Further Developing the Word Rain Text Visualisation Technique in a Digital History Project

Maria Skeppstedt[1,*], Magnus Ahltorp[2], Gijs Aangenendt[1,3] and Ylva Söderfeldt[3]

[1]*Centre for Digital Humanities and Social Sciences Uppsala, Department of ALM, Uppsala University*
[2]*Language Council of Sweden, Institute for Language and Folklore*
[3]*Department of History of Science and Ideas, Uppsala University*

Abstract

The development of the Word Rain text visualisation technique started as a theoretical project, in which insights from previous NLP and text visualisation research were used for creating a visualisation that addressed some of the problems associated with the classic word cloud. In particular, the problem of the word clouds being unsuitable for the analytical tasks to which they are often applied within digital humanities – such as categorising words and comparing texts – was addressed. The next step in the development of the Word Rain visualisation technique consisted of applying the technique to real use cases. One of these use cases studied longitudinal content change in periodicals from diabetes patient organisations. This work, which was carried out in a (partly digital) history project, resulted in several features being added to the Word Rain visualisation, e.g., features related to word positioning, colour use and word prominence calculations. We here describe this work of practically evaluating and further developing the Word Rain text visualisation technique, as well as present and motivate the new features of the visualisation technique that the work resulted in.

Keywords

Text visualisation, Digital history, Word embeddings

1. Introduction

The classic word cloud visualisation is often used within digital humanities as a tool for analysing and comparing texts, despite the technique not being suitable for these tasks (Hicke, Goenka, and Alexander 2022). We have therefore created the Word Rain text visualisation technique, which aims to retain the simplicity of the classic word cloud, while also supporting typical analytical text exploration tasks (Skeppstedt et al. 2024). These typical tasks include the categorisation of words, the exploration of interesting areas of the graph, and the comparison of pairs of graphs that visualise different texts.

The first version of the Word Rain technique was theoretical, in the sense that it was based on natural language processing (NLP) and text visualisation research, without any connection to a specific real-world application. Instead, we envisioned three different example use case scenarios. These consisted of the creation of word rain visualisations for i) comparing texts

The 9th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2025), March 5–7, 2025, Tartu, Estonia.
*Corresponding author.
maria.skeppstedt@abm.uu.se (M. Skeppstedt); magnus.ahltorp@isof.se (M. Ahltorp); gijs.aangenendt@idehist.uu.se (G. Aangenendt); ylva.soderfeldt@idehist.uu.se (Y. Söderfeldt)
© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
Digital Humanities in the Nordic and Baltic Countries Publications – ISSN: 2704-1441

from different genres, ii) studying longitudinal content change in a temporal corpus, and iii) investigating the coverage of a specialised vocabulary in relation to a text and finding new word candidates to include in the vocabulary. For all three use case scenarios, we used texts and vocabulary related to climate change. More specifically, we contrasted IPCC reports to a report on thought structures related to delay in climate change mitigation, studied longitudinal content change in Nature and Science editorials on the topic of climate change, and visualised the coverage of the climate change glossary GEMET in relation to IPCC reports. We also carried out a user study, in order to investigate the usability and usefulness of the Word Rain visualisation. This study resulted in a number of useful suggestions for how to further improve, use and describe the Word Rain visualisation (Skeppstedt et al. 2024).

However, as a next step, to find further suggestions for improvements and features to add, we applied the visualisation on real use cases from ongoing research and development projects. One of them consisted of working with dictionary developers, to study the coverage of a Yiddish dictionary and to find candidates for new words to add (Ahltorp et al. 2025). The other one, which is the work we describe here, consisted of applying the Word Rain visualisation in a (partly digital) history project to study the longitudinal content change in periodicals from diabetes patient organisations.

We will here describe the additions to the Word Rain visualisation technique that emerged from the use case. That is, the aim here is not to describe the results of the analysis of the word rains produced. Instead, we will report on the work where a visualisation method – developed on the basis of research within NLP and text visualisation – is applied practically in a digital humanities project, and what changes and additions to the original technique this resulted in.

Visualisation techniques used within digital humanities are often borrowed from other disciplines such as data science and the natural sciences without necessarily being adjusted to the research and source materials of the humanities (Drucker 2015). Developing the visualisation technique further within a digital humanities project gives the opportunity to align the technique with humanities methods and critically assess its usefulness for research within the humanities.

2. Background

The word cloud is a commonly used text visualisation technique (Brath and Banissi 2016), which is also employed for analytical tasks, for instance within digital humanities (Hicke, Goenka, and Alexander 2022). The aim of the technique is to provide a visual overview of a text by displaying its top n most prominent words. There are a number of services and programming packages for producing different types of word clouds.

The popularity of the word cloud has been attributed to the simplicity that comes with it being self-explanatory and not requiring any additional visual artifacts (Cao and Cui 2016). While there are a number of versions of the classic word cloud that build on interaction and/or animation (Collins, Viégas, and Wattenberg 2009; Koh et al. 2010; Lohmann et al. 2012; Liu, Shen, and Hu 2015; Wang et al. 2018; Xie et al. 2024), we aimed to create a visualisation that would still consist of a graph in the form of a static image, and as far as possible retain the simplicity of the classic word cloud.

