New Year's Evening speeches given by the Queen of Denmark, years 2018-2019 and 2020-2021

GOD BLESS DENMARK (and Sentiment Analysis) / Sentiment Analysis of New Year's Eve speeches held by the Queen of Denmark from 2018 and 2019, and 2020 and 2021 / RStudio

Anne Louise Bank Hornung

Keywords: *sentiment analysis; queen of Denmark; NYE speech; RStudio*

Part A

1. Introduction

When the coronavirus pandemic broke out in March 2020 it maintained its tight grip on Denmark and the world throughout 2020 and 2021. It was impossible to avoid hearing about it, and it was often not spoken of in very positive terms. As coronavirus intruded the lives of all Danes, it was inevitable for the Queen of Denmark, HM Queen Margrethe II, to address it, which she did in her New Year's Evening (NYE) addresses given in both 2020 and 2021. Queen Margrethe II NYE speeches are usually motivational and positive speeches, but by doing a sentiment analysis, I hypothesize that her NYE speeches from 2020 and 2021 will have more negatively categorized words than her NYE speeches from 2018 and 2019. Furthermore, by attempting to do a sentiment analysis on both her original speeches in Danish, and the version translated into English, it will be evident that words can go lost in translation, resulting in misinterpreted and/or odd words in the sentiment analysis.

2. Problems and background / Context

Queen Margrethe II's NYE speeches have huge cultural and historical meaning for Danes. According to Carsten Madsen, many Danes see Queen Margrethe II as in epitome of 'Danishness', and her NYE speeches can be seen as her staging of her perception of Danish national identity.[1] Furthermore, even though around 95% of Danes watch the NYE speech every year,[2] Madsen notices how little the literary structure, rhetoric, and political and legal

[1] Madsen, Carsten. 2017, "Magtens repræsentation Af Danskhed. Konstitueringen Af Nationalidentitet I Dronning Margrethes nytårstaler". *Passage - Tidsskrift for Litteratur Og Kritik* 31 (76). https://doi.org/10.7146/pas.v31i76.25233, 129

[2] Dehn-Nielsen, Henning. 2002, *Dronning Margrethes Nytårstaler. Kommenteret og sat i historisk relief*, Kgs. Lyngby: Holkenfeldt 3. 10, qtd. in Madsen 130

status of the speech has been analyzed throughout the years.[3] By doing a sentiment analysis on the NYE speeches it will contribute to a further insight into the attitudes, opinions, and feelings of the current state of Denmark, which are expressed through the country's monarch. As Denmark anno 2020 and 2021 was largely characterized by the coronavirus, I hypothesize that by doing a sentiment analysis, it will illustrate that the speeches' sentiment is different from the years before.

3. Software Framework

I wrote the code for this project on my 8-year old MacBook Air, 4 Gb RAM, which runs on macOS Mojave (version 10.14.2) operating system. I worked in the desktop version of R (4.2.2) and RStudio (4.2.2). To work in RStudio, it is first necessary to download R, as RStudio will otherwise not work. I followed the instructions given on https://datacarpentry.org/r-socialsci/setup.html. Here it also states that "[i]t is also a good idea to install XQuartz (needed by some packages)",[4] which was necessary for my computer. XQuartz runs on version 2.8.2 on my computer.
When opening RStudio, it consistently gives me the following warning message in the console: "R graphics engine version 15 is not supported by this version of RStudio. The Plots tab will be disabled until a newer version of RStudio is installed." I have not found this to be an issue.

4. Data acquisition and processing

All my data, Rmarkdowns etc. can be downloaded through GitHub here: https://github.com/Digital-Methods-HASS/au601065_hornung_anne
All my data has been extracted from kongehuset.dk, where the all the Queen's NYE speeches are available in both English and Danish. This made it easy for me to copy and paste the individual speeches from the webpage into a Word document and then convert them into four different PDF-files. One PDF-file contained the translated English NYE speeches from 2018[5] and 2019[6] and another one the translated English NYE speeches from 2020[7] and 2021[8]. A third PDF-file one contained the Danish NYE speeches from 2018[9] and 2019,[10] and the last

