

Project Report

by Batmanova Zlata, Fischer Leona, Löschnauer Maria and Wagner Sarah

CLIMATE CHANGE DISCOURSE IN AUSTRIAN PARLIAMENTARY DEBATES

As the topic of climate change grows more important by day, the main objective of this project is to find out how this growth of importance is detectable in Austrian politics. To further narrow it down the focus is on the climate change discourse in Austrian parliamentary debates in particular.

Climate change refers to long-term shifts in temperatures and weather patterns on a global level. Over the past two centuries, human activities have been the primary driver of these changes. We have reached a critical threshold, where the consequences of prolonged inaction become increasingly apparent. The environment is affected in numerous ways, including rising global temperatures, drought, floodings, wildfires and other extreme weather events. These phenomena have a direct impact on essential systems and services such as water supply, energy infrastructure, transportation, biodiversity, agriculture, ecosystems and public health. Mitigating the impacts and reducing their severity requires substantial investments by governments, industries and other stakeholders. Politicians not only play a key role in policy-making but also in shaping the public narrative around climate change. Their communications strategies, whether fact-based or emotionally driven, can significantly influence public perception, awareness and action.

1. Detailed Description of the Research Questions

With this background, we are seeking to answer the question of how the frequency of climate change terminology has changed in Austrian parliamentary debates over time. We compiled a predefined list of commonly used terms related to climate change. This process will be further explained in the chapter on data collection. Furthermore, we decided to examine the debates of the years 1996 to 1999 and from 2019 to 2022, in order to limit the scope of the material and allow for a more focused analysis. The most recent data available stems from the year 2022.

The reason why we chose these years in particular is rather simple as the data from 1996 is the first available and the one from 2022 is the most recent data. Therefore we chose to compare the first four years of available data to the last four years.

Originally we were going to examine how climate terms are contextually framed. This was with the intention to see if they are used in the context of policies, economy, health or crisis for example. And also then highlight what terms co-occur most frequently with these keywords. However as we started working with the data we soon realised that this would exceed the possible workload and would be too complex for the project scope as it was nearly impossible to determine each usage of each term clearly. Due to the nature of parliamentary debates, the boundaries between the context of policies, economy, health and also crisis are often blurred. So we decided to cut down our research questions to keep it in a feasible frame. These adjustments have led to the following finalized research questions:

- How has the frequency of climate change terminology changed in Austrian parliamentary debates over time?

- How does the use and framing of climate change terminology vary across different political actors and party affiliations?

2. Project Outcome

The analysis clearly shows a significant increase in the use of climate-related terminology in Austrian parliamentary debates between 1996–1999 and 2019–2022. This not only reflects the growing political relevance of climate change but also the rise of environmental issues. It is apparent that the frequency of terms such as „Klima“, „CO2“, and „erneuerbar“ has increased and also that the discourse overall has become more differentiated and policy-specific. The findings indicate that climate change has evolved from a peripheral concern into a central topic of legislative debate, shaped by party affiliation and speaker demographics. These results confirm that political discourse in the Austrian parliament has been influenced by the urgency of the climate crisis, particularly in recent years.

3. Finalized Report with a Description of the Data Collected

For our project, we used data from the ParlaMint corpus. ParlaMint is a CLARIN initiative that provides a collection of uniformly encoded and comparable corpora comprising transcriptions of parliamentary debates from various countries and autonomous regions (CLARIN, 2024).

Each corpus in ParlaMint is structured by parliamentary session days and includes extensive metadata. This metadata provides information about the terms, sessions, and meetings in which the debates took place. Each speech is attributed to a speaker, with metadata typically including the speaker's name, sex, role, and political

party affiliation. The transcripts also include transcriber comments such as applause, interruptions, or pauses (CLARIN.SI, 2024).

ParlaMint's latest version, 4.1, is available in three formats. The first is a plain-text version without linguistic annotation. The second, which is used in our project, is a linguistically annotated version featuring syntactic parsing and named entity recognition. The third format includes both linguistic annotation and machine translation of the speeches into English. All ParlaMint corpora are XML-encoded following a customization of the TEI Guidelines, which ensures interoperability across languages and time periods (CLARIN, 2024).

For our project, we used the ParlaMint AT corpus. It contains transcriptions of Austrian parliamentary debates from 1996 to 2022. The data is provided by the Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH) at the Austrian Academy of Sciences (ÖAW) (Austrian Academy of Sciences, Austrian Centre for Digital Humanities and Cultural Heritage, n.d.). The corpus is freely available for download and further processing.

To analyze the evolution of climate change discourse over time, we focused on two time periods: the earliest four years of the corpus (1996–1999), and the most recent four years (2019–2022). The TEI-encoded files were imported into the Oxygen XML Editor for further analysis. Due to the high quality and consistency of the data, no additional cleaning or pre-processing was required before beginning the analytical steps.