Previous word cloud developments that take the approach of producing a static image tend to

focus on only one of the problems associated with the classic word cloud, that is, the problems that make them unsuitable to use for analytical tasks within digital humanities (Hicke, Goenka, and Alexander 2022). The problem of the words lacking a meaningful positioning is typically addressed by retaining a layout similar to the classic word cloud, but positioning words that have a similar meaning closer to each other in the graph, and emphasizing their similarity with a colour coding (Wu et al. 2011; Barth, Kobourov, and Pupyrev 2014; Xu, Tao, and Lin 2016). The difficulty of comparing word clouds is addressed in other studies, with series of word clouds where the words always receive the same position (Herold et al. 2019; Cui et al. 2010), or with word clouds that simultaneously visualise several texts (Diakopoulos et al. 2015; Burch et al. 2014; Lee et al. 2010). We are not aware of any previous approaches that – as the Word Rain – aim to solve both these problems, while also retaining the simplicity of the static word cloud.

The most important novelty of the Word Rain is that it – instead of applying the random or alphabetical word positioning of the classic word cloud – utilises both the vertical and horizontal word position for conveying information. The horizontal word positioning is determined by the meaning of the word, in the sense that similar words are positioned close to each other on the x-axis. "Similar words" here means words that often occur in similar contexts and that could typically be substituted for each other in a sentence, i.e., the type of word similarity that is often referred to as *paradigmatic* similarity. (The pair "tea" and "coffee" forms an example of paradigmatically similar words, which, e.g., can be substituted for each other in the sentence "In the morning I drink [coffee]". The words "coffee" and "drink" – which occur together in the same sentence – instead forms an example of *syntagmatic* similarity.) The semantic word positioning of the Word Rain is achieved by using a word embedding model. Each word that is to be visualised in a series of word rains is represented by a multidimensional word embedding vector, which encodes the distributional semantics properties of the word (Cuba Gyllensten 2023). These multidimensional vectors are projected onto one dimension, using t-SNE dimensionality reduction (Maaten and Hinton 2008), and the projected point determines the word's position on the x-axis.

This semantically ordered word positioning makes it possible to zoom in on interesting areas of the graph to read words displayed with a small font, based on the position of more prominent words that are semantically similar. Since all graphs in a visualisation series use the same t-SNE projection – i.e., a word receives the same semantically motivated position on the x-axis for all graphs in a series created – it is much easier to compare pairs of word rains than to compare two classic word clouds. Also, since the words are semantically ordered, similar words create clusters along the x-axis, and it is thereby easy to categorise words and compare them to others that are semantically close.

The vertical position is instead a reflection of the word's prominence. In general, the more prominent a word, the higher its position along the vertical axis. The meaning of "word prominence" is user-defined, but in the simplest case it is determined by the word frequency in the text visualised. That is, the more frequent the word, the more prominent it is. While the horizontal position for a word is determined beforehand for all graphs that are to be generated in a series of word rains, the vertical position is adapted to the other words included in the graph. Therefore, a word receives a vertical position in the graph that is a high as possible, as long as it does not collide with a more prominent word. If there is a collision, the less prominent

word “rains down” along its vertical position until there is a free spot in the graph. Similar to a classic word cloud, the prominence is also indicated by font size. In addition, the graphs include a row of vertical bars – one bar associated with each word – where a bar has a height corresponding to the prominence of the word it represents.

Figure 1 shows an example of two word rains generated for the work described here. The corpus used for these graphs consists of periodicals issued by the British Diabetes Association, between 1950 and 1990. A series of graphs – one graph for each year in the corpus – was generated, and Figure 1 shows the graphs for the years 1958 and 1959. The word rain series uses tf, term frequency, as the prominence measure, and each graph shows the top 600 most prominent words for the year (with stop words excluded). The corpus was lemmatised with spaCy, before the word rain series was generated. This lemmatised version of the corpus was also used for training the word embedding model used for positioning the words along the x-axis.

Vertical clusters of paradigmatically similar words are formed along the x-axis of the graphs in Figure 1. For instance, to the left (in red) there is a cluster of titles and names, a bit to the right (in green) there is a cluster of persons such as “patient”/“child”/“people”, then a bit further to the right (in blue) there is a cluster of words on nutrition and “(blood) sugar” such as “carbohydrate”/“calorie”, and finally to the very right (in red) there is a cluster of words denoting food. The semantic word positioning does not only facilitate a manual categorisation of the words, but also directs the user to potentially interesting areas in the graph. For example, many of the food items are shown in a font size too small to read without zooming in into the graphs, but the more prominent word “fruit” guides the user to the food area in the graph. The horizontal word positioning also guides the user when comparing the two graphs, and for the pair of word rains in Figure 1, it can be noted that the set of words that are most prominent does not change very much between the two adjacent years.

Figure 2, which instead compares 1958 to 1987, shows a larger difference between the texts. For instance, in the 1987 graph, the nutrition/blood sugar cluster puts more focus on blood sugar and sugar level tests, than on nutrition. Just to the left of that cluster, a new cluster (in blue) with medical words (such as “pain”/“virus”/“infection”) has emerged. Also the cluster to the right (in violet), with words relating to insulin, has changed, in that words concerned with the methods of administering the hormone have become more frequent, e.g., “syringe”, “needle” and “injection”. To the left in the graph (in yellow), the area which in 1958 contains the occupations “doctor” and “nurse”, has in 1987 expanded to a cluster that also contains “dietitian”/“specialist”/“consultant”.

The implementation of the Word Rain visualisation technique is provided as open-source Python code at: https://github.com/CDHUppsala/word-rain. There is also code[1] for running Word Rain as a web service (Ahltorp and Skeppstedt 2024), making it possible for users without programming skills to generate word rains. An instance of this web service is deployed at: https://wordrain.isof.se.