[3] Madsen. "Magtens repræsentation", 130
[4] Juanfung et al., "datacarpentry / r-socialsci", https://datacarpentry.org/r-socialsci/setup.html
[5] Kongehuset.dk. 2018, "Her Majesty The Queen's New Year Address 2018"
[6] Kongehuset.dk. 2019, "New Year's Address 2019"
[7] Kongehuset.dk. 2020, "New Year's Address 2020"
[8] Kongehuset.dk. 2021, "Read HM The Queen's New Year's Address 2021"
[9] Kongehuset.dk. 2018, "Læs H.M. Dronningens nytårstale 2018"
[10] Kongehuset.dk. 2019, "Læs H.M. Dronningens nytårstale 2019"

one contained the Danish NYE speeches from 2020[11] and 2021.[12] I chose to make four PDF-files with each two NYE speeches in them, because I figured it would be the best way to compare speeches, if they are of roughly similar length.
Since Queen Margrethe II's NYE speeches are short, cohesive, and easily located online, it was not necessary to "clean" them through another software, and the only data preparation necessary was copy-pasting them into a Word document and converting to a PDF-file (as explained above). I have attached the four PDF-files to the "Exam" -> "Data"-folder in my GitHub repository.

5. Empirical results

The main elements and codes followed the structure of the SentimentAnalysis[13] tutorial on GitHub, which provided an easily executed sentiment analysis on the translated English NYE addresses. For the sentida package, I followed the tutorial[14] provided by the creators on GitHub, which, after some trial and error, worked after its intent.
The results of my investigation are mainly comparable when looking at the translated English NYE speeches, as I was not capable of running the same sentiment analysis (afinn, nrc and bing) packages to the Danish speeches.
All the word clouds from all four projects can however be directly compared (as done briefly below), since they all followed the same codes and represent the top 100 most frequent words said in the speeches.

[11] Kongehuset.dk. 2020, "Læs H.M. Dronningens nytårstale 2020"
[12] Kongehuset.dk. 2021, "Læs H.M. Dronningens nytårstale 2021"
[13] adivea, "SentimentAnalysis / W11Rmd", https://github.com/Digital-Methods-HASS/SentimentAnalysis/blob/main/W11.Rmd
[14] Guscode, "Sentida", https://github.com/Guscode/Sentida

Figure 1 Word cloud of top 100 most frequent words in NYE_d_18_19

Figure 2 Word cloud of top 100 most frequent words in NYE_d_20_21

Looking at the word clouds in figure 1 and 2, it is clear how the words "år" (year) and "danmark" (Denmark) dominate both speeches in frequency of words, which is not surprising

considering it's NYE speeches given in Denmark. However, in figure 4 it is evident that "coronaen" (the corona) and "krisen" (the crisis) are also words dominating the top 100, indicating what historical context Denmark was dealing with at the time.

