DHQ: Digital Humanities Quarterly
Preview
2022
Volume 16 Number 1

A Text Network Analysis of Discursive Changes in German, Austrian and Swiss New Year's Speeches 2000-2021

Kimmo Elo <kimmo_dot_elo_at_utu_dot_fi>, University of Turku

Abstract

New Year's speeches held by leading politicians–mostly prime ministers or presidents–have a long and firm tradition in Europe and have become an institution instead of being a crowing event for conciliatory efforts. From the perspective of political communication, New Year's speeches fulfil a triple function in the intersection of the past, the present and the future. First, they summarise the past year from the perspective of the political leadership and, hence, recall, reconstruct and remind the most important events of the year. Second, New Year's speeches describe the present and, thus, can be understood and analysed as reality constructions, as windows to the current state of affairs. And third, New Year's speeches serve as road maps to the future, into the new year. In this sense, a New Year's speeches summarises the most important future challenges, expectations and opportunities.

This article stems from the assumption that a nonlinear analysis of textual data based on network analysis could provide us with new ontological understanding about structural coherence and holes within a document corpora. It adopts a different viewpoint to discourse analysis based on social network analysis. The paper introduces a nonlinear way to analyse texts as networks in order to visualise and analyse how concepts are connected and to explore structural closeness and holes within a corpus of unstructured textual documents.

The results indicate a discursive turn in Germany, Austria and Switzerland after the breakout of the global financial crisis in 2008. But at the same time, the results evidence similarities across the three countries how the crisis and its impact was framed to discourses. In all countries, the use of concepts related to crisis and insecurity has increased dramatically since 2008. However, this vocabulary is not solely limited to the financial crisis. The "insecurity and crisis" frame referred both to the financial crisis, the armed conflict in Ukraine and to the refugee crisis.

1 Introduction

During the past decade advances in text mining have demonstrated that political issues and topics can be identified from unstructured textual documents with good accuracy. Examples from digital parliamentary studies or computational discourse analysis have brought up evidence how especially topic modelling can be applied on large document corpora to gain overviews over thematic structures of political texts, but also how these structures change over time. Most of these studies analyse large, even huge textual corpora with thousands of lengthy documents. Such "big text" analysis is often seen as typical example of how digital methods can – even should – be applied in social sciences and the Humanities. 1

In this article I focus on the problem of using digital methods and tools on a smaller corpus of unstructured textual documents, namely New Year's addresses, with varying political content and as addressed to a wide public audience, the citizens in general. Although politicians often favour longer speeches and verbal battles when it comes to struggle for political issues, especially the growing relevance of social media driven political communication has strengthened the role of short textual documents like tweets, postings, statements or commentaries in political discourse. Since most 2 short textual contributions are designed to take stand on current affairs and actual political questions, they offer an interesting empirical material to explore not only how politicians take stand on current affairs, but also the structure and dynamics of issue-driven political discourses.

This article is based on the understanding that sort political texts can be a valuable source when it comes to identifying topics and issues that really matter under the current political circumstances. Rising a current topic in a short political text can be seen as an act of politicalisation [Palonen 2007b, 62–66], i.e. an action marking the issue as political by detecting its political potential and thus expanding the presence of the political and presenting a particular representation of “reality” [Koller 2014, 164]. This is especially relevant for addresses and speeches given by high representatives of a nation like president or prime minister. All issues taken up in such texts are automatically labelled as being politically significant, so that the speech itself can be read as an attempt to persuade the target audience to accept this position [Shim et al. 2015, 56]. 3

The empirical results of this article suggest that New Year’s speeches offer a valuable and interesting source for the exploration of topical issues enjoying political and societal relevance. The argument here is rooted in the understanding that if an issue is taken up in a New Year’s address, the issue is considered as important and becomes politicised [Palonen 2007a]. Hence, the corpus of German, Austrian, and Swiss New Year’s speeches offers an interesting source to explore topical discourses from both national and comparative, cross-national perspective. 4

From the methodological perspective the application of digital text mining and analysis tools on a corpus of short texts consisting of several hundred words in average requires an adapted approach. This, because most tools for text mining have been developed for large-to-huge text corpora with millions of words and use statistical methods to model the corpus. Although these methods produce results also when applied to small corpora, the results tend to be biased and their reliability compromised. In order to circumvent these problems I use a rather novel approach of Text Network Analysis (TNA) using a non-linear approach to re-construct textual documents as a network of words appearing close to each other in the original documents [Paranyushkin 2011]. One of the key advantages of TNA is that since TNA is rooted in (Social) Network Analysis the resulting network can be visualised and analysed by methods and tools developed for network analysis. Another advantage is that networks can be created without any arduous preprocessing, although lemmatisation and stopword removal are strongly encouraged. One main objective of this article is the introduction of TNA as an alternative, relatively easy to use explorative tools for both text mining and topic modelling. 5

The analysis and results presented in this article should evidence the power of TNA as a topic modelling and exploration tool for short(er) text documents. The analysis introduces a non-linear way to analyse texts as networks and to present results based on graphical visualisations making it easier to capture and understand structural characteristics hidden in the corpus. The results, in turn, provide interesting insights in thematic continuity and change in political topics in the analysed three German-speaking countries. Further, a detailed discussion of the results confirm the accuracy of the main findings of my TNA analysis, thus providing additional support for the usefulness of the method. 6

The article is structured as follows. The first main section discusses TNA as an alternative digital tool for topic modelling. The second section introduces the empirical corpus used in this article and describes the data preparation process, as well the analytical steps needed for TNA. The corpus consists of New Year’s addresses given by political leaders in Germany, Switzerland and Austria between 2000 and 2021. All speeches are given in German language. The third section presents and discusses the main findings by reflecting them against contextual factors and previous research. The article is rounded up with concluding remarks critically summarising the main findings and achievements. 7

2 Text Network Analysis

Since the first decade of the 21st century digital research tools and methods have gained a stronger role and impact among social scientists interested in parliamentary debates, political speeches or other textual materials on political affairs. A great majority of such kind of contributions has focused on the identification of the main content through the analysis of commonly used words, by focusing on the content different words as used, or through the analysis of co-occurring words.[1] 8

A known challenge with political speeches is that such texts normally tackle several topics or issues. However, changes from one topic to another can be be hard to detect with methods based on a bag-of-words approach. This, because such an approach transforms the document in a word frequency table. Although such a table is rather simple to handle and analyse, the other side of the coin is that such a transformation destroys the original structure of the text, making it impossible to gain understanding about where different words appear in the original text [Hearst 1994]. 9

[Jelveh et al. 2014] present evidence for an existence of latent shared ideological positions among experts. In this article I expand this underlying assumption to include politicians as well, and make the initial assumption that politicians tend to what I would call “thematic recycling”, i.e. the endeavour to take up the same – or similar – issues over and over again on different occasions. Hence, a corpus of political textual documents from a politician can be assumed to share similar topics across different documents. Further, as in the case of expert texts we can, with some caution, generalise this assumption and expect to find shared topics by politicians sharing same or highly similar political insights rooted in certain ideological positions. Accordingly, we can expect to find shared latent topics and political positions also in documents delivered by different political actors sharing a similar political landscape. 10

Thus far most studies dealing with the problem of identifying topics across documents in a document collection have applied methods of topic modelling, mostly Latent Dirichlet Allocation (LDA) [Blei et al. 2003]. There are many good reasons for this, maybe the most important being its flexibility in defining the “document,” allowing the researcher to condition the topic distribution on paragraphs or time chunks, for example. However, as a method based on statistical modelling LDA also has some important caveats and pitfalls limiting its analytical power especially as regards shorter text documents and smaller corpora. One limitation being highly relevant for this article is that most LDA-based applications require datasets of several hundred documents, at minimum, to ensure validity and reliability of the results [Light 2014, 117]. 11