[1] https://github.com/sprakradet/wordrain-service

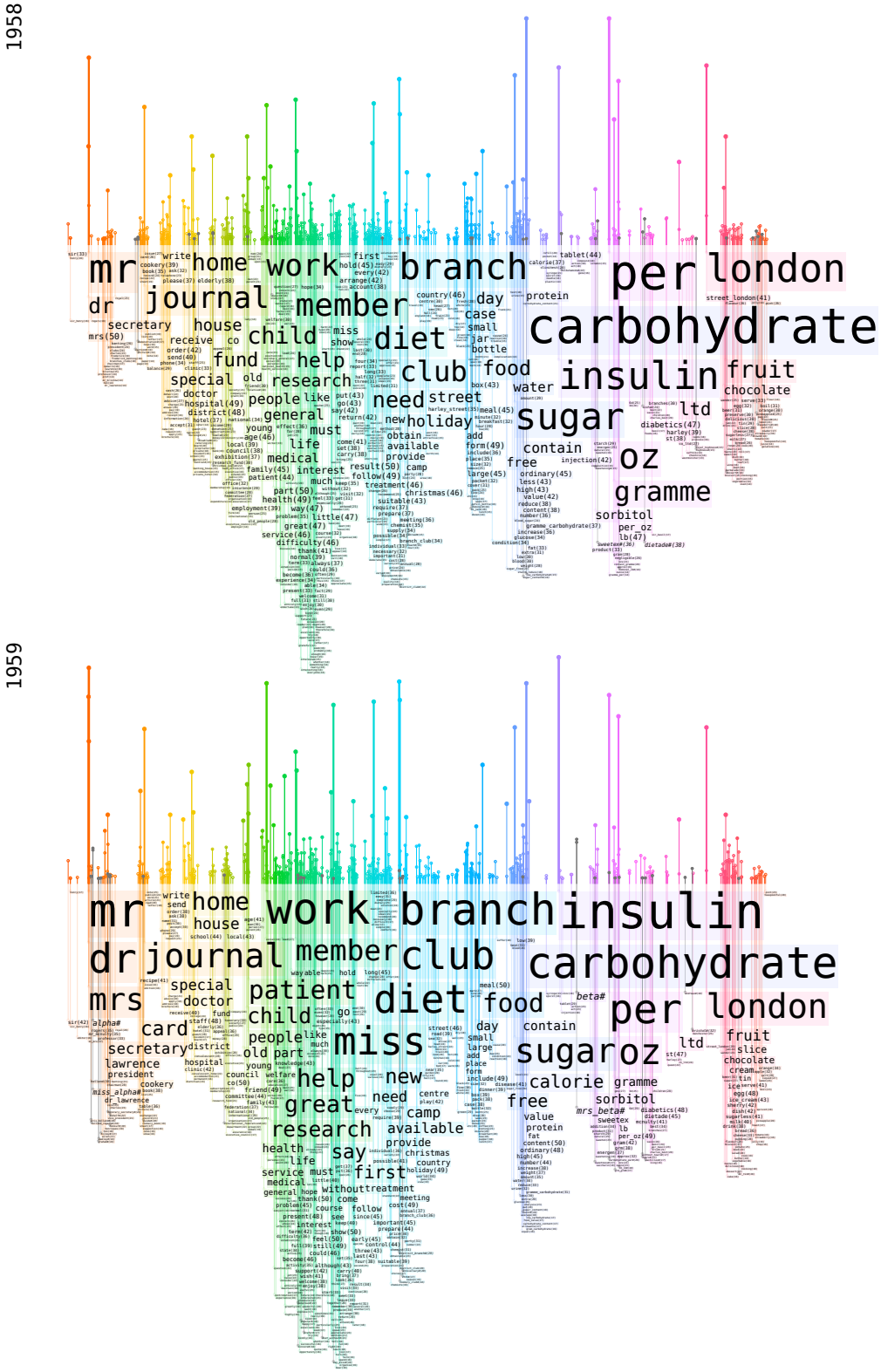

Figure 1: Two word rains generated for periodicals published by the British Diabetes Association (using term frequency as the prominence measure).

