23. https://doi.org/10.17576/3L-2020-2602-01

[2] Aslam, F., Awan, T. M., Syed, J. H., Kashif, A., & Parveen, M. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. Humanities and Social Sciences Communications, 7(1), 1–9. https://doi.org/10.1057/s41599-020-0523-3

[3] Almazán-ruiz, E , Orrequia-barea, A. "The British Press' Coverage of Coronavirus Threat: A Comparative Analysis Based on Corpus Linguistics". Cankaya University Journal of Humanities and Social Sciences 14 (2020): 1-22 <https://dergipark.org.tr/en/pub/cankujhss/issue/55223/730360>

[4] Rijksoverheid Nederland. (sd.) Documenten. https://www.rijksoverheid.nl/documenten

registered and thus not included in the respective corpus, this would have otherwise resulted in unfair comparisons in the analysis below.

Belgian corpus

Since no similar Belgian government archive contained the written COVID-19 speeches relevant for this study, other sources were consulted. While the competent ministers in government Wilmès and De Croo were also contacted, the most useful answers were offered by the two biggest Flemish broadcasting companies VRT and VTM.

As each program broadcasted on TV is subtitled via Teletext 888, both VRT and VTM made the Dutch texts available in function of this research. I would like to thank Veerle Haverhals (VTM) and Catherine Etcheverry (VRT) for their help. Since both parties however were not able to find a text for the first COVID-19 speech by prime minister Sophie Wilmès, I instead used the BE-alert text of 12 March 2020.[5]

The resulting 22 texts were cleaned since the Teletext metadata has no added value in this study. To illustrate this with an example: below a speech by minister Sophie Wilmès on the 6th of May.

```
SUBTITLE: [9]      TIMEIN: 10:05:54:24      DURATION: 02:17      TIMEOUT:
10:05:57:16
De crisis is enorm,

SUBTITLE: [10]     TIMEIN: 10:05:57:18      DURATION: 02:15      TIMEOUT:
10:06:00:08
net zoals de reikwijdte van onze beslissingen.
```

Methodology

To conduct the keyword analysis, the R package `mclm` is used. Keyword analysis is conventionally used to compare a certain type of text to a larger reference corpus. For the first research question the Dutch corpus will be used as the target corpus and the Belgian corpus as the reference corpus and vice versa. Since `mclm` offers a wide variety of measures that indicate to what extent words are more distinctive for a certain corpus with regards to another, we will choose three: Pointwise mutual information (PMI), the G-squared (G^2) and the Chi-squared test (X^2).

To conduct the time-analysis, i.e. which keywords appear more within a certain time frame, both the Belgian and the Dutch corpus will be divided in a number of 'time folds'. Each fold will be a target corpus compared to the rest of the folds acting as reference corpus. This is illustrated for the Netherlands in the code below. As you can see, we can freely choose the parameter fold size.

---

[5] Sn. (2020). Coronavirus: Fase 2 gehandhaafd, overgang naar de federale fase en bijkomende maatregelen. https://crisiscentrum.be/nl/news/crisisbeheer/coronavirus-fase-2-gehandhaafd-overgang-naar-de-federale-fase-en-bijkomende

```
#size = preferred fold size
#total = total amount corpus files
#amt = amount of folds
#corona_NL = filenames of covid speeches in the Netherlands

size <- 4
total<- nrow(as.data.frame(corona_NL))
amt <- ceiling(total / size) %>% print()

for(i in 0:(amt-1)){
  print("This is timecluster")
  print(i+1)
  print("With the following target speeches")
  print(corona_NL[(1+size*i):(size+size*i)])
  coronaNL_window_flist <-
freqlist(corona_NL[(1+size*i):(size+size*i)])
  coronaNL_reference_flist <-
freqlist(corona_NL[-((1+size*i):(size+size*i))])
  assoc_scores(coronaNL_window_flist, coronaNL_reference_flist)
%>%
    filter(!(type_names(.) %in% stop_list)) %>%
    print(sort_order = "G_signed", n = 50)}
```

To mitigate the effects of stopwords in the analyses in this paper, a stop word list was used that contains Dutch stop words such as 'de' 'een' or 'tussen'.[6]

To plot the distinct keywords along the COVID-19 curves for both countries, I used the public epidemiological data from Sciensano for Belgium[7] and Rijksinstituut voor Volksgezondheid en Milieu (RIVM) for the Netherlands.[8] As the case number proved to be not a good measure for indicating the epidemiological level of the coronavirus due to limited testing in the first wave, I used hospitalisation data as a more representative alternative. For Belgium this is the number of people per day in the hospital due to COVID-19, for the Netherlands it is the number of new hospitalizations per day. The `ggplot2` library was used for the accompanying visualizations.