Although identifying different topics is an interesting challenge of explorative text analysis, this article tackles a bit different challenge: how to explore thematic pathways across documents in a text document corpus. Accordingly, the aim is to explore, on the one hand, how words are used within statements, and how thematically similar statements “glue” different texts together across contexts, on the other. These thematic pathways building on statements and connecting different contexts should help us to understand the deep thematic backbone underlying the document corpus [Busse 2013, 43ff]. 12

The analysis presented in the next section is an application of Text Network Analysis (TNA), a method originally presented by Paranyushkin (2011) and rooted in theoretical foundations of Social Network Analysis (SNA). According to Paranyushkin TNA seeks to discover “repetitive patterns derived from the text’s structure, using their connectivity and the intensity of interactions between them as the only criteria for their belonging together” [Paranyushkin 2011, 5]. It is a non-linear approach transforming texts to a network of words by respecting the original structure of the texts. [Shim et al. 2015, 58], who use TNA to identify policy frames of nuclear energy policy across different countries, summarise the method by stating that it seeks “to identify salient words and concepts in order to extract underlying meanings and frames from the structure of concept networks.” As the contributions of both Paranyushkin and Shim et al. evidence TNA offers a very powerful tool with a robust fundament in network analysis and a rather simple and straightforward application. Further, text networks can be created as flexibly conditioned on e.g. sliding window techniques, sentences, or paragraphs. This gives a researcher additional flexibility to test different approaches to figure out the most suitable text unit for the analysis. 13

This article stems from the assumption that TNA as a non-linear analysis of textual data could provide us with new ontological understanding about structural coherence and holes within a document corpora. This new ontological understanding, however, is not only expected to shed light on how concepts connect in order to form statements. Instead, it can also help us to explore structures relevant for the understanding of how statements are organised within and across contexts. Through the application of network visualisations and the analysing of structural properties of text networks the underlying data can be explored from alternative perspectives in order to trace back discursive patterns within the text corpus. 14

One key difference of TNA compared to more traditional tools and techniques of text mining is that TNA understands its input as network data consisting of *nodes* (or *vertices*) connected by *edges* and having an own structure (topology) [Paranyushkin 2011] [Ruohonen 2013] [Scott 2013]. A common technique is to use word-based units – single words or n-grams - as nodes and co-occurrences as edges. Although this is the most common way, nothing actually prevents the researcher from using e.g. phrases as nodes, just to mention another example illustrating the analytical flexibility of TNA. Since TNA works based on network theory, it does not require any predefined ontological structure or pattern of text, nor does it superimpose any extra layer of ontologies on top of the textual data, which may - as [Paranyushkin 2011, 4] correctly states – create “a strong subjective (and even cultural) bias into the structure of the resulting text graphs.” Further, the researcher is not limited to the use of text theories or text mining tools (although nothing prevents her to use this kind of theories or tools in the analysis), but can apply a wide range of network analysis tools and methods to explore, analyse and visualise structural properties of the text networks. In other words, TNA gives the researcher the possibility to explore and analyse textual units conditioned to a certain structure and, thus, to explain why different textual units take different positions and roles by focusing on the network structure of the text. 15

3 Data and Workflow

The empirical case study in this article is based on New Year’s speeches of the three German-speaking countries, Germany, Austria and Switzerland[2] covering the years from 2000 to 2021. These three countries were selected, first, based on the shared language providing us the possibility to compare the use of language of high politics across these countries. Second, all three countries belong to the same cultural-political space, Central Europe, with a strong, continental orientation and a centuries-long experience and tradition of German cultural, economic, and political influence, but also dominance [Jordan 2005]. Besides the two more or less uniting criteria, the third selection criterion was a dividing factor: Germany and Austria are members in the European Union (EU), whereas Switzerland is only indirectly involved in European integration through its membership in the European Free Trade Association (EFTA) and a set of bilateral agreements with the EU. Switzerland is also member of the Schengen area allowing any person, irrespective of nationality, to cross the internal borders without being subjected to border checks. Currently, most of the EU countries plus the non-EU countries Iceland, Norway, Switzerland and Liechtenstein have joined the Schengen area. Against this background I expect these three countries to form an “imagined community” (Benedict Anderson) and to share certain values, interests and world views reflected by the New Year’s addresses. Further, I expect German and Austrian New Year’s speeches to share discursive patterns related to EU affairs, especially to the euro crisis (since 2008), the conflict in Ukraine (since 2014), and the refugee crisis (since 2015).[3] I also could explore the impact of the global Covid-19 pandemic disease on both the content and vocabulary across the three countries yet this impact is visible only in the speeches for the year 2021. 16

New Year’s speeches held by leading politicians–mostly prime ministers or presidents–have a long and firm tradition in Europe and “have become an institution instead of being a crowing event for conciliatory efforts” [Portman 2014, 89]. From the perspective of political communication, New Year’s addresses fulfil a triple function in the intersection of the past, the present and the future. First, they summarise the past year from the perspective of the political leadership and, hence, recall, reconstruct and remind the most important events of the year. Second, New Year’s speeches describe the present and, thus, can be understood and analysed as reality constructions, as windows to the current state of affairs. And third, New Year’s speeches serve as road maps to the future, into the new year. In this sense, a New Year’s speeches summarises the most important future challenges, expectations and opportunities. 17

There exist only a small number of dedicated studies on New Year’s speeches [Heikkinen 2006] [Portman 2014] [Purhonen and Toikka 2016]. I could not identify any specific reason for this, despite that New Year’s speeches are dominantly given in non-English speaking countries in Europe. As a text genre, however, New Year’s addresses are comparable with the Presidential Inaugural speeches of the U.S. and this article methodologically benefits from the study of [Light 2014] on the United States’ Inaugural Addresses. From the perspective of social sciences I am not primarily interested in the use of language, but issues mentioned and described in the addresses. This viewpoint is rooted in the assumption that a political phenomenon, topic or event gains in importance if mentioned in a New Year’s address. Here I also follow the idea that New Year’s speeches offer insights into shifts in political priorities, interests, 18 and opinions [Light 2014]. Against this background I am convinced that a longitudinal, computational analysis of German, Austrian and Swiss New Year's addresses can help us to a better understanding of how the political landscape has changed between 2000 and 2021, but also where the differences and similarities can be found between these three countries.

The speech corpus covers a period of 22 years and contains 64 speeches. There are two missing years – 2002 and 2017 – in the Austrian speech collection. In Austria Federal President delivers the New Year's speeches and between 2000 and 2004 the speeches were delivered by Thomas Klestil, between 2005 and 2016 by Heinz Fischer, and since 2018 by Alexander Van der Bellen. Although speeches from 2005 onwards can be downloaded from the official website of the Austrian Federal President. Fortunately, the speeches of 2000-2001 and 2003-2004 could be found in different Austrian media archives. As regards the year 2002, neither the President's office nor the National Archive of Austria could provide any information about where this particular New Year's speech could be found, so that I was forced to drop this year from the corpus. In 2017 Austria had an extraordinary political situation with no Federal President in office, so that no New Year's speech was given. 19

In Germany, the New Year's speeches are held by Federal Chancellor and full texts are available on the official website. [4]. The first six speeches (2000-2005) were held by Chancellor Gerhard Schröder, the remaining sixteen from 2006 to 2021 by Chancellor Angela Merkel. In Switzerland, the annual New Year's speech is delivered by the President of the Confederation and all speeches since 1980 are available online on the official website of the President of the Confederation. As the term of the Swiss President is just one year all New Year's speeches are held by a different person (except Micheline Calmy-Rey, Pascal Couchepin and Moritz Leuenberger, all of them re-elected and with two speeches). 20

Despite the varieties in the use of language in the German-speaking Europe, formal speeches by top-level politicians can be seen as an appropriate source of data, as policy discourses are comprised of policy addresses and speeches about policy issues [Van Dijk 1977]. 21