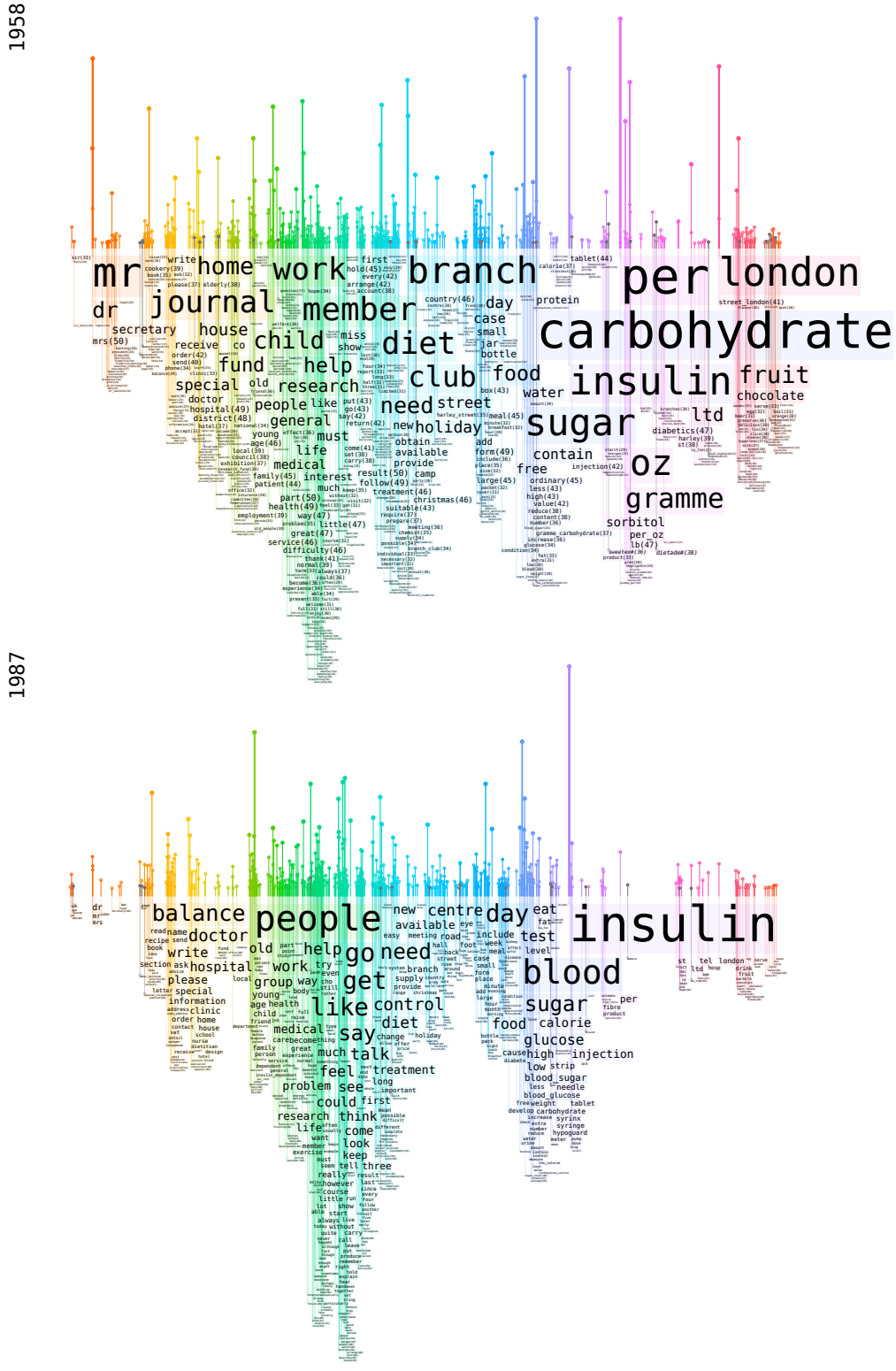

Figure 2: Two word rains generated for periodicals published by the British Diabetes Association (using term frequency as the prominence measure).

3. Method for evaluating and further developing the visualisation technique

The work described here of evaluating and further developing the Word Rain technique was carried out within the ActDisease[2] project. This project studies the development and impact of patient organisations and has as its main sources digitised versions of newsletters and magazines issued by about 10 different organisations in four countries.

The interdisciplinary nature of the project team – which consisted of historians and researchers/engineers with a background in NLP and digital humanities – made it possible to practically evaluate the usability of word rains produced using the corpora studied, and to implement suggestions for how to solve the problems detected.

For the work described here, we applied the Word Rain visualisation to three corpora consisting of periodicals from Swedish, German and British diabetes organisations (Aangenendt, Skeppstedt, and Söderfeldt 2024). In order to support analysis in this historical research project, we needed visualisations that could show longitudinal change. We therefore produced one visualisation for each year in the corpora, and studied how the visualisations changed over time.

Our previous user evaluation of the hypothetical use case of the Word Rain visualisation took place during a very limited time period (one hour for each of the three participants), and followed a formal structure, with both predefined task-specific questions posed to the users and standard usability questions. The work described here instead took place over the course of a year, and did not follow one particular structure. Instead, we produced many different word rain graphs we hypothesised would be useful, and discussed these within different constellations, mainly among the authors of this paper, but also with other researchers either involved in the development of the Word Rain visualisation or in the ActDisease project. The discussions could both take the form of sitting together discussing the word rains produced, and of writing and commenting on texts about the word rains. In addition, we also presented the Word Rain technique at seminars/workshops to receive feedback from external researchers with different backgrounds.

The discussions centred around the usefulness of the words rains as a tool for showing temporal content change and differences between corpora, but also around the technical design and implementation of the visualisation technique. Questions on how the word rains were produced were asked to better understand what the visualisations represent, leading to discussions on the selected preprocessing steps, the impact of the word embedding model used to project the words, and the meaning of proximity of the words appearing next to each other horizontally and vertically. Having regular meetings about the word rains also allowed for clarification of potential misunderstandings from the previous time and for follow-up questions after members of the project team had worked independently with a series of visualisations or new features were implemented.

[2] https://www.actdisease.org

4. Feature Additions

As a result of the application of the Word Rain text visualisation technique on the diabetes patient organisation corpora, a number of new features were implemented.

4.1. Creating more distinct semantic clusters

For the hypothetical use case scenarios used for the initial development of the Word Rain visualisation, texts with quite large content differences were compared. For instance, the temporal content change in Nature and Science editorials was studied using time chunks of multiple years and the scientific reports studied belonged to very different subfields within climate change research. These differences resulted in distinct clusters of words along the x-axes, which made it easy to compare the different time periods and the different types of scientific reports.

In contrast, in our study on diabetes patient organisation magazines, the content difference when comparing two years was not that large, particularly not when comparing word rains from two adjacent years. As a result, distinct clusters of semantically similar words were not always created, making it more difficult to discern word categories in graphs and to compare the difference in semantic content between pairs of word rains.