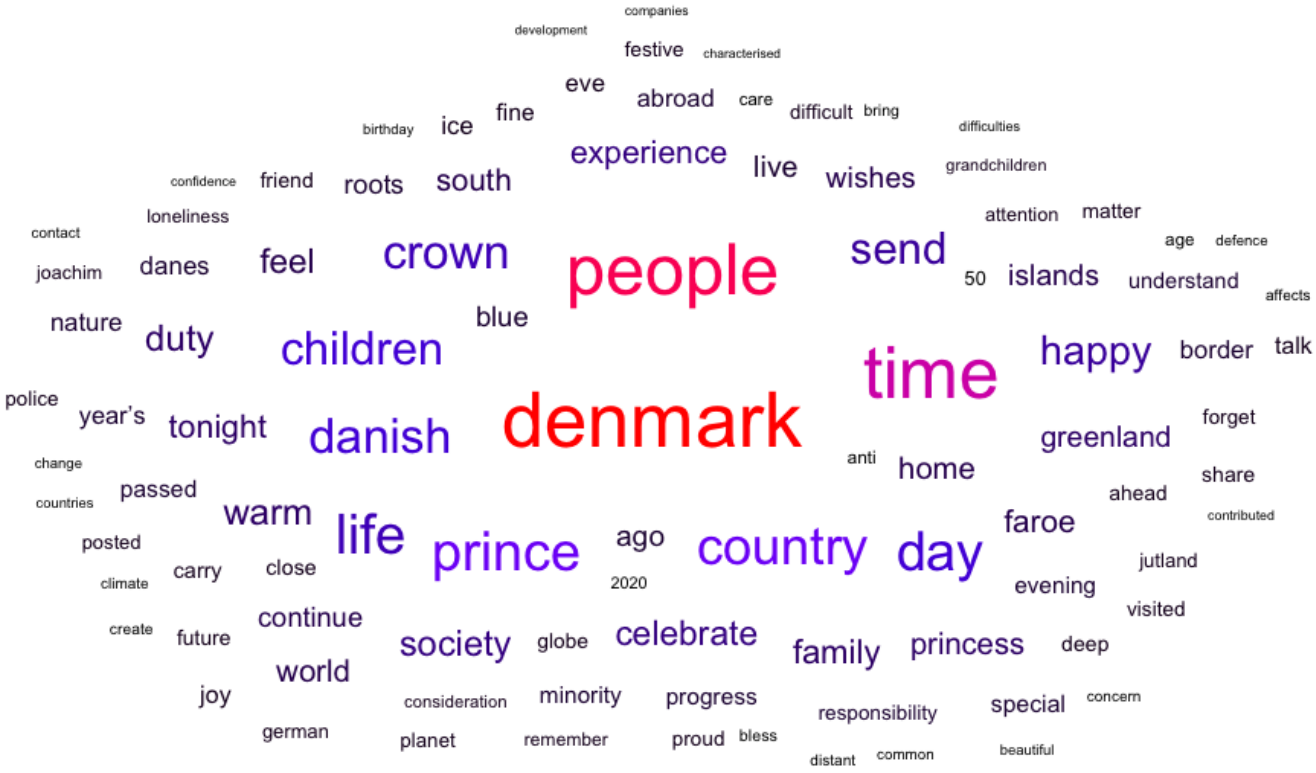

Figure 3 Word cloud of top 100 most frequent words in NYE_q_18_19

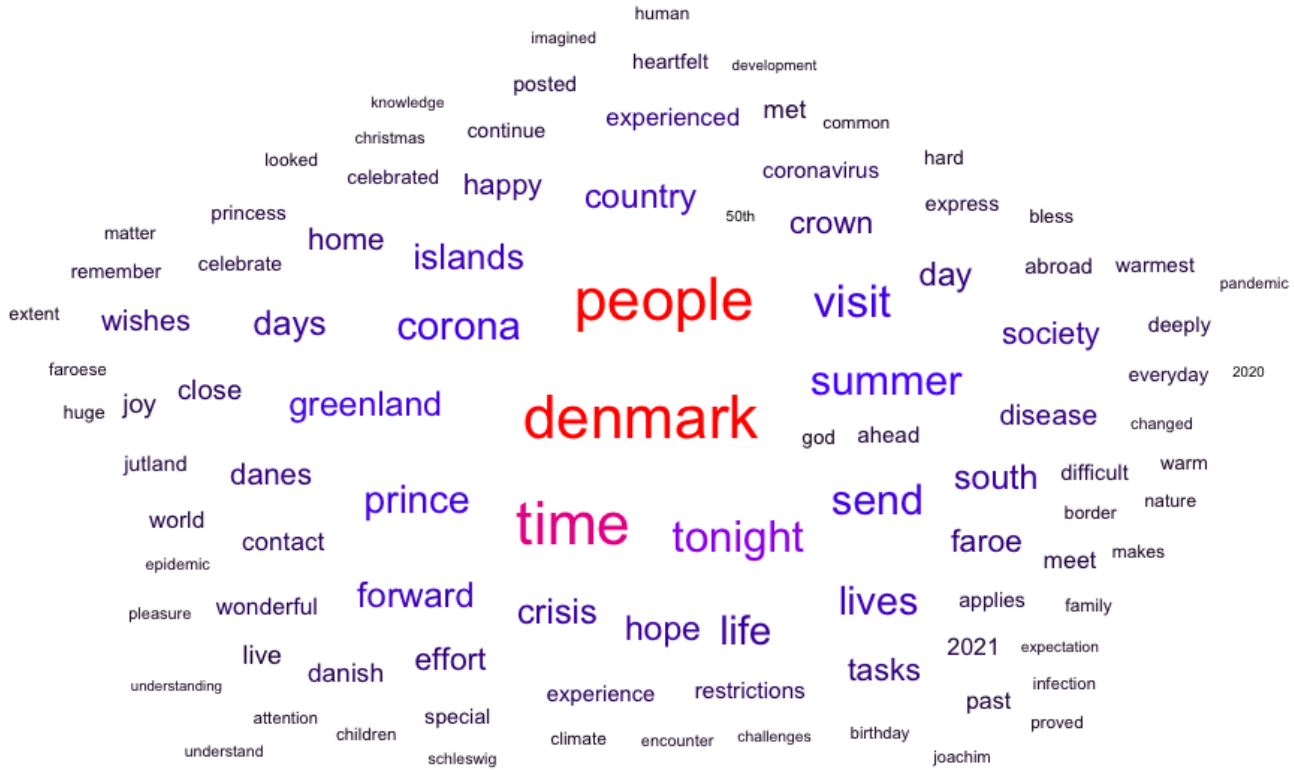

Figure 4 Word cloud of top 100 most frequent words NYE_q_e_20_21

Looking at the word clouds in figure 3 and 4, it is evident that "Denmark" and "time" dominate both word clouds; these words can be directly translated to the Danish words "Danmark" and "år", as are the most frequent words in figure 1 and 2. However, the word "people" is dominating both English word clouds, and when comparing to figure 2 and 1, such a similarly frequent word in Danish is nowhere to be located. This can be since many Danish words roughly translates into "people", indicating a loss in translation.

	value	n
1	-3	10
2	-2	19
3	-1	29
4	1	37
5	2	32
6	3	33
7	4	6

Figure 5 NYE_e_20_21_afinn_hist

	value	n
1	-3	3
2	-2	20
3	-1	21
4	1	37
5	2	44
6	3	27
7	4	2

Figure 6 NYE_e_18_19_afinn_hist

The figures 5 and 6 above show the data that underwent a sentiment analysis using the afinn-package, and these are the NYE speeches translated into English. The findings support my hypothesis that the NYE speeches given by Queen Margrethe II in 2020 and 2021, included more negative polarity words than the ones she gave in 2018 and 2019, at least when translated into English. In 2020+2021, 10 words ranked -3 and 29 words ranked -1 on the afinn sentiment scale, whereas only 3 words ranked -3 in 2018+2019, and 21 words ranked -1 on the sentiment scale.

I struggled to work out how to remove the stop words in the Danish speeches, but after trying various codes, I located Max Odsbjerg Pedersen's article "Åbne data – Ringe OCR kvalitet",[15] which helped me remove the Danish stop words easily.
The sentinda package turned out to be somewhat useful for my findings, but I was not able to comprise the exact same analyses using the sentida package as using the afinn, nrc and bing packages. Sentida is a package that can be installed into RStudio, and "is a lexicon comprised of the existing Danish sentiment lexicon AFINN and a list of new words".[16]

6. Critical evaluation

I originally intended to compare the Queen of Denmark's NYE speeches with the Prime Minister (pm) of Denmark's NYE speeches, all translated into English. But while looking into the PM's speeches, it became evident that comparing a largely political speech with a speech where the speaker is politically neutral, was not a feasible comparison. Instead, it seemed more feasible to compare the original Danish speeches to translated English speeches and explore the effect coronavirus-related words had on the speeches, as well as what may go lost in translation during a sentiment analysis.