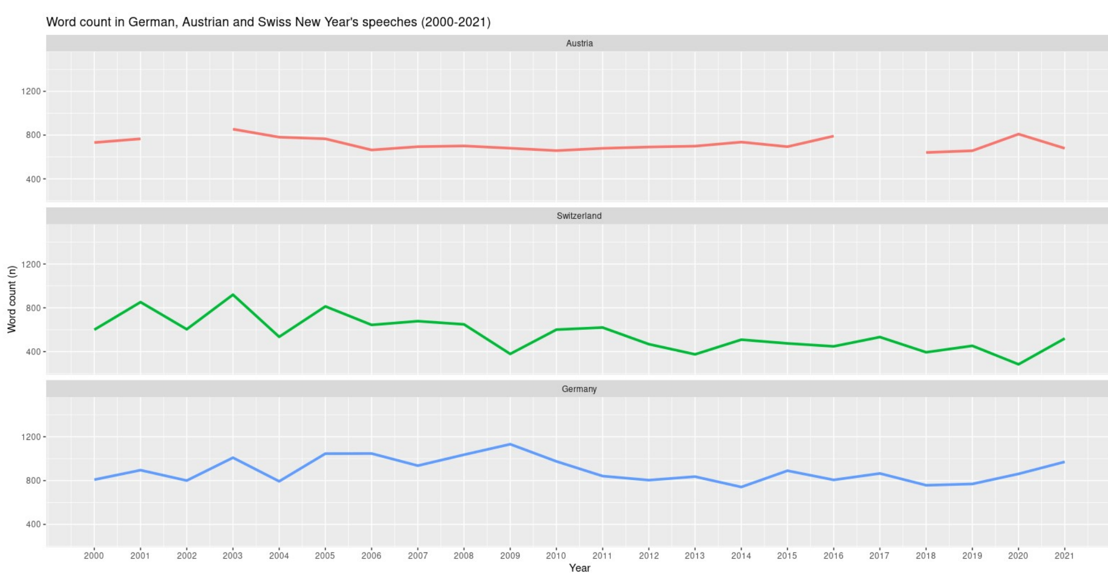

Figure 1. Length (# of terms) of New Year's speeches in Germany, Austria and Switzerland 2000-2021.

New Year's speeches are traditionally broadcasted over TV or radio within a relatively short time of 10 to 15 minutes. Consequently, these speeches are relative compact in size. The average speeches length of a German new New Year's speech was 892 tokens and standard deviation (sd) was 111. In Austria and Switzerland the average lengths were 719 (sd=58) and 562 (sd=159) respectively (Figure 1). Quite understandably the amount of unique tokens is somewhat lower, since certain words are used several times during a speech. For German New Year's addresses, there were 362 unique tokens in average (sd=33), for Austrian and Swiss speeches the average number of unique tokens was 317 22 (sd=27) and 255 (sd=60) respectively. Swiss New Year's speeches seem to have become shorter over time, whereas both German and Austrian New Year's speeches remain relative stable over time (Figure 1). Generally speaking a country's size seem to matter here, as German New Year's speeches are slightly longer than the Austrian or Swiss addresses. My explanation for this observed difference stems from the differences in political and economic importance. Germany's central role in European politics results in broader political agenda, so that there are more issues and topics to be addressed to in a New Year's address.

As regards the analysis workflow, all collected speeches were saved in separate files in plain text format. These files 23
were then imported into RStudio, a graphical end-user environment for the statistical package R. The research data was created and analysed in the four steps. First, I used the package 'udpipe'[5] to tokenise, part-of-speech tag, and lemmatise the documents. After this step, the data was structured as a data table consisting of 53893 words and 5277 unique lemmata, enriched with descriptive metadata about the country and the year, as well sentence indices. In the third step, I used text mining and explorative data analysis tools from the package 'tidytext' to carry out different text analyses and to transform textual data to different sets of text network data. The applied text mining methods will be described in the next section. For the text network network creation I used both bi-grams by linking two consecutive words, as well skipgrams of length two by combining non-adjacent words that were three, four or five words apart.[6] Both the bi-grams and skipgrams were created within the boundaries of the same sentence based on the sentence index included in the data structure. (2) in the case of concepts used in a single country, only concepts which occur more than three times are included. Text networks were visualised and analysed with 'visone', a fully-fledged, platform independent software offering powerful layouts for network visualisations and a comprehensive set of tools for network manipulation, transformation and analysis.

4 Results

Figure 2. New Year's speeches similarity network (2000-2021).

Figure 2 presents the addresses network for the whole corpus based on the similarity of the New Year's addresses. 24
Similarity is here calculated as cosine measure, a term-based comparison between documents in a document corpus, and has a value range from 0 (no similarity at all) to 1 (full similarity).[7] I reduced noise in the document-term matrix by removing both stopwords and all terms except nouns, verbs, pronouns, and proper nouns. In the network visualisation nodes represent different New Years speeches and are coloured with country-specific colours so that German addresses are coloured with yellow, Austrian with red, and Swiss speeches with white. I have also added labels to help the reader to identify different addresses. Nodes are connected by edges, of which width is visualised proportional to the cosine measure of the two speeches connected by that edge. In my data the cosine measures ranged from 0.103 (between the Swiss speeches 2006 and 2020) to 0.641 (between the German speeches 2009 and 2010), but I have removed edges with cosine measure below 0.400 in order to improve the readability of the network visualisation. Further, I used a force-directed layout and optimised it to position speeches with higher similarity closer, those with lesser to each other so that the clusters of more similar speeches are easier to identify.

As regards the similarity network two observations seem appropriate. First, the New Year's speeches evidence a 25
stronger similarity within, and weaker similarity across countries. This is clearly visible in the network structure as well, as speeches from different countries form clear and distinct cluster in the visualisation (see Figure 2). Overall, the average similarity for cross-country speeches is 0.318 (min=0.114, max=0.514), for national speeches the average similarity is 0.384 (min=0.103, max=0.614). If we set the threshold value for cosine measure to 0.5, only two cross-national combinations, namely Austria 2018 – Switzerland 2020 (0.505) and Germany 2014 – Switzerland 2018 (0.514), exceed this limit. Varieties of national political agenda and circumstances seem to explain these observed differences, thus underlining the primacy of national topical issues. An interesting result is also that the cosine similarity between German and Austrian speeches is significantly higher (mean=0.355) than that between German/Austrian and Swiss

speeches (mean=0.277). This difference is at least partly explained by the EU membership, as certain topical issues both German and Austrian speeches tackle are connected to EU politics.

26

And second, similarity seems to be stronger between between chronologically close speeches and weaker between speeches with greater temporal distance. I tested this with a simple Pearson's correlation test and calculated the correlation between the cosine measure and the temporal distance in years between the New Year's addresses. The correlation was -0.085 (p<0.001), thus confirming the hypothesis that the similarity measure decreases when the temporal distance between compared speeches increases. This seems logical when reflected against the fact that New Year's speeches tackle topical issues at the end of a year. The longer the temporal distance between two addresses, the more probable there has been significant changes in political circumstances affecting the vocabulary used in the addresses. Hence, this observation also gives support to the argument that New Year's speeches offer an interesting and valuable perspective on a year's prevailing political reality.

