Proceedings of ParlaCLARIN III @ LREC2022 pages 56–60
Marseille, 20 June 2022

Visualizing Parliamentary Speeches as Networks: The DYLEN Tool

Seung-bin Yim[1], Katharina Wünsche[1], Asil Cetin[1], Julia Neidhardt[2], Andreas Baumann[3], Tanja Wissik[1]
[1]Austrian Academy of Sciences,[2] TU Wien,[3] University of Vienna
Vienna, Austria
[1]{seung-bin.yim, katharina.wuensche, asil.cetin tanja.wissik}@oeaw.ac.at
[2]julia.neidhardt@tuwien.ac.at, [3]andreas.baumann@univie.ac.at

Abstract

In this paper, we present a web based interactive visualization tool for lexical networks based on the utterances of Austrian Members of Parliament. The tool is designed to compare two networks in parallel and is composed of graph visualization, node-metrics comparison and time-series comparison components that are interconnected with each other.

Keywords: parliament data, lexical networks, network visualisation, diachronic comparison

1. Introduction

Analyzing and visualizing the dynamics of language change is of interest to a broad field of studies including linguistics, natural language processing, digital humanities (DH) and computer science. In the project Diachronic Dynamics of Lexical Networks (DYLEN), we investigated lexico-semantic change based on two large corpora and how this change can be measured and visualized (Baumann et al., 2019). In this paper we will describe one component of the DYLEN tool[1] [2] that was developed within the DYLEN project, which visualizes lexical networks based on the utterances of the Austrian Members of Parliament (MPs) between 1996 and 2017 and enables diachronic comparison. The development of a new tool was necessary because no out of the box network-visualisation tool was specifically designed to allow diachronic network comparison and supported all the functions which were identified by potential users during a dedicated workshop. The tool adds a web based interactive visualization tool to the digital humanities toolbox that can be used to explore and compare different lexical networks over time and visualize related data.

2. Related Work

There are two common approaches for visualizing lexical networks: word clouds and graph-based visualizations. While word clouds (or collocation clouds) as for example described by Beavan (Beavan, 2008), Rayson (Rayson, 2008)), Heimerl et al. (Heimerl et al., 2014) and Xu et al. (Xu et al., 2016) display collocated words as a sorted list and represent statistical measures such as frequency or Mutual Information via a word's font size, color or brightness, graph-based visualizations usually show network graphs with a node for each word and edges representing collocations and/or semantic associations. Additional statistical information such as word frequency or the Mutual Information of two words can be encoded via the edge thickness or node size, whereas color is often used to represent a word's part-of-speech tag (cf. (Laußmann et al., 2011); (Lee and Jhang, 2013); (Brezina et al., 2015)). For the visualization of even more information like time series or multiple statistical values at once, (Rayson et al., 2017) propose multidimensional visualizations that combine multiple visualization components such as line charts, network graphs or tables. Network visualizations are not the only way of presenting linguistic, and in particular parliamentary, data: journalists of the German news magazine Zeit Online combine line charts, video snippets, pictograms and textual annotations in an online tool that allows users to analyze German parliamentary speeches from 1949 to 2019 (Zeit Online, 2019). For browsing and searching German political speeches from 1990 onward, (Barbaresi, 2018) provides the tool politische-reden.eu that offers basic visualizations for word-frequency distributions as well as the option to read full speeches containing a selected word.

3. Data

For analyzing and visualizing the lexical repertoire of MPs in the Austrian parliament in the DYLEN tool we used lexical networks (Marakasova et al., 2022). The constructed networks are based on the ParlAT Corpus, containing the proceedings of the Austrian Parliament from 1996 to 2017 (Wissik and Pirker, 2018). The ParlAT data was split into 376 subcorpora, containing all utterances of a single MP, out of which networks were constructed. Not all the MPs had enough utterance data to construct networks. Thus, not all MPs are searchable in the tool. Here, nodes represent lexical words (nouns, proper nouns, verbs and adjectives). Two nodes are

[1]The DYLEN tool is available at https://dylen-tool.acdh.oeaw.ac.at/

[2]Source code is available at https://github.com/acdh-oeaw/dylen-tool and https://github.com/acdh-oeaw/dylen-backend

linked if they share similar contexts (based on pairwise similarities between the respective word embeddings). For the general networks of parties, the subcorpora of grouped MPs according to their party affiliation were created. In such a general network nodes stand for the most frequent words that constitute the selected corpus of an MP or party.

4. Use Case and Tool Description

The objective of the DYLEN tool was to support diachronic comparison of lexical networks. In the following, we will present a use case for the application of the tool, describe the tool development process and its components.