The data I used followed the FAIR guiding principles for data as outlined by Wilkinson et al., who argue that data must be Findable, Accessible, Interoperable, and Reusable.[17] The findability of the data I used in my project is present in the sense that the "metadata clearly and explicitly include[s] the identifier of the data it describes",[18] since the date, heading and

[15] Pedersen, Max Odsbjerg. 2020, "Åbne data – Ringe OCR kvalitet", https://library.au.dk/fileadmin/www.bibliotek.au.dk/LibLab/Eksempler/old_newspapers_1849_korr.html#Analyse
[16] Lauridsen, Gustav Aarup et al. 2019, "SENTIDA: A New Tool for Sentiment Analysis in Danish", Vol. 4, no. 1, 40
[17] Wilkinson, Mark D. et al. 2016, "Comment: The FAIR Guiding Principles for scientific data management and stewardship", DOI: 10.1038/sdata.2016.18. 1
[18] Ibid, 4

subheading identifies the data it describes in each NYE speech. The data is also accessible in the sense that the speeches can be easily retrieved by their identifiers, as well as open and free for the public.[19] The speeches that were retrieved in English, are also interoperable, as English is a "broadly applicable language".[20] The original Danish speeches may qualify as less interoperable, though. Lastly, the data is also reusable.

Using the nrc package, it is evident that there are some misinterpreted words presented on the sentiment scale: "bear" presented in both the "anger" and "fear" category on the sentiment scale, as shown in figure 7 and in NYE_q_e_18_19. But when locating the word "bear" in the actual speech, it is presented in the sentence "we who are adults should also bear it in mind",[21] which implies that there are no correlations to the animal bear that might entail anger and fear. This is an example of the many words that go lost in translation when translating NYE speeches from Danish to English, and of the sentiment analysis' many limitations.

[19] Ibid, 4
[20] Ibid, 4
[21] Kongehuset.dk. "New Year Address 2018"