It should be noted that the fact that the word rains were able to show the similarity of the texts, is one of the core strengths of the visualisation technique. It is not necessarily the case that the exact same words are included in the texts over the time period studied. But the texts are similar in the sense that there are very often (or for some categories always) clusters of, e.g., person names, medical problems, food-related terms, cities, and so on. The exact content of these clusters might vary over the time period studied, but the high-level similarity of the texts – which would not be apparent in, e.g. a word frequency list – is visible thanks to the semantic positioning of the Word Rain visualisation. However, when we projected words belonging to semantic clusters that exist in all/many of the created graphs on the x-axis, the visual distinction between the different clusters were not always very clear, making it difficult to read the graphs.

In order to address this issue, we first attempted to vary the parameters of the scikit-learn implementation (Pedregosa et al. 2011) for the t-SNE dimensionality reduction used, in order to make it create more distinct clusters of words along the x-axis. However, these attempts did not result in the visually distinct clusters we were aiming for. We, therefore, instead used agglomerative clustering of the word embeddings representing the words to be visualised. This was added as a configurable feature option, where the user defines the number of semantic clusters they want the words to be divided into. As a first attempt, we let all the words in a cluster be represented by the centroid of the cluster, so that they all received the exact same position on the x-axis. This, however, totally obscured the semantic difference that exists between the words within a cluster. We, therefore, instead took the following approach: We moved the vectors representing the words halfway from their original position in the vector space towards the centroid of its cluster, to create a new vector representing the word. The t-SNE dimensionality reduction was then applied on these new vectors. This procedure resulted in the visual output we aimed for. Words belonging to the same cluster are closer to each other on the x-axis, making the clusters more visually distinct, but the semantic differences

between the words within the same cluster are still partly retained. For instance, in figures 1 and 2, the word rains have been configured to use 30 clusters. The figures thereby consist of around 30 vertical "stripes" that approximately correspond to how the words were clustered with agglomerative clustering. (Approximately 30, since the application of the t-SNE algorithm might combine several clusters into one, or create additional ones).

4.2. Updating colour scheme and colour use

As mentioned in the previous section, from the discussions, the need emerged to improve the readability of the word rains and specifically the need to emphasize which words belonged to the same cluster. Also the colour scheme and use were therefore adjusted to visually aid the user in discerning which words were more similar.

In a word rain, the plotted word starts at its assigned x-coordinate, i.e., its horizontal position determined by the t-SNE projection. (For a left-to-right writing system, this means the point just to the left of the word, and otherwise it means the point just to the right of the word.) However, the horizontal extension for some of the words is quite long, in particular for words that are displayed with a large font size and/or contain many letters. Therefore, the extension of a word might span several word clusters along the x-axis. This makes the word clusters less distinct, even with the addition of the agglomerative clustering described in the previous section. Our suggestion for solving this problem is to add a coloured background to the words, where the colour is decided by the word's assigned x-coordinate. Thereby, words belonging to the same cluster will have a similar background colour, even if their horizontal extension spans several cluster positions along the x-axis. In Figure 1, there are a number of words displayed with a large font, whose extension spans several clusters, e.g., "carbohydrate" and "insulin".

In the previous Word Rain version, the horizontal position of the words was emphasised through a colour gradient along the x-axis that spanned from cyan to magenta. We have now updated the default gradient, so that the full colour spectrum along the x-axis instead is used. The clusters are then separated by a larger colour spectrum, making the difference between clusters more evident. These changes also have the effect that we now use the same colour at the left-most position of the graph as we do at the right-most position. With these changes, we aim to decrease the risk of the user interpreting the x-axis as a scale of meaning, rather than as a one-dimensional projection of unordered semantic clusters.

4.3. A more flexible configuration option for the word prominence measure

Different word prominence measures can show different properties of the represented texts. For example, displaying the top n most frequent words suggests, at a glance, the main topics of the texts, whereas measuring frequency differences of words between different parts of the corpus – such as different time periods – gives an amplified view of change over time or other contrasts. When generating word rains on the diabetes patient organisation periodicals, we experimented with different configuration options for measuring word prominence to see if we can highlight vocabulary change or stability over time.

The prominence measures available in the first version of the Word Rain implementation were term frequency – *tf* – and term frequency-inverse document frequency – *tf-idf* – (a measure

that downweights words appearing in many documents). In addition to deciding which top n most prominent words to include in the word rain graph, the prominence measure also decides the word's prominence in the graph, i.e., its font size, the height of the corresponding vertical bar, and position on the vertical axis. When combining the two implemented word prominence measures with other configuration options available, such as cut-off limits for the maximum number of years in which a word could occur to be included in the visualisations, we were able to use the word rains for visualising longitudinal corpus change from a number of different angles, for example to study words that occur only a brief period of time, words that remain stable throughout the time period, and the introduction of new words and their development over time (Skeppstedt and Aangenendt 2024).

However, this method of combining tf or tf-idf with different cut-offs still did not capture all relevant aspects of longitudinal change in the patient organisation corpora. In particular, the most frequent words often did not change that much between two adjacent years in the corpora studied. In order to not only be able to visualise the similarity between two texts, we also needed prominence measures that would emphasise the differences between texts. Statistical measures commonly used in corpus linguistics (e.g., log-likelihood and odds ratio), is one example of prominence measures that can be used to emphasize differences. We were also introduced to the tf-bidf measure, which is a version of tf-idf adapted to corpora that stretches over a time period, and which gives more prominence to a word the less it has been used previously (La Mela, Frankemölle, and Tell 2024). Thereby, it can be used to visualise the innovativeness of a word. Yet a prominence measure for emphasising contrast is the word frequency difference between two adjacent years visualised, i.e., either to visualise the top n words whose frequency had increased the most compared to the previous year, or – the opposite – to visualise those with the largest frequency decrease compared to the previous year.