4.1. Use Case

Does the general content of MPs' speeches remain stable or does it change over time? Quantitative analysis on different levels have shown that changes in governing coalitions have an impact on the lexical usage of parties in the Austrian parliament (e.g. (Hofmann et al., 2020), (Kern et al., 2021)). Can these dynamics also be seen via the visualization and comparison of lexical networks of single MPs? As an example we have chosen an MP of the SPÖ (Social Democratic Party of Austria) who was MP from 1983 to 2017: Dr. Josef Cap. The available data visualisations are covering the period from 1996 to 2016. We will look in more details at the network of the year 2003, where SPÖ was in opposition and at the network of the year 2014, where the SPÖ was in governing coalition with the ÖVP (Austrian People's Party). In the chosen time period, the MP had similar roles, in 2003 he was chairman of the parliamentary group and in 2014 he was deputy chairman of the parliamentary group.

4.2. Tool Development Process

The features of the tool were derived from user stories identified in a workshop that was conducted at the beginning of the project (Knoll et al., 2020). The dual track agile approach (Sedano et al., 2020) was followed for the development of the tool, two usability test rounds were conducted to improve the user experience. Vue.js and d3.js were used for the frontend, a Java Spring boot backend service was developed and made available for query via GraphQL interface.

4.3. Tool Components

When the user selects the network to explore and clicks on the submit button, a dashboard appears with three visualization components (Fig. 1), namely network-graph visualization, node-metrics comparison and time-series components. The network-graph visualization component visualizes networks of words used by a certain MP or party. The node-metrics comparison component lets users compare different networks based on their metrics [3] (e.g., degree centrality or betweenness centrality). The time-series component shows the change in difference in similarity measures relative to specific years (first, last, previous). The visualizations can be used to explore different aspects of the lexical networks. Each of the components were designed to let users compare two networks in parallel. In this paper, we will focus on the description of the network graph visualization and the node-metrics comparison based on the use case example.

4.3.1. Network Graph Visualization

Network-graph visualizations consist of nodes and edges that connect the nodes. Different information can be encoded using the size, thickness, colors of nodes and edges. In the DYLEN tool, nodes are most frequently occurring words within the selected corpus (i.e. single MP or party), and two nodes are connected by an edge if they share similar contexts. The similarity between the nodes are calculated by first training word embeddings with skip-gram model, then applying pairwise cosine similarity function. The size of nodes are determined by word frequency, while the thickness of edges represent contextual/semantic similarity of the words. The colors of the labels represent different part-of-speech tags.

A major aspect to consider when visualizing network data is the size of the network and visualization performance. Theoretically, a lexical network can have as many nodes as the size of the vocabulary and all the connections between them, such large graphs can be hard to explore and inefficient to visualize. Different graph filtering algorithms exist to tackle these issues (Hu and Lau, 2013), which can be divided into two groups, stochastic and deterministic graph filtering algorithms (Von Landesberger et al., 2011). Early user tests have shown that the lexical networks of parliamentary data are too large for visualization. The tool provides a deterministic graph filter based on eight different network metrics, such as betweenness centrality. In addition to the graph filter, a timeout mechanism is implemented to prevent users from unpleasant user experience when they select a large range of filter values. Another problem that makes interpretation and exploration of large graph visualizations difficult is edge cluttering. To account for this issue, the tool provides an option to change the visibility of edges with a slider based on similarity values alongside with Pan, Zoom in/out functionalities. Since the tool is designed for diachronic comparison, it allows to visualize two networks at the same time. There is a year slider to select a specific year to be visualized. For example, the lexical networks of Dr. Josef Cap in two different years are quite different in size. Based on this, one could hypothesize that speakers of the opposition party

[3] Detailed descriptions about the metrics can be found on the *Technical Details* page of *Node Metrics Comparison* tab at the start page of the tool

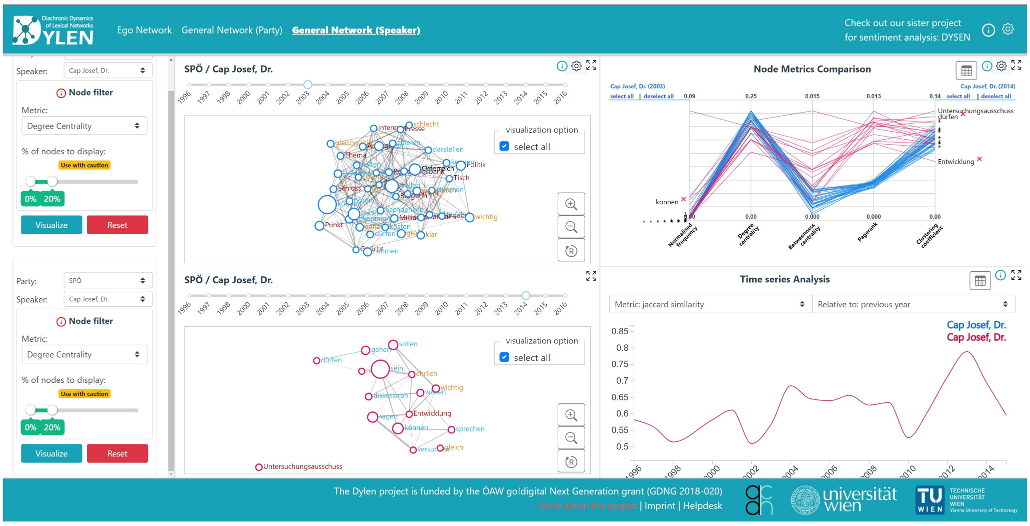

Figure 1: Dashboard view with all components.