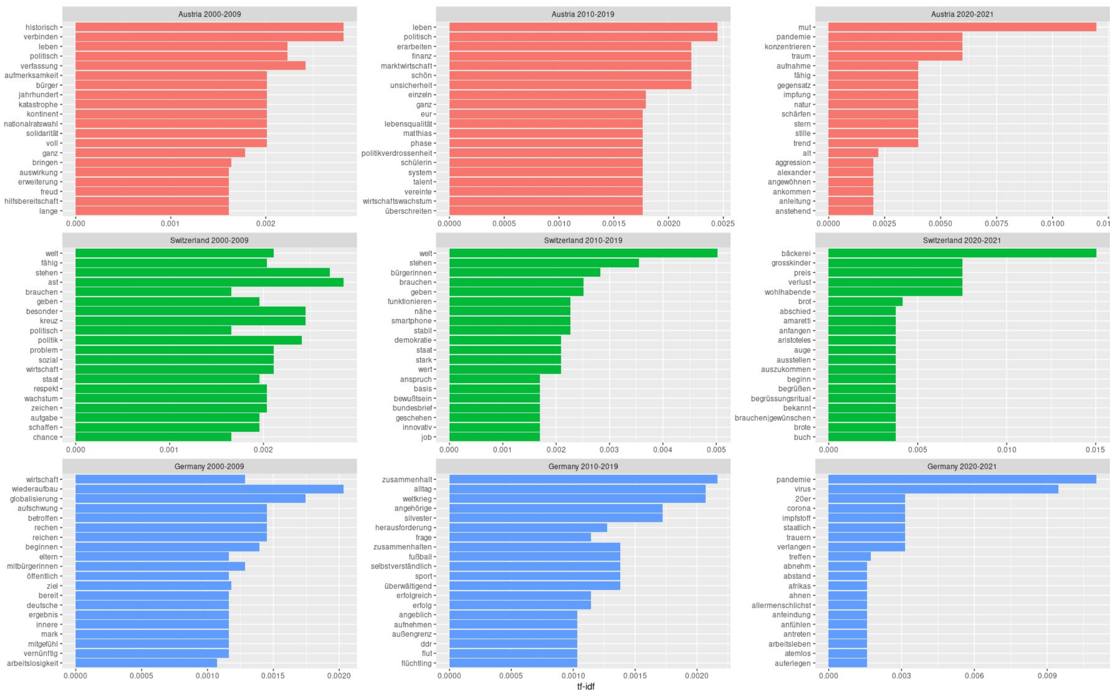

Figure 3. The 20 most important words per country in different periods (tf-idf analysis).

27

Results from the document-level TNA suggested the existence of variance in vocabulary between the countries and over time. In order to better understand this variance I applied a specific text mining method, called term frequency, inverse document frequency *(tf-idf)* analysis, to adjust the frequency of a word for how rarely it is used in the collection of New Year's speeches of each country. A word's inverse document frequency (*idf*) was then applied to explore words that are not used very much during the whole period from 2000 to 2021. I divided each national collection of New Year's speeches into three distinct periods of 2000-2009, 2010-2019, and 2020-2021. Here the idea was to capture changes in the vocabulary over time within a country, most probably caused by changes in topical issues or political circumstances. The results of this *tf-idf* analysis are presented in Figure 3. In this barplot we can see for each country the most important words in each distinct period of time. Hence, the Figure 3 highlights temporal changes within each country, but preserves national differences. In my opinion three central findings from this analysis are worth being discussed in detail.

28

First, the connection to topical political issues and circumstances is especially evident in Germany across the time periods, but also clearly visible in Austrian speeches of the last two periods. As regards Germany, during the first period (2000-2009) words like “wiederaufbau” (rebuilding), “betroffen” (upset), and mitgefühl“” (sympathy, compassion), together with such words like “naturkatastrophe” (natural disaster), “region” (region), or “erschüttern” (shock, shake) belonging to the top-30 *tf-idf* words in this period, tackle the 2002 European flood. Germany was the hardest hit country in Europe and the flood resulted in heated political discussions in the eve of the Federal elections 2002. In Austrian speeches of this period words like “jahrhudert” (century), “katastrophe” (catastrophe), and “hilfsbereitschaft” (helpfulness) document the presence of the 2002 flood – which hit also Austria quite severely – also in Austrian political discourses.

Another issue in this period was the outbreak of the sovereign debt crisis in the eurozone in 2008/2009. This economic crisis is present in German speeches through words like “globalisierung” (globalisation), “wirtschaft” (economy), “arbeitslosigkeit”, (unemployment). Contrary to Germany, the eurozone crisis is not very present and visible in Austrian speeches in this period of time. A possible explanation to this difference might be found in the different roles the two countries played in the early stage of the eurozone crisis. Germany as the strongest economy in the eurozone played a central, designing role from the very beginning of the eurozone crisis [Müller-Brandeck-Bocquet 2010] [Jsnning and Möller 2016] [Kundnani 2016] [Bulmer 2018], whereas Austria were not strongly hit by, nor a central player in this crisis. In addition to that I would also stress the economy-oriented political culture in Germany. The strong role of economic questions as a central shaping factor of German domestic and European politics increases the weight of topical economic issues in public debates. Against this background it is not surprisingly that the eurozone crisis occupied a more significant space in public debates in German compared to Austria [Allen 2005] [Maull 2018]. 29

During the second period (2010-2019) the focus of German New Year’s speeches shifts from economy to refugee politics. In 2015 Chancellor Merkel decided to keep German state borders open to refugees coming mainly from the Middle East. Although this political turn was captured in positive political slogans like “Willkommenskultur” (culture of welcoming) or – most prominently – “Wir schaffen das” (we can do it/ we manage it), the fact that words like “flüchtling” (refugee), “zusammenhalt” (sticking together), “herausforderung” (challenge), “außengrenze” (outer border), or “aufnehmen” (receive, welcome) are among the top-20 terms underline the change the refugee crisis caused in German public and political discourses during this period of 2010-2019. From the methodological point of view the identification of terms with a clear connection to the refugee crisis evidence the usefulness of *tf-idf* analysis when it comes to explore discursive changes. Contrary to Germany, the refugee crisis was not central in Austrian New Year’s addresses. Instead, the national political and economic setting seem to dominate Austrian addresses. Terms like “marktwirtschaft” (market economy), “lebensqualität” (quality of life), or “finanz”(finance) tackle economic consequences of the eurozone crisis. The word “politikverdrossenheit” (jadedness with politics) refers to domestic political turbulences behind the rise of the right-wing populist Freedom Party of Austria (FPÖ) in the mid-2010s. Although this rise was partly supported by the refugee crisis, it was also an expression of disillusionment with the political elite. 30

Second, as expected the outbreak of the Covid-19 pandemic disease in Europe in the first half of 2020 resulted in a clear change, especially as regards the German and Austrian New Year’s addresses. Words like “pandemie” (pandemic), “impfung”/“impfstoff” (vaccination), “abstand” (distance), or “corona” (coronavirus) poke from the results as indicators for this discursive change in political and public discourse in Germany and Austria. Once again, the results also confirm the usefulness of *tf-idf* analysis for the exploration of discursive changes through changes in vocabulary over time. 31

And third, the overall results confirm the hypothesis that Germany and Austria are, what comes to political agenda and circumstances, as well to topical issues, closer to each other than to Switzerland. It is, however, a somewhat odd observation that Swiss speeches seem to *avoid* topical political issues and to concentrate on the construction of neutral narratives. A possible explanation for this might be found in the fact that in Switzerland, as opposed to Germany and Austria, the federal president giving the New Year’s speech is not the head of state, but just the head of Switzerland’s Federal Council, and only carries out some representative duties. Hence, the federal president is merely a mediator between the different parts of the Swiss federal state. 32

The results presented thus far indicate two main aspects. First, New Year’s addresses offer a valuable and reliable source to explore topical issues and tackle changes in political and public discourses. And second, although the national political agenda and circumstances seem to have a stronger impact on the content of the addresses, significant European political incidents and events, e.g. the refugee crisis or the Covid-19 pandemic disease, seem to cause changes across countries. This leads us to the final analysis, an attempt to explore shared topics. For this analysis I 33 constructed text networks based on bigrams, i.e. word co-occurrences of two adjacent words, and of skipgrams, i.e. combinations of two non-adjacent words that are three to five words apart [Jelveh et al. 2014, 1805]. The bi- and skipgrams were created for words occurring within the same paragraph. Since the idea is to analyse topics across countries all word pairs co-occurring in New Year's speeches of one single country only were removed. In the last step I calculated a co-occurrence weight measure by dividing the number of speeches a word pair co-occurs by the total number of speeches (64) in the multi-national corpus. The interpretation of this weight measure is rather straightforward: the higher the value, the more significant the word pair for the cross-country content.

Figure 4. New Year's speeches core bigram network across the three countries (2000-2021).

Figure 5. New Year's speeches core skipgram network across the three countries (2000-2021).