Results

Keyword analysis Belgium versus the Netherlands

In what follows we will go over the first research question: to what extent Belgian and Dutch speeches differ in terms of content? We will do this through the different measures: PMI, G-Squared and Chi-Squared.

[6] Sn. (2019). Gemoderniseerde stopwoorden lijst. https://eikhart.com/nl/blog/moderne-stopwoorden-lijst
[7] Sciensano. (2021). COVID-19. https://epistat.wiv-isp.be/covid/
[8] RIVM. (2021). Ontwikkeling COVID-19 in grafieken. https://www.rivm.nl/coronavirus-covid-19/grafieken

PMI

Dutch corpus versus Belgian corpus

	Keyword	PMI
1	klachten	0,538435
2	anderhalve	0,53476
3	hugo	0,533694
4	kabinet	0,529221
5	verpleeghuizen	0,529221
6	ggd[9]	0,52661
7	brabant	0,525995
8	besluiten	0,523121
9	jonge	0,523112
10	meter	0,522907
11	drukte	0,522278
12	pakket	0,520402
13	intensive	0,519354
14	ruim	0,518222
15	nederland	0,517916

Belgian corpus versus Dutch corpus

	Keyword	PMI
1	code	1,636256
2	nationale	1,636256
3	mondmasker	1,631514
4	beslist	1,620155
5	bubbel	1,620155
6	telewerk	1,620155
7	veiligheidsraad	1,61742
8	voorbije	1,61742
9	epidemie	1,614314
10	federale	1,614314
11	intensieve	1,614314
12	toepassing	1,614314
13	veiligheidsafstand	1,614314
14	's	1,610756
15	epidemiologische	1,60664

To allow for a more easy comparison, we can distinguish three categories of words based on both their individual meaning and its use given the context of the corpus: words that describe new safety measures or decisions (in green), words that describe the epidemiological situation (in blue), and words that describe governmental instances and political terms (in yellow).

Certain words such as 'klachten', 'jonge' and 'hugo' however can not be subsumed into these categories. The importance of these two last words is due to the fact that the Dutch minister of Health is called Hugo De Jonge. The words 'drukte' and 'toepassing' were classified as a safety measure since they were often used in the context of 'vermijd drukte' and 'maatregelen van toepassing' respectively.

From the tables we can observe that Belgium, compared to the Netherlands, more distinctively communicates about the safety measures and decisions. This may be due to a more strict focus on

---

[9] GGD stands for Gemeentelijke Gezondheidsdienst (English: Municipal Health Service)

these keywords in their texts, or due to generally stronger measures in Belgium, or a combination of both. For example, mouth masks weren't obligatory until the end of 2020 in the Netherlands while in Belgium they were since July.[10] Therefore, the last possibility seems the most reasonable.

In addition, the Netherlands have more attention for COVID-19 symptoms in their speeches with 'klachten' being their top keyword according to the PMI measure. Further distinctive focus mainly goes to epidemiological indicators such as the situation in nursing homes, health instances ('ggd') and intensive care.