After selecting the relevant subcorpora from ParlaMint AT and conducting an initial review of the TEI-encoded files, we compiled a list of climate-related terms that

would serve as the basis for all further queries and the resulting database. These terms were selected based on their relevance to the discourse on climate change.

We identified two academic sources that had previously analyzed climate discourse in German-language media. They formed the starting point for our collection of climate-related terms:

- Tereick (2016) conducted a diachronic corpus-assisted discourse analysis of German newspapers, magazines, YouTube videos, and television broadcasts over a 16-year period (1995–2010). Her study includes a search string used to identify climate-related texts, which provided a foundational set of lexical items for our own lemma list.
- Müller (2021) examined how the framing of climate change shifted over time in German print media (2003–2019). His list of keywords used to identify climate-related articles was also integrated into our term collection.

To further expand our lemma list, we consulted three official glossaries that systematically define climate- and environment-related terminology to provide a logically structured and clearly delineated overview of central environmental concepts, enhancing the scientific, political, and public discourse on the protection of natural resources:

- Glossar zum Ressourcenschutz (Glossary on Resource Conservation) (German Environment Agency, 2012)
- Climate Glossary of the Second Austrian Assessment Report on Climate Change (AAR2) (APCC, 2025)

- BMIMI Climate Glossary (Austrian Federal Ministry for Innovation, Mobility and Infrastructure, 2022)

Based on these sources, we compiled a finalized lemma list consisting of German terms that refer directly or primarily to climate change or environmental issues. This list can be found in the appendix of this report. These lemmata were used in our XQuery-based extractions (see Chapters 4 and 5 for details). In these queries, we focused not only on the base form of each lemma, but employed Corpus Query Language (CQL), a formal query language used to define complex search patterns. The queries were constructed carefully, using trial and error to refine patterns and ensure that as many relevant results as possible were included.

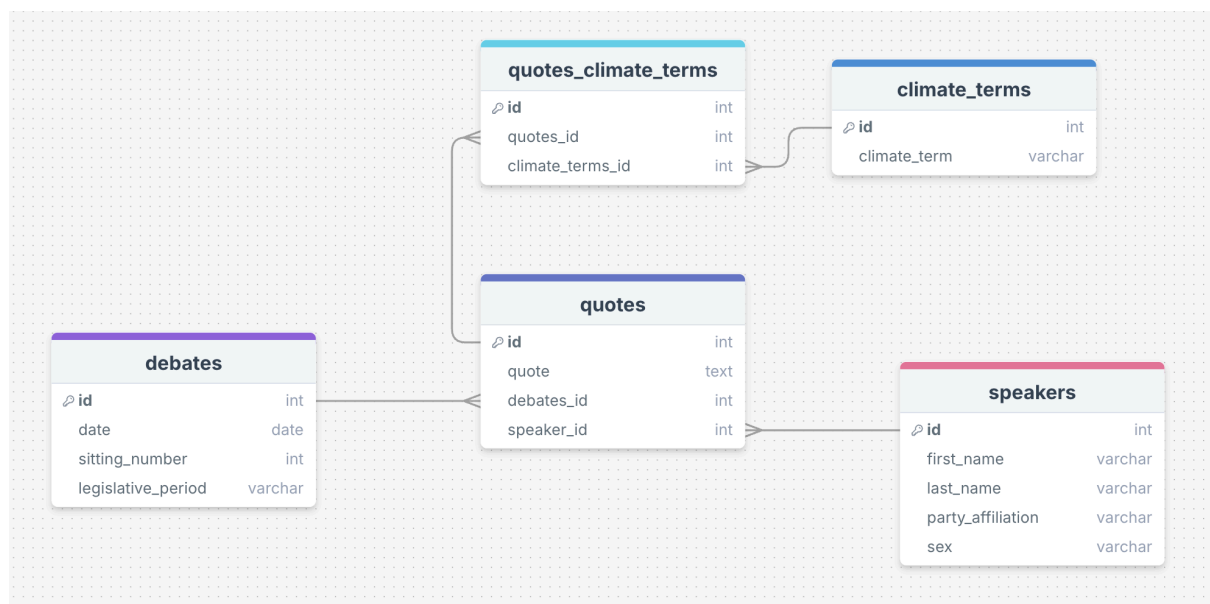
Using CQL allowed us to capture a wide range of related forms, such as plurals, derivations, and compounds. However, the broader scope of the CQL queries also introduced semantic noise, as some matches appeared in non-climate contexts - for example, the word *Klima* in *Arbeitsklima* or as the personal name of the former Austrian politician *Viktor Klima*. To address this, we reviewed a sample of results and adapted our CQL queries to exclude common false positives using negative pattern matching and logical constraints. This resulted in a curated list of CQL expressions, which is included in the appendix.