Figure 4 visualises word co-occurrences network across countries based on bigrams, whereas Figure5 visualises word 34
co-occurrence network across countries based on skipgrams. Hence, both visualised text networks present shared word co-occurrences central for the multi-country corpus of German, Austrian, and Swiss New Year's addresses, i.e. word pairs that can be interpreted as fundamental for the – put in Benedict Anderson's (1991) famous terminology – imagined discursive community of these three German-speaking countries. The bigram-based core text network consists of 336 words and 565 co-occurrences, whereas the skipgram-based core text network consists of 423 words and 1157 co-occurrences.

In order to visually highlight the most important structural properties both visualisations apply identical visualisation 35
effects:

- Node size is mapped to the node's degree centrality value and node label size is mapped the nodes betweenness centrality measure. In network theory centrality, in general, indicates a node's position in the network and can be calculated either relative to a node's direct neighbours or the whole network. A node's degree is the simplest centrality measure and equals to the number of connections the node has to other nodes. Betweenness, as the term itself indicates, defines centrality by analysing where a node is placed within the network. Consequently, a node's betweenness centrality score is computed by taking into consideration the rest of the network and by looking at how many times a node sits on the shortest path linking two other nodes together, thus helping to identify nodes having "a high probability of occurring on a randomly chosen shortest path between two randomly chosen vertices" [Hasu and Kao 2013] [Prell 2012, 103–104]. Considering meaning circulation across the entire network, the latter capability is assumed to be more relevant, since betweenness centrality "shows the variety of contexts where the word appears, while high degree shows the variety of words next to which the word appears" [Paranyushkin 2011, 13]. This difference is important to keep in mind and therefore I use betweenness centrality to measure a concept's general status in the text network. As both [Paranyushkin 2011] and [Shim et al. 2015] point out concepts with high betweenness and degree centrality play a meaning circulation role across texts in the document corpus, whereas concepts with low centrality measures are peripheral concepts and, thus, typical only for a certain part of documents. Between these two extremes are located concepts with high betweenness but low degree centrality and concepts with low betweenness but high degree centrality. The former play an important role as bridging concepts between local communities, the latter, in turn, are local hubs within a cluster [Shim et al. 2015, 59f]. This mapping strategy allows us to easily identify the most important and influential words in the graph.
- Node colour is mapped to the topic cluster the node belongs to. One of the proposed key advantages of TNA is bound with the possibility to exploit network community detection techniques.[8] The underlying idea

here is that when textual documents are reconstructed as text networks a word's position in the network is not random but determined relative to the context it appears in. Consequently, words belonging to the same (or similar) context are assumed to cluster across documents, so that we could be able to identify groups of words in the network structure having dense connections *within* the group, but sparse connections to other parts of the text next. These groups are here interpreted as topics describing the main thematic content of the document corpus. By leaning on the promising results of [Paranyushkin 2011] and [Shim et al. 2015] I applied a modularity-based community detection method called "Louvain method" based on the assumption, that nodes being more densely connected together than with the rest of the network construct a network community [Blondel et al. 2008]. The method identified in total 13 clusters - in my interpretation: topics - in the bigram-based and 12 topics in the skipgram-based text network.

- Edge width is mapped to the weight of the co-occurrence of the word pair. The more often two words co-occur in the document corpus, the wider the connecting edge is visualised in the network graph. Since the graph includes only word pairs shared by at least two countries, edge effects help us to identify word pairs used together frequently across countries.
- Layout used for the visualisation is a radar-like centrality layout, in which a node's position is determined by its topic cluster and links are arranged according to their weights. The underlying idea of this layout is to gather nodes belonging to the same cluster on the same circumference. Further, the visualisation makes also connections between the clusters – i.e. word pairs, of which words belong to different topic clusters – easier to identify. As regards the meaning circulation across the multi-country text network word pairs connecting two topics can be analysed from the perspective of structural holes. These connections bridging two separate topics by closing a structural hole in a network can make us aware of possibly existing latent similarities between these two topics [Burt 2004].

Bigram text network core communities	**(Total # of clusters: 13, modularity: 0.55)**
Cluster	*Core content words of the topic*(a)
#1: Solidarity / Sense of community (55 words in total)	bürgerin/bürger, frieden, geschichte, hilfbereitschaft, sicherheit, stabilität, verantwortung, wiederaufbau, zukunft, zusammenhalt
#2: Crisis recovery (41)	einsetzen, ereignis, erinnern, freuen, generation, glauben, krise, verbinden, weltkrieg, überwinden
#3: Covid-19 era (36)	erfolgreich, familie, friedlich, meistern, pandemie, phase, resignation, schwierig, zeit, zusammenleben
#4: Europe (32)	entwicklung, erweiterung, europäisch, hoffnung, krieg, mitgliedstaat, parlament, terror, union, wählen
Skipgram text network core communities	**(Total # of clusters: 12, modularity: 0.35)**
Cluster	*Core content words of the topic*(a)
#1: Crisis recovery (62 words in total)	erinnerung, friedlich, krieg, mauer, notwendig, optimismus, pandemie, vergessen, verlust, wirtschaftskrise
#2: Faith in the future (51)	aufschwung, bemühen, gestaltung, neugier, selbstvertrauen, technik, wirklichkeit, wissenschaft, zufriedenheit, zuversicht
#3: Economy (46)	arbeitsplatz, beitrag, einsatz, erreichen, global, idee, leisten, offen, unternehmen, wirtschaft
#4: Solidarity / Sense of community (45)	alltag, bürgerin/bürger, demokratie, entwicklung, familie, frieden, gemeinsam, gemeinschaft, stehen, welt

Table 1. Most important topics in the New Year's speeches across countries.

(a) Core content words include ten (10) most influental words relative to the topic of the cluster (listed in alphabetical order). Bold words indicate top ranking words both in degree and betweenness centrality, i.e. words being central for meaning circulation across the text network.

Table 1 summarises the most significant results of the community detection analysis of both the bigram based and the skipgram based text network by focusing on the four largest communities found by the Louvain algorithm. As expected there is some variance across topics between the bigram and skipgram networks. I consider this variation as normal when the differences in network data creation is taken into account. Actually, against this background, the two networks should not be considered as exclusionary but complementary to each other. I have given each community a descriptive label based on a manual evaluation of the vocabulary linked to the community. 37

There are two shared topics. I have labelled the first “Solidarity / Sense of community” and it is the largest topic in the bigram and fourth largest topic in the skipgram text network. Both includes words “bürger/bürgerin” (citizen) and “frieden” (peace). In the bigram network this topic is with respect to content more focused on helpfulness and security, indicated through words like “hilfebereitschaft” (helpfulness), “sicherheit” (security), “stabilität” (stability), “verantwortung” (responsibility), and “zusammenhalt” (togetherness, solidarity). The skipgram network, in turn, has in this topic a stronger bias towards democracy and society embodied by words like “alltag” (everyday life), “demokratie” (democracy), “familie” (family), and “gemeinschaft” (community). 38

This topic appears in different contexts over time. In the German New Year’s speech in 2006 the topic is used to emphasise the inter-linkages between Germany and the international, global community: 39

[9]

> *Deutschland ist keine Insel. Wir stehen in einem internationalen Qualitätswettbewerb, der alle Bereiche unseres Zusammenlebens betrifft: Welcher Nation gelingt es am besten, die schöpferischen Kräfte ihrer Menschen zu wecken? Wie offen ist eine Gesellschaft für Neues? Was bietet sie jungen Familien? In welchem Land gibt es die besten Schulen und Hochschulen? Wie gut gelingt das Miteinander von Einheimischen und Zuwanderern?* (Germany 2006)

In the German speech of 2015, the topic is used in the context of the refugee crisis to foster solidarity among citizens, but also to strengthen confidence and commitment to the community and its values: 40