G-signed

Dutch corpus versus Belgian corpus				**Belgian corpus versus Dutch corpus**		
	Keyword	**G²**			**Keyword**	**G²**
1	anderhalve	77,394		1	code	56,51834
2	meter	73,14818		2	nationale	56,51834
3	nederland	58,69336		3	regels	54,36695
4	klachten	56,18616		4	mondmasker	47,45263
5	jonge	36,92015		5	activiteiten	41,07454
6	hugo	35,34583		6	één	40,79922
7	geldt	28,87529		7	contacten	38,70787
8	natuurlijk	25,88171		8	betreft	35,76361
9	kabinet	25,75805		9	beslist	33,91879
10	verpleeghuizen	25,75805		10	bubbel	33,91879
11	advies	24,92806		11	telewerk	33,91879
12	echt	24,46789		12	regel	33,63056
13	testen	24,23166		13	protocollen	32,12616
14	thuis	23,12638		14	veiligheidsraad	31,67385
15	ggd	22,09651		15	voorbije	31,67385

Following the same classification as we did in the previous section, we can confirm earlier observations: the main Dutch safety measure is the one and a half meter social distance versus a wide variety of measures in Belgium. However, with the G² - test, COVID-19 testing and staying at home is also included in the table for the Netherlands.

---

[10] Verstraete, A. (2020). Nederland gaat in "gedeeltelijke lockdown": cafés en restaurants moeten dicht, geen verkoop van alcohol na 20 uur. https://www.vrt.be/vrtnws/nl/2020/10/13/nederland-nieuwe-maatregelen/

Additionally, the wide variety of Belgian measures is extended with the words 'één', 'contacten' and 'activiteiten'.

Chi-squared

Dutch corpus versus Belgian corpus

	Keyword	x^2
1	anderhalve	50,24546
2	meter	49,2846
3	nederland	40,06617
4	klachten	35,96099
5	jonge	24,85589
6	geldt	23,92396
7	hugo	23,0295
8	natuurlijk	22,16914
9	echt	21,03273
10	thuis	20,2784
11	houden	19,08529
12	testen	18,91509
13	advies	18,05977
14	kabinet	17,032
15	verpleeghuizen	17,032

Belgian corpus versus Dutch corpus

	Keyword	x^2
1	regels	59,72634
2	code	55,40803
3	nationale	55,40803
4	mondmasker	4,674174
5	activiteiten	45,17746
6	één	44,9417
7	contacten	42,68099
8	betreft	39,41654
9	regel	36,56616
10	leerlingen	34,69719
11	moeilijk	34,06398
12	beslist	33,75036
13	bubbel	33,75036
14	telewerk	33,75036
15	protocollen	33,73344

Once again using the same encoding of keywords, generally the main observations in the previous sections still hold. For the Netherlands, we observe that with this test less weight is given to epidemiological indicators and more noisy keywords such as 'geldt', 'natuurlijk' and 'echt' come into play. The Dutch social distancing importance clearly remains however.

While the keyword 'code' was the top Belgian keyword for PMI and G^2, it is the second top keyword for X^2. Belgium seems to use this word very distinctively, compared to the Netherlands across, all of the three measures. Since the code stands for a color code indicating the epidemiological level, this says a lot about how the government gives priority to trying to communicate concisely about it.

Identifying top keywords for time clusters along epidemiological curves

In this section we will analyze at what point in time certain topics were focused on more than others for both of the countries.

Since the number of speeches by the Netherlands is not equal to the number of speeches by Belgium, and the speeches are distributed over different dates, we need to choose the fold size wisely for each country. As a result, we obtained five folds for the Netherlands with seven texts in the first four time folds and six texts in the last time fold. For Belgium we obtained four folds with six texts in the first three folds and four texts in the last time fold.

Each set of keywords belonging to a time fold represents the top 10 distinctive keywords for each time interval compared to the rest of the time considered from March until the end of December 2020. The top 10 keywords belonging to a certain time fold are plotted in a descending G^2 order on the median date between the first and the last speech of the considered fold.