Given these large number of suggestions for how to tailor the word prominence measure to different needs, we decided not to try to implement any of these as a built-in configuration option, but to still limit the pre-implemented configurable measures to the two standards tf and tf-idf. Instead, we added a more flexible configuration option to the Word Rain implementation that allows the user to define their own function for calculating word prominence. This option requires more programming skills from the user, but it also provides a very large degree of flexibility, and makes it possible to implement prominence measures we have not yet envisioned would be relevant.[3]

Figure 3 provides an example of when relative word frequency difference compared to the previous year is used as the prominence measure. That is, the first graph shows the words that have increased the most between 1957 and 1958, and the second graph those that increased the most between 1958 and 1959. (The graphs also exclude words present every year in the corpus studied.) The graphs show, for instance, that in 1959, the medication "Tolbutamide" is mentioned more often than the previous year. This is also the case for a number of sweeteners/sugars, such as "Sweetex", "Sorbitol", and "Dextrosol".

[3]We have started to provide example code for creating different user-defined prominence measures, and we intend to expand that library with additional examples: https://github.com/CDHUppsala/word-rain/tree/main/handbook

4.4. Additional features

We also made a number of smaller feature additions and updates, of which we will only mention two here. The first regards the visual appearance of the word rains. When producing many graphs in a series of word rains, some became very dense and some very spacious, depending on the frequency distribution among the top n words for the year visualised. It might be a possible future project to automatically create a visually appealing balance between density and spaciousness. However, in the meantime, we added a parameter, which makes it possible to manually configure this balance for each word rain graph in a series.

Another simple improvement, but one that was very important when practically using the word rains, regarded the visual marker for signalling when a word appears for the first time in a temporal series of graphs. When using the word rain practically, we discovered that being able to search in the word rains was an important feature, for words in general, but also for when they first appear in a temporal word rain series. Finding the first appearance of a particular word can be valuable information when determining when certain topics, people, places or things enter into a particular discourse. In early versions, the first appearance of a word was initially marked in bold. However, due to the different sizes of the words it was sometimes hard to tell whether a word was marked in bold or just large, and more importantly it was not possible to search for bold words. In the next version, these words were therefore marked with an asterisk (*). However, for the pdf readers we tried, it was not possible to use asterisk in a search term,[4] which made it impossible to search for the first occurrence of a word. Finally, we opted to instead use a hash (#) at the end of the word as the indicator of the word appearing for the first time.

5. Conclusion and future work

Applying the Word Rain visualisation in a digital history project meant the technique could be further tailored to the needs of its intended users and the possible application areas within the digital humanities. As a result of regular discussions among the members within the project, relevant features were suggested and evaluated, and new features that enhanced the usability and readability of the word rains were added.

Future work on the implementation of additional features can be taken in different directions. There are two additional features we noticed were missing when practically applying the Word Rain visualisation. However, we have not yet determined whether these features should be integrated into the Word Rain visualisation or if they are better suited as visualisations of their own. The first feature concerns the fact that the Word Rain visualisation now only focuses on paradigmatic similarity, as it sorts the words according to the extent to which they occur in similar contexts. However, it might also be relevant to explore syntagmatic similarity, i.e., the extent to which the words co-occur. One possible solution might be to produce word rains where the horizontal axis represents syntagmatic similarity instead of paradigmatic. Another option might be to try to represent both types of word similarities in the same graph, for instance by using lines that connect words that often co-occur.

[4] Possibly due to asterisk often being used as wildcard in search terms.

The second feature regards the granularity of the paradigmatic similarity shown by the Word Rain visualisation technique. The point of using word embeddings for positioning the words is to provide semantic information that can support the user in exploring the most prominent words in a text. The point is not to support the study of fine-grained differences in word similarity. However, it might be relevant to also carry out such an investigation of paradigmatic word similarity on a more detailed level. It could, e.g., be useful to investigate whether words belonging to a specific semantic category (e.g., for our corpora, categories such as diseases, symptoms or occupations) occur in similar contexts, or if they occur in different semantic contexts, and thereby can be divided into several semantic clusters. It could also be relevant to study longitudinal changes in the contexts in which a word occurs. Therefore, it might be useful to supplement the Word Rain technique with additional text visualisation techniques that retain more of the semantic information encoded in the multidimensional word embeddings.

6. Acknowledgments

The ActDisease project is funded by the European Union (ERC ActDisease ERC-2021-STG 101040999). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

The development of Word Rain as a digital humanities research infrastructure resource is funded by the Swedish Research Council: Huminfra (2021-00176), InfraVis (2021-00181) and Swe-CLARIN/The National Language Bank of Sweden (2017-00626).

References

Aangenendt, Gijs, Maria Skeppstedt, and Ylva Söderfeldt. 2024. "Curating a historical source corpus of 20th century patient organization periodicals." In *Proceedings of the Huminfra Conference,* 76–82. HiC '24. January 4, 2024. Accessed January 17, 2024. https://doi.org/10.3384/ecp205011.