> *Genauso klar ist: Nur mit offenen Diskussionen und Debatten können wir Lösungen finden, die langfristig Bestand haben und von Mehrheiten getragen werden. Wir sind es, die Bürger und ihre gewählten Repräsentanten, die entwickeln und verteidigen werden, was dieses unser liberales und demokratisches Land so lebenswert und liebenswert macht. Wir sind es, die Lösungen finden werden, die unseren ethischen Normen entsprechen, und den sozialen Zusammenhalt nicht gefährden. Lösungen, die das Wohlergehen der eigenen Bürger berücksichtigen, aber nicht die Not der Flüchtlinge vergessen.* (Germany 2015)

The Austrian speech of 2007, in turn, highlights the role of persons in different positions of trust for the security, stability and confidence of the community: 41

> *Lassen Sie mich abschließend die Gelegenheit benutzen, um den vielen Frauen und Männern herzlich zu danken, die in den verschiedensten Funktionen für unser Gemeinwohl, für unsere Sicherheit, für unsere Gesundheit und für den Zusammenhalt unserer Gesellschaft buchstäblich oft Tag und Nacht beruflich, oder auch als freiwillige Helferinnen und Helfer tätig sind. Der Wert dessen, was hier geleistet wird, kann gar nicht hoch genug eingeschätzt werden.* (Austria 2007)

And finally, the following two quotes from the Swiss speeches of 2017 and 2021 establish connections between the economic welfare, sovereignty and stability of the country and the solidarity and togetherness of the society: 42

> *Stabil und erfolgreich wollen wir auch in Zukunft sein. Dazu braucht es mehr denn je den Zusammenhalt. Unsere Gesellschaft ist stark, weil wir erprobt sind im Versöhnen von Ansprüchen.* (Switzerland 2017)

> *Wir Schweizerinnen und Schweizer müssen zusammenstehen. Nur so können wir als Land einstehen für die Interessen von uns allen: für unsere Gesundheit und unser wirtschaftliches Wohlergehen, für Frieden*

und Verbundenheit, für Freiheit und Unabhängigkeit, – kurz: für alles, was uns seit Langem lieb und teuer ist. (Switzerland 2021)

The second shared topic is labelled "Crisis recovery". The content vocabulary of this topics indicates that this topic is not limited to the Covid-19 pandemic disease only. Instead, this topic seems to be used over time to embed a current, topical crisis into a wider, historical context. Although the clustering algorithm allocates topical words like "pandemie" (pandemic) or "wirtschaftskrise" (financial crisis) to this cluster, in my interpretation other content words like "generation" (generation), "krieg / weltkrieg" (war / world war), "überwinden" (overcome), "optimismus" (optimism), "mauer" (wall), and "erinnern" (remember) stand for this wider historical context. 43

Both in the Swiss speech in 2010 and German speech in 2011 this topic was used to frame the eurozone crisis. Both quotations manifest a clear optimism that the crisis has been successfully fought and the country will recover rather quickly: 44

Wir verfügen über eine starke Wirtschaft und eine solide Finanzpolitik. Wir haben es geschafft, trotz grosser Wirtschaftskrise die Schulden nicht übermässig wachsen zu lassen. Das wird uns im Aufschwung stärken. (Switzerland 2010)

Deutschland hat die Krise wie kaum ein anderes Land gemeistert. Was wir uns vorgenommen hatten, das haben wir auch geschafft: Wir sind sogar gestärkt aus der Krise herausgekommen. Und das ist vor allem Ihr Verdienst, liebe Mitbürgerinnen und Mitbürger. (Germany 2011)

In 2020, the Austrian speech used this topic in a totally different context combining the most important crises and challenges of the current time, i.e. climate change, digitalisation, and migration, as well questions related to gender equality and welfare state reforms. 45

Zur Klimakrise kommen weitere große Herausforderungen: Wie werden wir künftig arbeiten? Welche Antworten geben wir in Österreich auf die Digitalisierung? Wie soll sich unser Wirtschaftsstandort entwickeln? Wie gehen wir mit Migration um? Und was tun wir, um Frauenrechte zu stärken? Haben wir ausreichend drüber nachgedacht, das Nötige im Bildungsbereich anzugehen? Welche Reformen sind im Gesundheits- und Sozialsystem notwendig, um soziale Sicherheit und sozialen Zusammenhalt für die Zukunft zu gewährleisten? (Austria 2020)

Both text networks produce also distinct topics. In the bigram network we can identify the unique topics labelled by myself "Covid-19 era" and "Europe", in the skipgram network such topics include "Faith in the future" and "Economy". As regards the bigram-specific topics, "Covid-19 era" not only embodies the strenuousness of the pandemic disease by content words like "pandemie" (pandemic), "resignation" (resignation), or the word combination "schwierig" + "zeit" (hard times), but also seeks to create hope through words like "erfolgreich" (successful), "familie" (family), "meistern" (control), "zusammenleben" (life together). The topic "Europe", in turn, tackles both current European issues like terrorism ("terror"), European parliamentary elections ("wählen" (vote), "parlament" (parliament)) and European integration as a wider context ("europäisch" + "union" (EU), "erweiterung" (enlargement), "mitgliedstaat" (member state)). This latter topic clearly evidences the role and status of the EU as the most important political, economic, and geographical context. 46

The topic "Covid-19 era" is a mixture of resignation and hope. The Austrian speech in 2021 well exemplifies this mixture: 47

Ein neues Jahr liegt vor uns. Wir spüren noch die Last des alten, die Last der Pandemie. Aber viele von uns spüren trotz allem eine hoffnungsfrohe Erwartung, wie sie nur am Beginn von etwas Neuem stehen kann, wenn alle Möglichkeiten offen und alle Träume noch frisch sind. (Austria 2021)

The German address, in turn, acknowledges the dramatic changes caused by the Covid-19 pandemic disease. At the same, however, the quotation has a positive trait, stressing the supportive and stabilising role of the federal state in the times of crisis: 48

Die Aufgaben, vor die die Pandemie uns stellt, bleiben gewaltig. Bei vielen Gewerbetreibenden, Arbeitnehmern, Solo-Selbstständigen und Künstlern herrschen Unsicherheit, ja Existenzangst. Die Bundesregierung hat sie in dieser ganz unverschuldeten Notlage nicht allein gelassen. Staatliche Unterstützung in nie dagewesener Höhe hilft. Verbesserte Kurzarbeitsregeln greifen. Arbeitsplätze können so bewahrt werden. (Germany 2021)

Contrary to Austria and Germany, the Swiss speech of 2021 is marked by negative ambience, even resignation underlining the deep societal impact the Covid-19 pandemic disease has caused in Switzerland: 49

Selten haben wir Vergleichbares erlebt: Unsere Tätigkeiten kamen zum Stillstand. Die ganze Gesellschaft befand sich in noch nie dagewesener Isolation. Wir mussten lernen, ohne Händeschütteln auszukommen. Dieses wichtige Begrüssungsritual gefährdete plötzlich unsere Gesundheit. (Switzerland 2021)

The topic "Europe" is mostly used, as the following quotations exemplify, to remind the citizens about the fundamental importance of European integration as a uniting community of Europeans, but also as provider for economic and political security in a global world: 50

Es geht nun darum, dass die Union nicht nur wirtschaftlich, sondern auch verstärkt politisch und emotional zusammenwächst. Die Bürger müssen spüren, dass es sich lohnt, in diesem erweiterten Europa zu leben und zu arbeiten. (Austria 2003)

In den vergangenen Jahren habe ich oft gesagt, dass es auch Deutschland auf Dauer nur dann gut geht, wenn es auch Europa gut geht. Denn nur in der Gemeinschaft der Europäischen Union können wir unsere Werte und Interessen behaupten und Frieden, Freiheit und Wohlstand sichern. (Germany 2020)

Wirtschaftlich sind wir weltweit verflochten. Geografisch, historisch und kulturell sind wir ein europäisches Land. Die Beziehungen zu unseren Nachbarländern und zu den anderen Mitgliedstaaten der Europäischen Union regeln wir in bewährter Weise bilateral. (Switzerland 2005)