The Netherlands
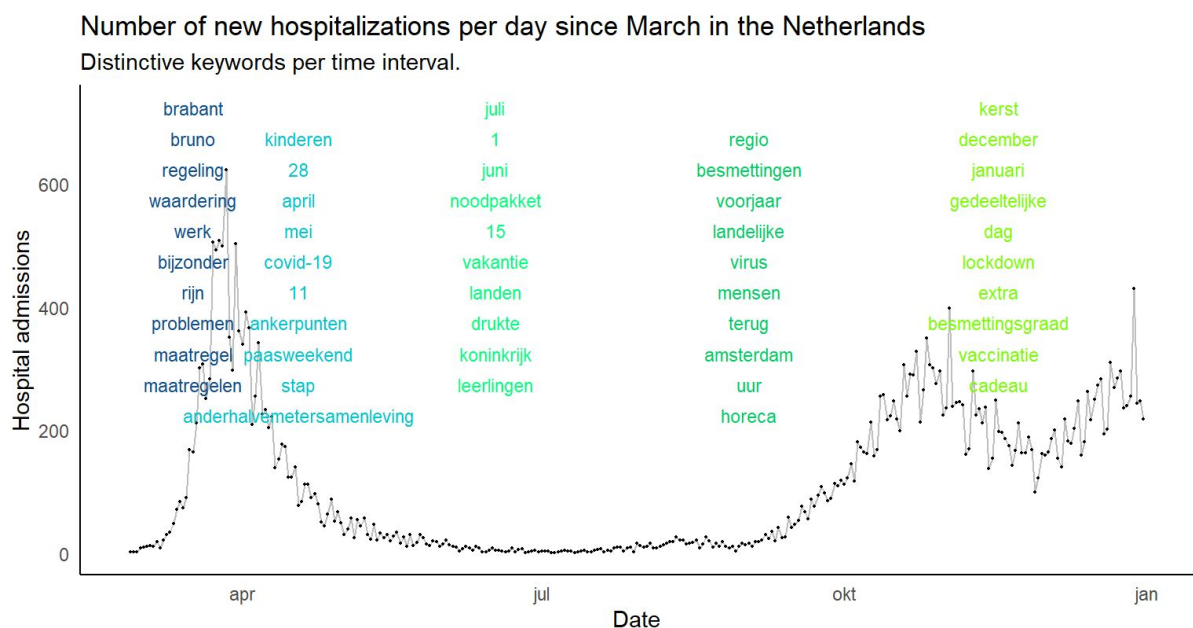
First of all we can see that each timefold fits the behavioural changes of the curve well.

While the data is somewhat noisy due to focus on certain locations 'Rijn' and 'Brabant', we can observe a general response to the spike in cases with both safety measures and appreciation towards care providers and the public in general in the first time interval ranging from 9 March to 25 March.

For the second time interval ranging from 31 March to 6 May along with a decrease in COVID-19 cases, the communication focus is more future driven than the reference corpus: what about the

kids and schools, when will the measures be relaxed, how does this one and a half meter society work, Easter break…

The third interval ranging from 13 May to 22 July is heavily characterized by its summer vacation elements and consequent plans. The public was trusted to be cautious with busy places ('drukte') during their vacation ('vakantie'), and was able to visit certain countries ('landen') after the borders reopened. Such communication was clearly justified by the epidemiological level.

Due to the, firstly local, rise in COVID-19 cases in the fourth interval ranging from 27 July to 13 October, the government acts by making people aware that the virus still is present and cases may increase further due to travelers combing back ('terug'). Also it decides to limit the opening hours of the hospitality industry: restaurants and bars ('uur', 'horeca').

The last interval ranging from 27 October to 8 December shows a story of good and bad news. The bad news is that due to an increasing infection rate a partial lockdown ('gedeeltelijke', 'lockdown') is initiated. However, the good news that is brought is that vaccination ('vaccinatie') can start in january. One can consider this news a Christmas present ('kerst', 'cadeau').

From this graph interpretation we can thus conclude that the top keywords for Dutch government communication about COVID-19 is obviously driven by the epidemiological situation, but also by the specific time context in which a speech is brought.