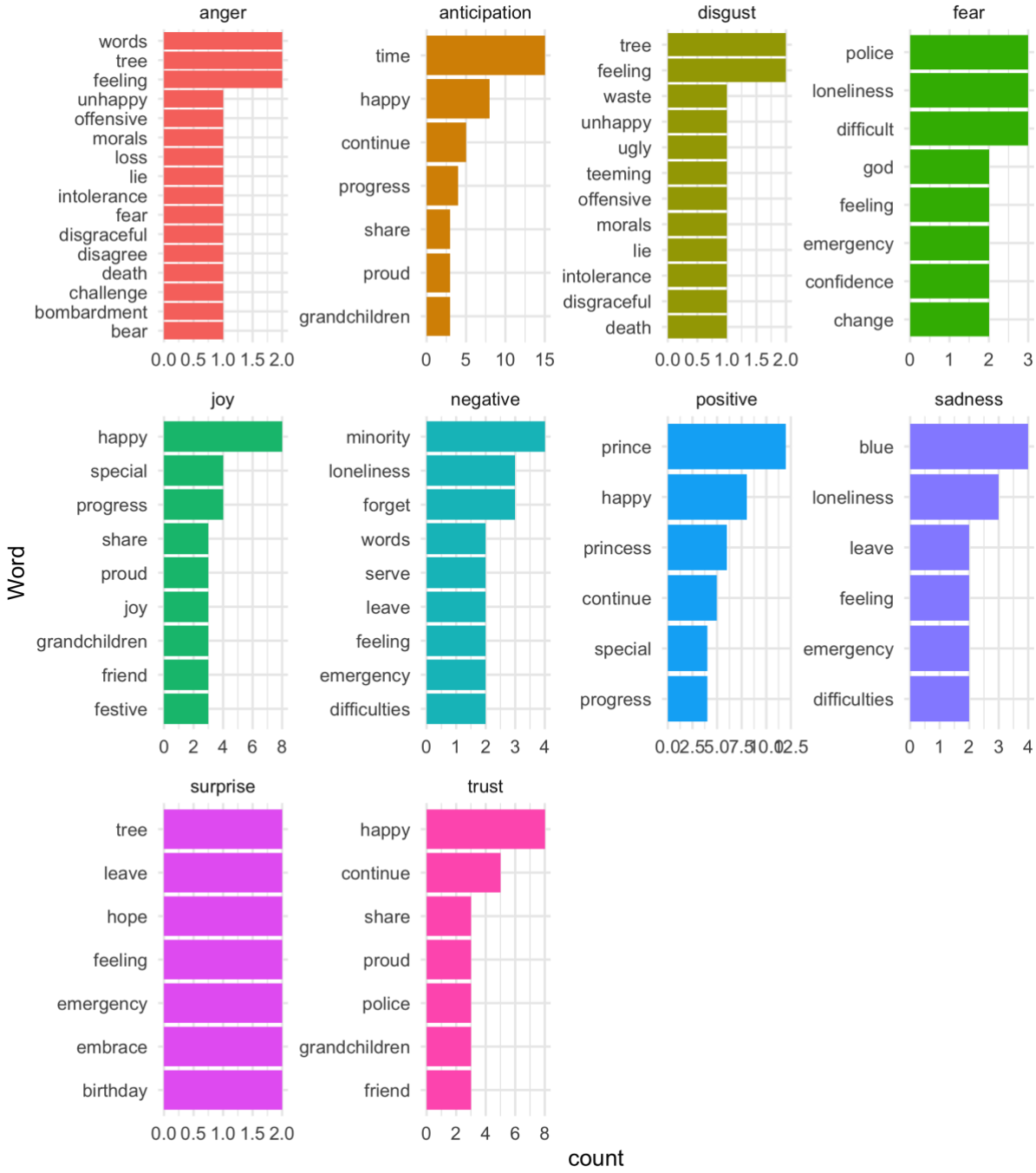

Figure 7 Sentiment analysis done on NYE_q_e_18_19

Lastly, I want to address my findings and exploration using the sentida package. I was able to use the sentida package to calculate the mean of words in the Danish speeches, which meant I could compare the mean findings to the afinn package used in the English translated speeches, as the sentida package "is a polarity strength lexicon similar to AFINN", meaning that they both span from -5 meaning "very strong negative emotion and a score of 5 mean[ing] very strong positive emotion".[22]

[22] Lauridsen, Gustav Aarup et al., "SENTIDA", 41

The afinn package was used to calculate the mean of the words in NYE_e_18_19 and NYE_e_20_21 that had underwent a stop word removal (see "# Bind words in `NYEs_q_18_19_stop` to `afinn` lexicon" in NYE_e_18_19.rmd and "# Bind words in `NYE_q_e_20_21_stop` to `afinn` lexicon" in NYE_e_20_21.rmd).
To make sure it was the same words that underwent a mean calculation in NYE_d_18_19 and NYE_d_20_21, I copied the English stop-word-removed words from NYEs_q_18_19_stop and NYE_q_e_20_21_stop data frames into a Word document, translated and sorted them manually into Danish (Word document provided in GitHub).
Hereafter, I pasted the Danish words into NYE_d_18_19 and NYE_d_20_21, respectively, where I ran the sentida codes (see the chunk "{r sentida speeches 2020+2021}" in NYE_d_20_21.rmd and the chunk "{r sentida speeches 2018+2019}" in NYE_d_18_19.rmd). This provided the following results: in NYE_d_18_19 it shows that the mean of the words is 0,96, and in NYE_d_20_21 shows that the mean is 1,02. This means that both speeches rank around "very weak positive emotion", according to the sentida coding scheme.[23] Surprisingly, the NYE_d_20_21 ranked (a little) higher than NYE_d_18_19; I hypothesized the opposite.
In the NYE_e_18_19 the mean of the sentiment is calculated to be 0,93 and the mean of NYE_e_20_21 is calculated to be 0,76. This indicates that in English, the Queen's speeches were a little less positive in 2020+2021, which is due to the increase in negative polarity words caused by coronavirus-related words.

I am somewhat satisfied with the final project, which involved analyzing the sentiment of speeches in Danish and English. I found that the sentida package was not as effective as the afinn, nrc and bing packages at handling larger amounts of text (I was at least not able to figure it out). I also wish I had been able to apply the same sentiment analysis to both the English and Danish speeches to better compare them and address the issue of what is possibly lost in translation.
Working with sentida was the most frustrating part of my project, but I learned from my mistakes, and I will try to find simpler solutions in the future.