Within the skipgram network, the topic "Faith in the future" represents a strong faith in technological and scientific development as the most important fundament for a better future. This is especially evident when we consider such content words like "technik" (technology), "wissenschaft" (science), "aufschwung" (boom, economic expansion), "selbstvertrauen" (self-confidence), and "zuversicht" (confidence). Interestingly, this topic also contains words with strong, dynamic connotation to the shaping the future ("gestaltung") and the putting oneself out ("bemühen"). How this topic is used to generate optimism towards the future is well illustrated by the following quotations: 51

Wie wichtig es ist, auch im scheinbar größten Durcheinander Gelassenheit, Mut und Zuversicht zu bewahren. (Austria 2020)

Zusammenhalt, Offenheit, unsere Demokratie und eine starke Wirtschaft, die dem Wohl aller dient: Das ist es, was mich für unsere Zukunft hier in Deutschland auch am Ende eines schweren Jahres zuversichtlich sein lässt. (Germany 2017)

Dennoch bin ich zuversichtlich. Denn ich glaube an die Fähigkeit des Menschen zu gestalten. Es gibt immer einen Weg. Manchmal braucht es Mut und Kraft. (Switzerland 2010)

Finally, the topic "Economy" is characterised by very concrete, economic content words like "arbeitsplatz" (job, post), "global" (global[isation]), "unternehmen" (company), and "wirtschaft" (economy). But this topic also has a forward-looking content embodied by words like "erreichen" (reach), "idee" (idea), or "leisten" (perform). An excellent example of how this topic is present in the speeches is the following quotation from the German speech of 2008: "Deutschland kann seine alte Kraft als das Land der Sozialen Marktwirtschaft wieder neu unter Beweis stellen, der Verbindung von Freiheit und Gerechtigkeit, Fleiß und Unternehmergeist". The same spirit can be found in the following Austrian and Swiss addresses: 52

> *Unsere Bevölkerung hat nicht nur ein Recht auf sichere Grenzen und innere Sicherheit, sie hat auch ein Recht auf ein möglichst hohes Maß an wirtschaftlicher Stabilität und sozialer Sicherheit.* (Austria 2013)

> *Starke Unternehmen sind der beste Garant für Arbeitsplätze. Und Arbeitsplätze sind die Basis für soziale Sicherheit und Wohlstand. Noch gehört die Schweiz zu den wirtschaftlich erfolgreichsten Ländern. Sorgen wir mit freiheitlichen Rahmenbedingungen dafür, dass das so bleibt.* (Switzerland 2016)

Overall, the topics identified by the Louvain method seem appropriate and reliable in the context of these three countries. The main topics discussed above not only connect well to the political and economic reality in these countries or in Europe in more general terms, but also evidence that TNA tools offer – as the previous studies of [Paranyushkin 2011], [Light 2014], or [Shim et al. 2015] already suggest – a reliable alternative to traditional methods of text mining. Further, the differences between topics identified in the bigram and the skipgram based network seem to give support to the study of [Jelveh et al. 2014] that the complementary use of both methods can help us to reliably identify topics that use similar vocabulary but show differences in the text structure. And finally, it is worth being noted that the most important content words of each topic contain only a few words with both a high degree centrality and a high betweenness centrality. This means that the most important content words are those enjoying higher relevance within the topic and, thus, are descriptive for the topic. A closer look at words used for meaning circulation across topics and countries reveals that these words include words related to peace ("frieden", "friedlich") and war ("krieg"), security ("sicherheit"), crisis ("krise"), community ("gemeinschaft"), economy ("wirtschaft"), and future ("zukunft"). All these "glueing" words are less topical and more contextual, thus framing and embedding the topical issues into a wider context of European (integration) history. 53

5 Discussion

This article has sought to exemplify the utility of text network analysis as a computational method for exploratory text and content analysis. Further, in the leaning on similar previous studies, this article also understands TNA as an promising and powerful alternative to traditional methods of topic modelling, especially to LDA. Against the background of these objectives and based on the results of the empirical analysis the following three conclusions are worth being discussed. 54

First, one major objective of this article was to present and evidence the usefulness of computational, exploratory text analysis for gaining new insights of the structure and dynamics of a collection of short textual documents. In this respect, the results are promising. Especially I would like to highlight the possibilities of using network clustering methods for the exploration of topics hidden in the corpus. As the result indicate, network clustering algorithms can identify contextually meaningful and relevant clusters. As I have shown the applied approach can help the researcher to identify appropriate text passages for different topics, serving as a starting point for an in-depth context analysis. 55

Second, results from the empirical analysis evidence the analytical power of the two computational methods applied. On the one hand, the rather traditional *tf-idf* method helped to gain insights and a better understanding of changes over time and of differences between the three countries. By increasing the weight of words that are more common in a certain period I could tackle changes in the content of the speeches over time. On the other hand, TNA proved to be a rather simple, yet powerful alternative to traditional topic modelling methods. The validation of the results of the community detection against samples from the original sources showed that the identified clusters not only had a meaningful content relevant for all countries, but they also were, according to their modularity values, valid and reliable. The two methods, when applied in a complementary manner, were very useful in identifying changes in the content over time and across nations, but also in exploring topics binding these three countries together. Here we could also identify one clear difference between the three countries. For Germany and Austria Europe seem to play an important role and many speeches reflected the fundamental importance of Europe, especially the European Union, as *the* political, economic, and historical context for both countries. Swiss New Year's speeches remain somewhat abstract and globally oriented in this respect, thus reflecting the long tradition of the neutral Swiss foreign policy [Gabriel and Fischer 2003] [Goetschel and Schwarz 2005] 56

Third, although the results are encouraging both in empirical and methodological sense, the limitations of text network analysis need to be addressed. As Diesner et al. (2012) points out, validation of the results can be difficult for densely connected large-scale networks. Further, techniques for text preprocessing, node identification, and link construction must be decided before mining network structure from text data, since these decisions “could strongly influence the structure of resulting networks” [Shim et al. 2015, 75]. As a comparison of the bigram and skipgram networks evidenced, different methods of network data creation may produce different results, so that data preprocessing and network analysis techniques should be selected with care and be closely aligned with research questions and objectives. 57

Overall, the analysis presented in this article is well in line with these previous works and supportive for the idea that text network analysis could offer an interesting alternative method for computational content analysis. Since the method seems to work quite reliable also with smaller data sets I can only encourage colleagues interested in this kind of analysis to test TNA tools. 58

Notes

[1] For a good overview of different methods for content mining, see e.g. [Aijmer and Stenström 2004] [Lee and Kim 2010] [Stuart and Botella 2009] [Weiss et al. 2010] [Diesner and Tambayong 2012] [Leetaru 2012] [Ignatow and Mihalcea 2016] [Jacobs and Tschötschel 2019]

[2] Switzerland is in fact a multilingual country with four official languages, German included.

[3] For a similar design, see [Shim et al. 2015] who apply text network analysis on nuclear policy documents from six countries in order to analyse how the Fukushima accident affected nuclear energy policy frames.

[4] It should be noted that in Germany also the Federal President (*Bundespräsident*) gives a speech on Christmas. This Christmas speech is, what comes to its function as a political speech act, very similar to the New Year’s speech. Since the purpose of this article is to compare New Year’s addresses, German Christmas speeches are not included in this study.

[5] The package is available and documented at:https://bnosac.github.io/udpipe/docs/doc2.html. For a technical and scientific description of the package, see Straková 2017.

[6] This technique was adopted from Jelveh et al. (2014).

[7] For details, see [Leydesdorff 2005]

[8] For a summary, see [Fortunato 2010] [Wu et al. 2013]

[9] References to speeches can be located at the following: Austrian New Years speeches 2000-2021. Available at the website of the Federal President of Austria, German New Years speeches 2000-2021. Available at the website of the German Chancellor's Office, Swiss New Years speeches 2000-2021. Available at the website of the Swiss Federal Council

Works Cited

Aijmer and Stenström 2004 Aijmer, K. and A.-B. Stenström (Eds.) (2004). *Discourse Patterns in Spoken and Written Corpora*. Amsterdam: John Benjamins Publishing.