Belgium

Number of hospital beds occupied since March in Belgium
Distinctive keywords per time interval.
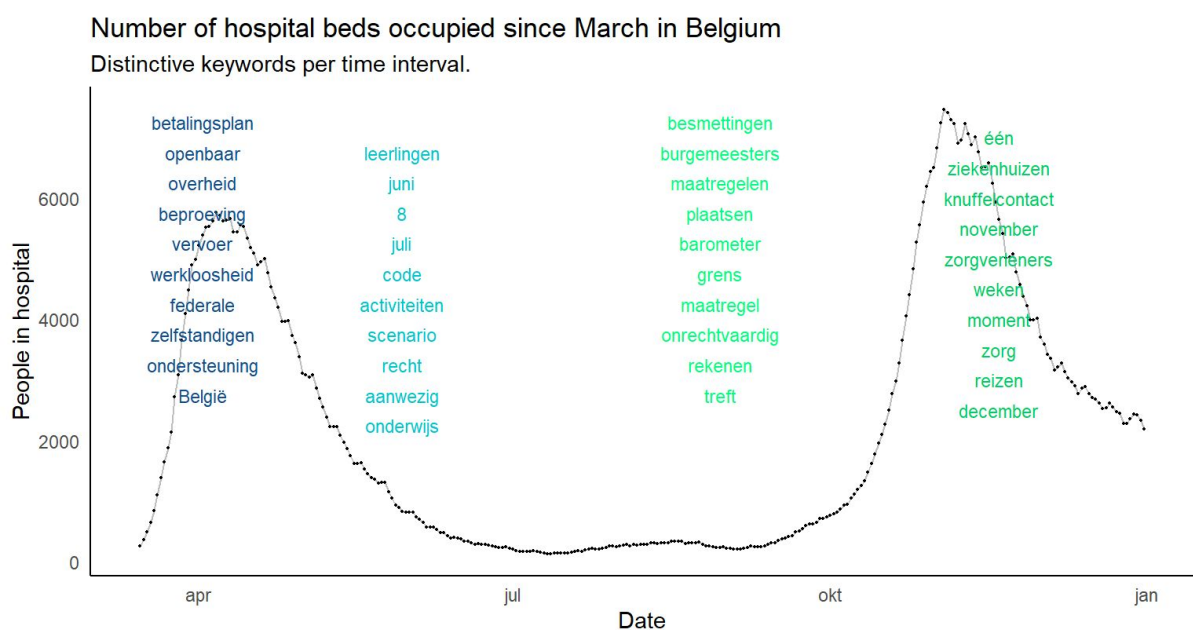
For Belgium, the first time interval represents the period when COVID-19 hit Belgium. From 12 March to 24 April, the time clustered top keywords indicate that the government was primarily

focused on the economical impact of its choice for a national lockdown with keywords such as 'betalingsplan', 'ondersteuning', 'werkloosheid' and 'zelfstandigen'.

When case numbers decreased in the second interval ranging from 6 May to 24 June, a similar 'looking ahead' pattern as in the Netherlands can be observed. Topics such as education ('onderwijs'), events ('activiteiten'), relaxations of the measures ('8', 'juni') and potential future scenarios ('scenario') were discussed. This is also the period when the colour code discussed in the chi-squared section was introduced.

The third time window ranging from 15 July to 16 October was characterized by local spikes of infections first[11] and later cases went up nation-wide resulting in the beginning of the infamous second wave. Due to the local spikes words like 'plaatsen', 'burgemeester' and 'besmettingen' appear. Interestingly, next to the prominent use of the word 'code' in the previous time interval, the word 'barometer' is now also introduced as an indicator.

Lastly, we consider the fourth interval ranging from 23 October to 18 December. Since this interval considers the period of the peak of the second and beyond, it comes as no surprise to find the measure 'één knuffelcontact' in the list. Furthermore a communication focus goes to the care providers ('zorgverleners', 'zorg'). More attention also goes to traveling.

Next to the general observation we made for the Dutch graph in the previous section which also holds for Belgium if we choose five time folds, we can observe a shift in the communication style of Belgium. The speeches during the first wave were more economically focused whereas in the second wave more attention went to clear communication about the safety measures and appreciation for the caregivers. This might also be attributable to the change of governments.[12]

Conclusion

When comparing the Dutch and Belgian COVID-19 government communication corpora, we can make several observations. Firstly, the Netherlands focused more on symptoms with the keyword 'klachten' appearing in the top 4 for all three chosen distinction measures. Secondly, the country also appears to give the social distancing safety measure of one and a half meter a central place in their speeches. Lastly they refer more to healthcare instances such as nursing homes, intensive care and general practitioners to give an insight in the epidemiological situation.