might be more active in the parliament. Hypotheses like this can be then tested with a larger number of individuals. As demonstrated with the use case described above, network visualization can be useful for comparison of networks in their size and complexity over time, and for deriving hypotheses based on analyses like this.

4.3.2. Node Metrics Comparison

The node-metrics comparison component can be used to explore the network topology. For example, different centrality measures can be used to extract key phrases (Boudin, 2013). Parallel coordinates are used as the visualization technique. The visualization consists of n parallel axes, where n is the number of node metrics to be visualized. The chart allows visualization of all nodes of a network in one chart and facilitates the comparison of node metrics. Parallel coordinates can be used for identification of outliers, trends or clusters (Johansson and Forsell, 2015). Some challenges had to be addressed to increase the usability of the visualization. One issue was the labels of the nodes. Lots of nodes had metrics values within a small range, simply displaying the full labels would lead to overlapping labels. This issue is addressed by replacing the labels by the asterisk (*) character to denote that multiple labels are overlapping. When the user hovers over the asterisk characters, the full label is displayed. Another way to deal with the issue is to reorder the axes and put the metrics with more even distribution to the most left and right axis, respectively. In addition to the parallel coordinates view, the node-metrics comparison component offers a table view, with which the user can select/filter/sort nodes. This is useful if the network is too large and has to be reduced in size for proper visualization. An example application of keyword extraction using centrality metrics for our use case is shown in Table 1. It is extracted by sorting the nodes according to the two centrality metrics that have been demonstrated to be useful for keyphrase extraction (Boudin, 2013) and taking the ranks of the nodes. The extracted keywords are fairly different between the networks in the years 2003 and 2014 and could be a subject of further analysis.

4.4. Interactivity Between Components

One of the tool's main strengths is the interactivity between the individual visualizations, i.e., the network graphs, the parallel coordinates plot as well as the table view, which supports the Visual Information-Seeking Mantra defined by (Shneiderman, 1996): "overview first, zoom and filter, then details on demand". While the graphs provide a first overview of the network structure and highlight the most frequent words based on their node sizes, users who are interested in the network metrics and their distribution can take a look at the parallel coordinates plot and highlight words by hovering over them in either the graph or the parallel coordinates visualization. Moreover, users have the option to select and deselect words by clicking on their labels or (un)checking them in the table view, which can reduce the complexity by filtering out words that are of minor interest.

4.5. Challenges and Limitations

There are some remaining challenges and limitations. One aspect is the size of the network and the performance of visualization. This could be addressed by either increasing the efficiency of the visualization itself by different algorithms/libraries/APIs or by reducing the number of nodes and edges with sophisticated filtering algorithms, such as edge filtering. Another challenge is the visualization of multidimensional data in the case of node metrics comparison. Parallel coordinates are useful when the user is interested in analyzing patterns or outliers, but presenting the node labels within a limited screen size remains a challenge.

Network	Top 10 Keywords (nouns) - Closeness Centrality	Top 10 Keywords (nouns) - Degree Centrality
Cap Josef, Dr. (2003)	Wahrheit, Situation, Österreicher, Kritik Abfangjäger, Teil, Tag, Art, Möglichkeit, Haider	Wahrheit, Situation, Österreicher, Kritik, Abfangjäger, Art, Möglichkeit, Haider, Entwicklung, Gegengeschäft
Cap Josef, Dr. (2014)	Regierung, Österreich, Weg, Möglichkeit Jahr, Land, Modell, Diskussion, Russland, Union	Regierung, Land, Modell, Österreich, Weg, Möglichkeit, Jahr, Diskussion, Russland, Union

Table 1: Keyword extraction using centrality metrics.

5. Future Work

The practical contribution of our work, the interactive visual analysis tool, has the potential to be used with other similar corpora for exploratory analysis in the future. Even though the scope of our current project focused on the Austrian Parliament data, the detailed requirement analysis conducted with the users (Knoll et al., 2020) and the successful technical implementation present a solid basis for future work. Further development in the database layer of the tool to import data in different structures and customization possibilities for visualization components will be the primary tasks to achieve this. Another functionality that could be useful is to add filters for each of the metrics axes and combine with brushing (Martin and Ward, 1995). A feature that was planned but not yet implemented is keyword-in-context, which could improve the value of the tool, since it was requested by multiple usability testers.