Allen 2005 Allen, C. S. (2005). “Ideas, Institutions and Organized Captitalism: Germany, Europe and 21st Century Economic Policy Models.” *Conference Papers – American Political Science Association*, 1–25.

Anderson 1991 Anderson, B. (1991). *Imagined communities: reflections on the origin and spread of nationalism* (Rev. and extended ed. ed.). London: Verso.

Blei et al. 2003 Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). “Latent Dirichlet allocation.” *Journal of machine learning research 3*(4-5), 993–1022.

Blondel et al. 2008 Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). “Fast unfolding of communities in large networks.” *Journal of Statistical Mechanics: Theory and Experiment* (P10008).

Bulmer 2018 Bulmer, S. (2018). “Germany and the european union: Post-brexit hegemon?” *Insight Turkey 20*(3).

Burt 2004 Burt, R. (2004). “Structural Holes and Good Ideas.” *American Journal of Sociology 110*(2), 349–399.

Busse 2013 Busse, D. (2013). “Linguistische Diskurssemantik: Rückschau und Erläuterung nach 30 Jahren.” In D. Busse and W. Teubert (Eds.), *Linguistische Diskursanalyse: neue Perspektiven*, pp. 31–53. Wiesbaden: Springer VS.

Diesner and Tambayong 2012 Diesner, J., K. M. Carley, and L. Tambayong (2012). “Extracting socio-cultural networks of the Sudan from open-source, large-scale text data.” *Computational and Mathematical Organization Theory 12*(3), 328–339.

Fortunato 2010 Fortunato, S. (2010). “Community detection in graphs.” *Physics Reports* (486), 75–174.

Gabriel and Fischer 2003 Gabriel, J. M. and T. Fischer (Eds.) (2003). *Swiss Foreign Policy, 1945-2002*. Houndmills: Palgrave Macmillan.

Goetschel and Schwarz 2005 Goetschel, L., M. Bernath, and D. Schwarz (2005). *Swiss Foreign Policy: Foundations and Possibilities*. London: Routledge.

Hasu and Kao 2013 Hsu, Y.-Y. and H.-Y. Kao (2013). “CoIN: a network analysis for document triage.” *Database*, 1–11.

Hearst 1994 Hearst, M. A. (1994). “Multi-Paragraph Segmentation of Expository Text.” *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL 1994)*.

Heikkinen 2006 Heikkinen, V. (2006). “Uudenvuodenpuheiden piirteitä 1935–2006 ja näkymiä vallan medioitumiseen.” In T. Nikko and P. Pälli (Eds.), *Kieli ja teknologia. Talous ja kieli IV*, Number B76 in Publications of the Helsinki School of Economics, pp. 173–194. Helsinki: Helsinki School of Economy.

Ignatow and Mihalcea 2016 Ignatow, G. and R. F. Mihalcea (2016). *Text Mining: A Guidebook for the Social Sciences*. London: SAGE.

Jacobs and Tschötschel 2019 Jacobs, T. and R. Tschötschel (2019). “Topic models meet discourse analysis: a quantitative tool for a qualitative approach.” *International Journal of Social Research Methodology 22*(5), 469 – 485.

Jelveh et al. 2014 Jelveh, Z., B. Kogut, and S. Naidu (2014). “Detecting Latent Ideology in Expert Text: Evidence From Academic Papers in Economics.” In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1804–1809.

Jordan 2005 Jordan, P. (2005). *Großgliederung Europas nach kulturräumlichen Kriterien. Europa Regional 13*(4), 162–173.

Jsnning and Möller 2016 Janning, J. and A. Möller (2016). “Leading from the centre: Germany’s new role in Europe.” ECFR Policy Brief, European Council on Foreign Relations.

Koller 2014 Koller, V. (2014). “From text to political positions via discourse analysis: Introduction.” In B. Kaal, I. Maks, and A. van Elfrinkhof (Eds.), *From Text to Political Positions. Text Analysis across Disciplines*, pp. 163–169. Amsterdam: John Benjamins Publishing.

Kundnani 2016 Kundnani, H. (2016). *German Power. Das Paradox der deutschen Stärke*. München: C.H.Beck.

Lee and Kim 2010 Lee, S., J. Song, and Y. Kim (2010). “An Empirical Comparison of Four Text Mining Methods.” *Journal of Computer Information Systems 51*(1), 1–10.

Leetaru 2012 Leetaru, K. H. (2012). *Data mining methods for the content analysts.* New York: Routledge.

Leydesdorff 2005 Leydesdorff, L. (2005). “Similarity measures, author cocitation analysis, and information theory.” *Journal of the American Society for Information Science & Technology 56*(7), 769–772.

Light 2014 Light, R. (2014). “From words to networks and back: Digital text, computational social science, and the case of presidential inaugural addresses.” *Social Currents 1*(2), 111–129.

Maull 2018 Maull, H. W. (2018). Reflective, Hegemonic, Geo-economic, Civilian...? The Puzzle of German Power. *German Politics 27*(4), 460–478.

Müller-Brandeck-Bocquet 2010 Müller-Brandeck-Bocquet, G. (2010). “Deutsche Europapolitik unter Angela Merkel: Enge Gestaltungsspielräume in Krisenzeiten.” In G. Müller-Brandeck-Bocquet, C. Schukraft, N. Leuchtweis, and U. Kessler (Eds.), *Deutsche Europapolitik: Von Adenauer bis Merkel*, pp. 253–349. Wiesbaden: VS Verlag für Sozialwissenschaften. 2., aktualisierte und erweiterte Auflage.

Palonen 2007a Palonen, K. (2007a). *Four Times of Politics: Policy, Polity, Politicking and Politicization*, In: K. Palonen (2007b), pp. 55–70.

Palonen 2007b Palonen, K. (2007b). *Re-thinking Politics. Essays from a Quarter-century*. Jyväskylä: University Printing House.

Paranyushkin 2011 Paranyushkin, D. (2011). “Identifying the Pathways for Meaning Circulation using Text Network Analysis.” Research report, Nodus Labs.

Portman 2014 Portman, A. (2014). *From Might to Mandata: Values in Rulers’ Public Speeches in Finland 1809-2000*. Publications of the Department of Social Research of the University of Helsinki 2014:4. Helsinki: Unigrafia.

Prell 2012 Prell, C. (2012). *Social Network Analysis: History, theory and methodology*. London: SAGE.

Purhonen and Toikka 2016 Purhonen, S. and A. Toikka (2016). “'Big Data' haaste ja uudet laskennalliset tekstiaineistojen analyysimenetelmät. Esimerkkitapauksena aihemallianalyysi tasavallan presidentin uudenvuodenpuheista 1935-2015.” *Sosiologia 53*(1), 6–27.

Ruohonen 2013 Ruohonen, K. (2013). “Graph theory.” Technical report, Tampere University of Technology.

Scott 2013 Scott, J. P. (2013). *Social Network Analysis* (3 ed.). London: Sage Publishing.

Shim et al. 2015 Shim, J., C. Park, and M. Wilding (2015). *Identifying policy frames through semantic network analysis: an examination of nuclear energy policy across six countries. Policy Sciences 48*(1), 51–83.

Straková 2017 Straková, M. S. A. (2017). “Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe.” Technical report, Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Prague.

Stuart and Botella 2009 Stuart, K. and A. Botella (2009). “Corpus Linguistics, Network Analysis and Co-occurrence Matrices. ”*International Journal of English Studies* (S1), 1–28.

Van Dijk 1977 van Dijk, T. A. (1977). *Text and context*. London: Longman.

Weiss et al. 2010 Weiss, S. M., N. Indurkhya, and T. Zhang (2010). *Fundamentals of Predictive Text Mining*. London: Springer.

Wu et al. 2013 Wu, Z., J. Cao, J. Wu, Y. Wang, and C. Liu (2013). “Detecting Genuine Communities from Large-Scale Social Networks: A Pattern-Based Method.” *The Computer Journal*, 1–15.