When we look at Belgium versus the Netherlands on the other hand, we observe that a wider variety of safety measures is included in the top distinctive Belgian keywords. This may indicate a stronger Belgian COVID-policy and/or a more repetitive use of these keywords in different

---

[11] Keersmaekers, L. (2020). Eerste avond met avondklok in Antwerpen.
https://www.vrt.be/vrtnws/nl/2020/07/30/eerste-avond-met-avondklok-in-antwerpen/
[12] Decre, H., Willems, F. (2020). https://www.vrt.be/vrtnws/nl/2020/09/30/bevoegdheden-regering-de-croo/

speeches. Furthermore we see that the Belgian government communicates the epidemiological level more concisely by the use of colour codes.

A time-analysis for the Dutch keywords in certain time intervals learns us that the top keywords for Dutch government communication about COVID-19 driven by both the epidemiological situation and the specific time context in which a speech is brought. Furthermore, Belgium appears to show a shift in terms of keywords. We can observe a less economical perspective on the facts in the second wave compared to the first.

This research only scratches the surface of differences between the two communication approaches. For future research, it could be interesting to further investigate collocates and cooccurrences of specific words such as 'piek', 'gevaar', 'lockdown', 'symptomen', 'klachten'. Another approach is to investigate certain topics and the presence of belonging keywords. For example: Which country focuses the most on education? Which country wants to solve this on a european level? Additionally, this epidemic reveals that in the Netherlands the words 'mondkapje' is primarily used for describing a mouth mask while in Belgium 'mondmasker' is mostly used. This could be studied from a variety perspective.

References

- Almazán-ruiz, E , Orrequia-barea, A. "The British Press' Coverage of Coronavirus Threat: A Comparative Analysis Based on Corpus Linguistics". Cankaya University Journal of Humanities and Social Sciences 14 (2020): 1-22. https://dergipark.org.tr/en/pub/cankujhss/issue/55223/730360
- Aslam, F., Awan, T. M., Syed, J. H., Kashif, A., & Parveen, M. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. Humanities and Social Sciences Communications, 7(1), 1–9. https://doi.org/10.1057/s41599-020-0523-3
- Decre, H., Willems, F. (2020). https://www.vrt.be/vrtnws/nl/2020/09/30/bevoegdheden-regering-de-croo/
- Etcheverry, C. "Re: Federale toespraken COVID-19." Message to Maarten De Winter. 30 December 2020. E-mail.
- Haverhals, V. "Re: Federale toespraken COVID-19." Message to Maarten De Winter. 24 December 2020. E-mail.
- Keersmaekers, L. (2020). Eerste avond met avondklok in Antwerpen. https://www.vrt.be/vrtnws/nl/2020/07/30/eerste-avond-met-avondklok-in-antwerpen/
- Tan, K. H., Woods, P., Azman, H., Abdullah, I. H., Hashim, R. S., Rahim, H. A., Idrus, M. M., Said, N. E. M., Lew, R., & Kosem, I. (2020). COVID-19 insights and linguistic methods. 3L: Language, Linguistics, Literature, 26(2), 1–23. https://doi.org/10.17576/3L-2020-2602-01
- Rijksoverheid Nederland. (sd.) Documenten. https://www.rijksoverheid.nl/documenten
- RIVM. (2021). Ontwikkeling COVID-19 in grafieken. https://www.rivm.nl/coronavirus-covid-19/grafieken
- Sn. (2020). Coronavirus: Fase 2 gehandhaafd, overgang naar de federale fase en bijkomende maatregelen. https://crisiscentrum.be/nl/news/crisisbeheer/coronavirus-fase-2-gehandhaafd-overgang-naar-de-federale-fase-en-bijkomende
- Sn. (2019). Gemoderniseerde stopwoorden lijst. https://eikhart.com/nl/blog/moderne-stopwoorden-lijst
- Sciensano. (2021). COVID-19. https://epistat.wiv-isp.be/covid/
- Verstraete, A. (2020). Nederland gaat in "gedeeltelijke lockdown": cafés en restaurants moeten dicht, geen verkoop van alcohol na 20 uur. https://www.vrt.be/vrtnws/nl/2020/10/13/nederland-nieuwe-maatregelen/