6. Conclusion

We have introduced the DYLEN Tool, a web-based interactive visualization tool for lexical networks of Austrian Parliament proceedings. The tool can be used to explore and analyze lexical networks via three different visualization components, such as graph visualization, parallel coordinates and time series visualization and demonstrated to be useful for deriving hypotheses and gaining an overview of network topologies. However, some challenges remain such as performance issues related to the size of the networks and more effective visualization for multidimensional data for comparing node metrics.

7. Acknowledgements

This research was funded by the ÖAW go!digital Next Generation grant (GDNG 2018-02).

8. Bibliographical References

Barbaresi, A. (2018). A corpus of German political speeches from the 21st century. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), pages 792–797, Miyazaki, Japan. ELRA.

Baumann, A., Neidhardt, J., and Wissik, T. (2019). Dylen: Diachronic dynamics of lexical networks. In *LDK (Posters)*, pages 24–28.

Beavan, D. (2008). Glimpses through the clouds: collocates in a new light, June.

Boudin, F. (2013). A comparison of centrality measures for graph-based keyphrase extraction. In *International joint conference on natural language processing (IJCNLP)*, pages 834–838.

Brezina, V., McEnery, T., and Wattam, S. (2015). Collocations in context:a new perspective on collocation networks.

Heimerl, F., Lohmann, S., Lange, S., and Ertl, T. (2014). Word Cloud Explorer: Text Analytics Based on Word Clouds. In *2014 47th Hawaii International Conference on System Sciences*, pages 1833–1842, January. ISSN: 1530-1605.

Hofmann, K., Marakasova, A., Baumann, A., Neidhardt, J., and Wissik, T. (2020). Comparing lexical usage in political discourse across diachronic corpora. In *Proceedings of the ParlaCLARIN II Workshop*, pages 58–65.

Hu, P. and Lau, W. C. (2013). A survey and taxonomy of graph sampling. *arXiv preprint arXiv:1308.5865*.

Johansson, J. and Forsell, C. (2015). Evaluation of parallel coordinates: Overview, categorization and guidelines for future research. *IEEE transactions on visualization and computer graphics*, 22(1):579–588.

Kern, B. M., Hofmann, K., Baumann, A., and Wissik, T. (2021). Komparative zeitreihenanalyse der lexikalischen stabilität und emotion in österreichischen korpusdaten. *Proceedings of Digital Lexis and beyond at OELT*.

Knoll, C., Cetin, A., Möller, T., and Meyer, M. (2020). Extending recommendations for creative visualization-opportunities workshops. In *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*, pages 81–88. IEEE.

Laußmann, J., Lux, M., Menßen, C., and Mehler, A. (2011). An online platform for visualizing lexical networks. In *Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on*, volume 1, pages 495–496, Los Alamitos, CA, USA, aug. IEEE Computer Society.

Lee, S.-M. and Jhang, S.-E. (2013). Visualization of Collocational Networks: Maritime English Keywords.

Martin, A. R. and Ward, M. O. (1995). High dimensional brushing for interactive exploration of multivariate data. In *Visualization Conference, IEEE*, pages 271–271. IEEE Computer Society.

Rayson, P. E., Mariani, J. A., Anderson-Cooper, B., Baron, A., Gullick, D. S., Moore, A., and Wattam, S. (2017). Towards interactive multidimensional visualisations for corpus linguistics. *Journal for language technology and computational linguistics*, 31(1):27–49.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4):519–549, January.

Sedano, T., Ralph, P., and Peraire, C. (2020). Dual-Track Development. *IEEE Software*, 37(6):58–64, November.

Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, VL '96, page 336, USA, September. IEEE Computer Society.

Von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J. J., Fekete, J.-D., and Fellner, D. W. (2011). Visual analysis of large graphs: state-of-the-art and future research challenges. In *Computer graphics forum*, volume 30, pages 1719–1749. Wiley Online Library.

Xu, J., Tao, Y., and Lin, H. (2016). Semantic word cloud generation based on word embeddings. In *2016 IEEE Pacific Visualization Symposium (PacificVis)*, pages 239–243, April. ISSN: 2165-8773.

Zeit Online. (2019). Darüber spricht der bundestag.

9. Language Resource References

Marakasova, A., Neidhardt, J., Baumann, A., and Wissik, T. (2022). Dylen: Diachronic dynamics of lexical networks.

Wissik, T. and Pirker, H. (2018). Parlat beta corpus of austrian parliamentary records. In *Proceedings of the LREC 2018 Workshop'ParlaCLARIN*, pages 20–23.