Ahltorp, Magnus, Jean Hessel, Gunnar Eriksson, and Maria Skeppstedt. 2025. "Visualisering av ett lexikons täckning av olika textgenrer: Experiment med en jiddischordbok (Visualising the coverage of dictionary for different text genres: Experiments with a Yiddish dictionary)." In *Nordiske Studier i Leksikografi (accepted for publication).*

Ahltorp, Magnus, and Maria Skeppstedt. 2024. "Word Rain as a Service." In *Proceedings from the CLARIN Annual Conference 2024.*

Barth, Lukas, Stephen G. Kobourov, and Sergey Pupyrev. 2014. "Experimental Comparison of Semantic Word Clouds." In *Experimental Algorithms,* 247–258. Springer International Publishing. https://doi.org/10.1007/978-3-319-07959-2_21.

Brath, Richard, and Ebad Banissi. 2016. "Using Typography to Expand the Design Space of Data Visualization." *She Ji: The Journal of Design, Economics, and Innovation* 2 (1): 59–87. https://doi.org/10.1016/j.sheji.2016.05.003.

Burch, Michael, Steffen Lohmann, Fabian Beck, Nils Rodriguez, Lorenzo Di Silvestro, and Daniel Weiskopf. 2014. "RadCloud: Visualizing Multiple Texts with Merged Word Clouds." In *Proceedings of the International Conference on Information Visualisation,* 108–113. IV '14. IEEE. https://doi.org/10.1109/IV.2014.72.

Cao, Nan, and Weiwei Cui. 2016. *Introduction to Text Visualization.* Vol. 1. Atlantis Briefs in Artificial Intelligence. Atlantis Press. https://doi.org/10.2991/978-94-6239-186-4.

Collins, Christopher, Fernanda B. Viégas, and Martin Wattenberg. 2009. "Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora." In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology,* 91–98. VAST '09. IEEE. https://doi.org/10.1109/VAST.2009.5333443.

Cuba Gyllensten, Amaru. 2023. "Quantifying Meaning." PhD diss., KTH Royal Institute of Technology.

Cui, Weiwei, Yingcai Wu, Shixia Liu, Furu Wei, Michelle Zhou, and Huamin Qu. 2010. "Context-Preserving, Dynamic Word Cloud Visualization." *IEEE Computer Graphics and Applications* 30, no. 6 (November): 42–53. https://doi.org/10.1109/MCG.2010.102.

Diakopoulos, Nicholas, Dag Elgesem, Andrew Salway, Amy Zhang, and Knut Hofland. 2015. "Compare Clouds: Visualizing Text Corpora to Compare Media Frames." In *Proceedings of the IUI Workshop on Visual Text Analytics.*

Drucker, Johanna. 2015. "Graphical Approaches to the Digital Humanities." In *A New Companion to Digital Humanities,* edited by Susan Schreibman, Ray Siemens, and John Unsworth, 238–250. New York: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118680605.ch17.

Herold, Elisa, Marcus Pöckelmann, Christian Berg, Jörg Ritter, and Mark M. Hall. 2019. "Stable Word-Clouds for Visualising Text-Changes Over Time." In *Digital Libraries for Open Knowledge,* 224–237. Springer International Publishing. https://doi.org/10.1007/978-3-030-30760-8_20.

Hicke, Rebecca M. M., Maanya Goenka, and Eric Alexander. 2022. "Word Clouds in the Wild." In *Proceedings of the IEEE Workshop on Visualization for the Digital Humanities,* 43–48. VIS4DH '22. IEEE. https://doi.org/10.1109/VIS4DH57440.2022.00015.

Koh, Kyle, Bongshin Lee, Bohyoung Kim, and Jinwook Seo. 2010. "ManiWordle: Providing Flexible Control over Wordle." *IEEE Transactions on Visualization and Computer Graphics* 16, no. 6 (November): 1190–1197. https://doi.org/10.1109/TVCG.2010.175.

La Mela, M., J. Frankemölle, and F. Tell. 2024. "Using document similarity in historical patents to study value in innovation (long paper)." In *Digital Humanities in the Nordic and Baltic Countries conference DHNB 2024.*

Lee, Bongshin, Nathalie Henry Riche, Amy K. Karlson, and Sheelash Carpendale. 2010. "SparkClouds: Visualizing Trends in Tag Clouds." *IEEE Transactions on Visualization and Computer Graphics* 16, no. 6 (November): 1182–1189. https://doi.org/10.1109/TVCG.2010.194.

Liu, Xiaotong, Han-Wei Shen, and Yifan Hu. 2015. "Supporting Multifaceted Viewing of Word Clouds With Focus+Context Display." *Information Visualization* 14, no. 2 (April): 168–180. https://doi.org/10.1177/1473871614534095.

Lohmann, Steffen, Michael Burch, Hansjörg Schmauder, and Daniel Weiskopf. 2012. "Visual Analysis of Microblog Content Using Time-Varying Co-Occurrence Highlighting in Tag Clouds." In *Proceedings of the International Working Conference on Advanced Visual Interfaces,* 753–756. AVI '12. Association for Computing Machinery. https://doi.org/10.1145/2254556.2254701.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using t-SNE." *Journal of Machine Learning Research* 9:2579–2605.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825–2830.

Skeppstedt, Maria, and Gijs Aangenendt. 2024. "Using the Word Rain Technique to Visualize Longitudinal Changes in Periodicals from the Swedish Diabetes Association." In *Proceedings of the Workshop on Visualization for Natural Language Processing*. Vis4NLP '24. The Eurographics Association. https://doi.org/10.2312/vis4nlp.20241132.

Skeppstedt, Maria, Magnus Ahltorp, Kostiantyn Kucher, and Matts Lindström. 2024. "From Word Clouds to Word Rain: Revisiting the Classic Word Cloud to Visualize Climate Change Texts." *Information Visualization* 23, no. 3 (July): 217–238. https://doi.org/10.1177/14738716241236188.