7. Conclusion

The coronavirus hugely impacted all our lives in a mostly negative way, which explained why the Queen of Denmark's NYE speeches were analyzed as having more negative sentiment (at least when the mean was calculated in the translated English speeches). It also became

[23] Ibid, 41

evident that words can go lost in translation, both in the process of translating speeches as seen in the word cloud figures, but also when binning emotions with the nrc package that misinterpreted some words. Although my project mainly confirmed my hypothesis, the speeches from 2020+2021 were only marginally less positive, which is likely due to the nature of the Queen's NYE speeches that are supposed to be motivational and positive, as leading the Danes into the new year.

Part B: Required Metadata

Table 1 – Software metadata

Nr	Software metadata description	
S1	Current software version	RStudio version 4.2.2 (R version 4.2.2)
S2	Permanent link to Github repository where you put your script or R project	https://github.com/Digital-Methods-HASS/au601065_hornung_anne
S3	Legal Software License	Kongehuset.dk (Copyright)
S4	Computing platform / Operating System	MacBook Air Early 2015 which runs on macOS Mojave version 10.14.2.
S5	Installation requirements & dependencies for software not used in class	Packages required: sentida pdftools textdata tidytext ggwordcloud get_sentiments(lexicon = "nrc") get_sentiments(lexicon = "afinn") get_sentiments(lexicon = "bing")
S6	If available Link to software documentation for special software	https://github.com/Guscode/Sentida
S6	Support email for questions	annebankh@gmail.com

Table 2 – Data metadata

Nr	Metadata description	
D1	NYE_q_18_19.pdf	A PDF-file that contains the Queen of Denmark's New Year's Address translated into English from 2018 and 2019. Extracted from kongehuset.dk on 8 December 2022
D2	NYE_d_18_19.pdf	A PDF-file that contains the Queen of Denmark's New Year's Address in Danish from 2018 and 2019. Extracted from kongehuset.dk on 5 January 2023
D3	NYE_q_e_20_21.pdf	A PDF-file that contains the Queen of Denmark's New Year's Address translated into English from 2020 and 2021. Extracted from kongehuset.dk on 4 January 2023
D4	NYE_d_20_21.pdf	A PDF-file that contains the Queen of Denmark's New Year's Address in Danish from 2020 and 2021. Extracted from kongehuset.dk on 5 January 2023

Bibliography

adivea, “SentimentAnalysis / W11Rmd”, https://github.com/Digital-Methods-HASS/SentimentAnalysis/blob/main/W11.Rmd

Dehn-Nielsen, Henning. 2002, *Dronning Margrethes Nytårstaler. Kommenteret og sat i historisk relief*, Kgs. Lyngby: Holkenfeldt 3.

Guscode, et al., “Sentida”, https://github.com/Guscode/Sentida

Juanfung et al., ”datacarpentry / r-socialsci”, https://datacarpentry.org/r-socialsci/setup.html

Kongehuset.dk. 2020, “Læs H.M. Dronningens nytårstale 2020”

---. 2021, “Læs H.M. Dronningens nytårstale 2021”

---. 2018, “Her Majesty The Queen’s New Year Address 2018”

---. 2019, “New Year’s Address 2019”

---. 2020, “New Year’s Address 2020”

---. 2021, “Read HM The Queen's New Year’s Address 2021”

---. 2018, “Læs H.M. Dronningens nytårstale 2018”

---. 2019, “Læs H.M. Dronningens nytårstale 2019”

Lauridsen, Gustav Aarup et al. 2019, ”SENTIDA: A New Tool for Sentiment Analysis in Danish”, Vol. 4, no. 1, 40

Madsen, Carsten. 2017. “Magtens repræsentation Af Danskhed. Konstitueringen Af Nationalidentitet I Dronning Margrethes nytårstaler”. *Passage - Tidsskrift for Litteratur Og Kritik* 31 (76). https://doi.org/10.7146/pas.v31i76.25233.

Pedersen, Max Odsbjerg. 2020, “Åbne data – Ringe OCR kvalitet”, https://library.au.dk/fileadmin/www.bibliotek.au.dk/LibLab/Eksempler/old_newspapers_1849_korr.html#Analyse

Wilkinson, Mark D. et al. 2016, ”Comment: The FAIR Guiding Principles for scientific data management and stewardship”, DOI: 10.1038/sdata.2016.18.