Wang, Yunhai, Xiaowei Chu, Chen Bao, Lifeng Zhu, Oliver Deussen, Baoquan Chen, and Michael Sedlmair. 2018. "EdWordle: Consistency-Preserving Word Cloud Editing." *IEEE Transactions on Visualization and Computer Graphics* 24, no. 1 (January): 647–656. https://doi.org/10.1109/TVCG.2017.2745859.

Wu, Yingcai, Thomas Provan, Furu Wei, Shixia Liu, and Kwan-Liu Ma. 2011. "Semantic-Preserving Word Clouds by Seam Carving." *Computer Graphics Forum* 30, no. 3 (June): 741–750. https://doi.org/0.1111/j.1467-8659.2011.01923.x.

Xie, Liwenhan, Xinhuan Shu, Jeon Cheol Su, Yun Wang, Siming Chen, and Huamin Qu. 2024. "Creating Emordle: Animating Word Cloud for Emotion Expression." *IEEE Transactions on Visualization and Computer Graphics* 30 (8): 5198–5211. https://doi.org/10.1109/TVCG.2023.3286392.

Xu, Jin, Yubo Tao, and Hai Lin. 2016. "Semantic word cloud generation based on word embeddings." In *Proceedings of the IEEE Pacific Visualization Symposium,* 239–243. PacificVis '16. IEEE. https://doi.org/10.1109/PACIFICVIS.2016.7465278.

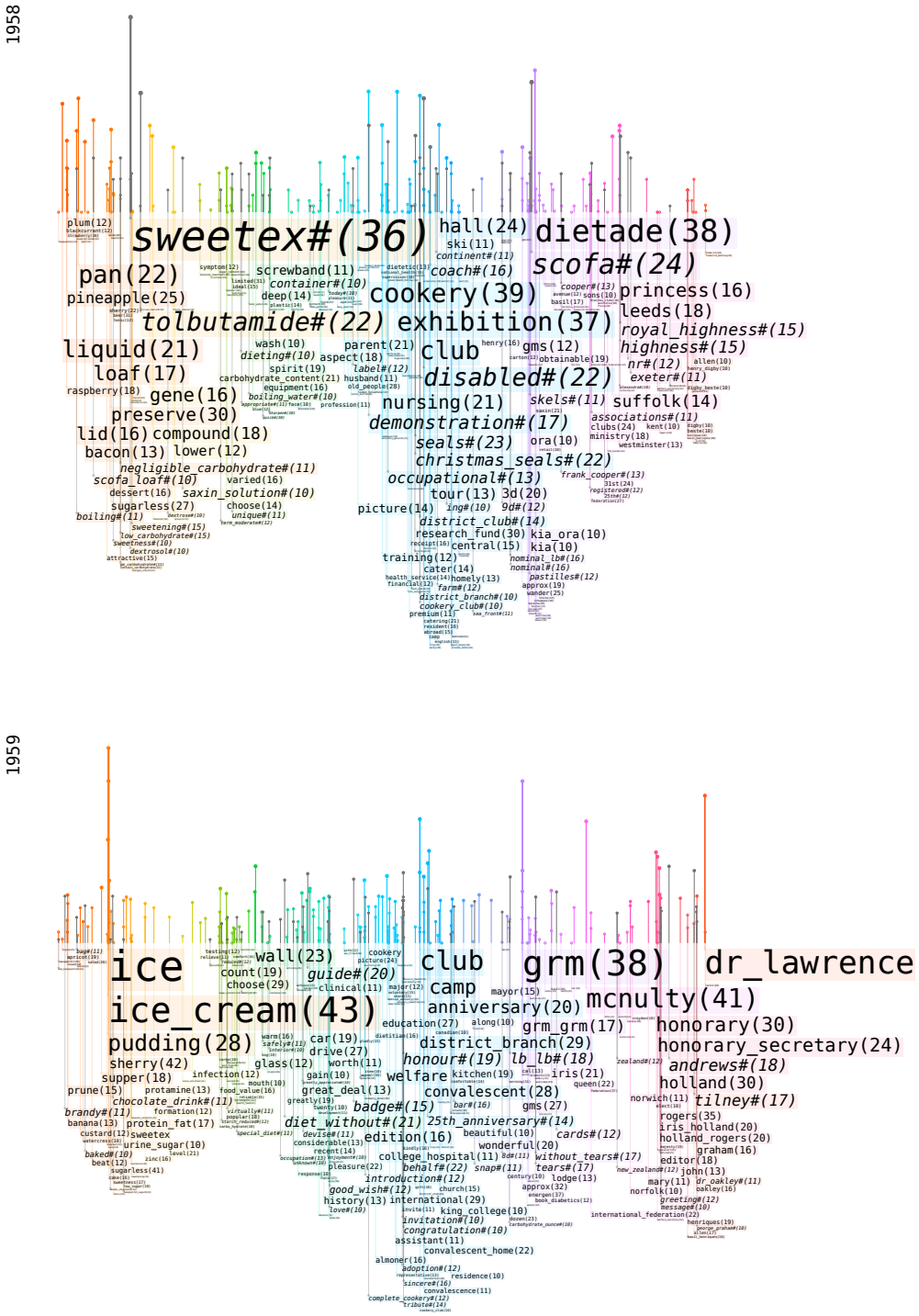

Figure 3: Two word rains generated for periodicals published by the British Diabetes Association (using term frequency difference compared to previous year as the prominence measure). Note that these stem from another series of graphs than the Figures 1 and 2, and use another word positioning.