Since our goal was to capture as many relevant climate-related terms as possible, we chose not to apply overly restrictive filters. Instead, we manually verified each result in the final output. Only those matches that were relevant to the topic of climate change were counted in our final frequency results.

4. Database Schema

To enable a structured and efficient analysis, a relational database was developed and implemented using Oxygen XML Editor, XQuery and DB Browser for SQLite. A detailed description of the setup and workflow can be found in the file *Database_Setup_and_Analysis.pdf*.

First, a project environment was created, with dedicated folders for input data, query scripts and output files. Next, a relational database schema was drawn in DrawSQL:



The database schema consists of five core tables: *debates*, *speakers*, *quotes*, *climate_terms* and *quotes_climate_terms*. Each table has a primary key (*id*). Some primary keys are used as foreign keys in other tables (e.g. *debates_id* in the table *quotes*). A one-to-many relationship was established between the *debates* and the *quotes* tables (= each debate consists of multiple quotes, while each quote belongs to one specific debate) and between the *speakers* and the *quotes* tables (= each speaker can have multiple contributions, but every quote is spoken by a specific speaker). As one quote can include more than one climate term

and one climate term can be included in many quotes, a many-to-many relationship was implemented between the tables `quotes` and `climate_terms`, using the associative table `quotes_climate_terms`.

In our current database schema, the `party_affiliation` is stored as a varchar field within the `speakers` table. This design choice was made for the sake of limited time available. However, for a more advanced analysis - such as evaluating party-related patterns in climate discourse - a separate, normalized table for this attribute would be beneficial. It would also allow the addition of metadata such as party colors. To allow quotes to be classified by thematic context, an additional *context* table could also be added.

SketchEngine

At the beginning of our project, the use of SketchEngine, which is commonly used to analyse text corpora, was decided. It is a tool to explore how language works, where algorithms analyze text corpora to identify different metrics such as collocations and word combinations, concordance, word frequency, keywords of the respective corpus as well as statistics of the whole corpus. As it is designed for text analysis, which is the nature of this project, the decision to make use of this tool regarding the research of climate terms frequency was made. This is also due to the clear structure and straightforward nature of this tool. After uploading all the files of the parliamentary debates from 1996 to 1999 and from 2019 to 2022 in txt format, a corpus called *Austrian parliamentary debates* was created. It contains a total of eight files. The whole corpus contains 19,685,603 words. Corpus Query Language (CQL) was used to query each climate term for each year. Then the frequency was noted in an EXCEL Sheet, where all frequencies for each year and term were collected.

As the project progressed, it became evident that more control over data structure, metadata integration and manual cleaning was required - particularly to correct false positives and manage speaker information. Therefore, we transitioned to using the Oxygen XML Editor to transform the corpus XML files into SQL files via XQueries, and used the DB Browser for SQLite for data refinement and analysis within a relational database framework.

Oxygen XML Editor

Using XQuery, two key scripts were developed and applied on the XML files in the ParlaMint AT corpus:

- *CCD_setup_tables.xquery* – executed only once – to initialize the database schema by creating tables with the previously defined keys and attributes. This script also inserted the predefined climate terms (see chapter 3), stored in a XML file (*climate_terms.xml*), in the corresponding tables
- *CCD_XML_to-SQL.xquery* – applied to all debate files at once – to extract structured content from the corpus.

As the corpus files are encoded in TEI, it is important to ensure that the XML namespace is correctly declared in both XQuery scripts prior to querying (Prefix: *tei* and URI: <http://www.tei-c.org/ns/1.0>).

SQLite Database

After generating the SQL scripts via XQuery, a SQLite database was created using DB Browser for SQLite. The *CCD_setup_tables.sql* file was executed to create the schema and insert climate terms, followed by *CCD_XML_to-SQL.sql* to populate the tables. Upon examination of the tables, it was observed that the regular expressions from the climate terms list were not fully incorporated within the XQuery

implementation. Furthermore, during reviewing the query outputs, several false positive entries were identified that do not relate to climate change (e.g., “Klima der Debatte”, “Klima der Studentenfeindlichkeit”). Consequently, manual data cleaning and refinement were required through the use of SQLite queries.

5. Summary of the data

The following table shows the number of entries before and after the cleanup:

Table	Before cleanup	After cleanup
<code>climate_terms</code>	32 entries	32 entries
<code>debates</code>	395 entries	395 entries
<code>quotes</code>	11.469 entries	10.141 entries
<code>quotes_climate_terms</code>	16.868 entries	14.959 entries
<code>speaker</code>	514 entries	514 entries

In the `speaker` table, it was also noted that individuals with two or more recorded names (for example, when a name changed due to marriage) had all name variants returned. This could have been addressed directly within the XQuery logic itself, but SQLite queries also proved effective for its resolution.

From this point forward, the database functions as the central framework for analyzing the findings further.

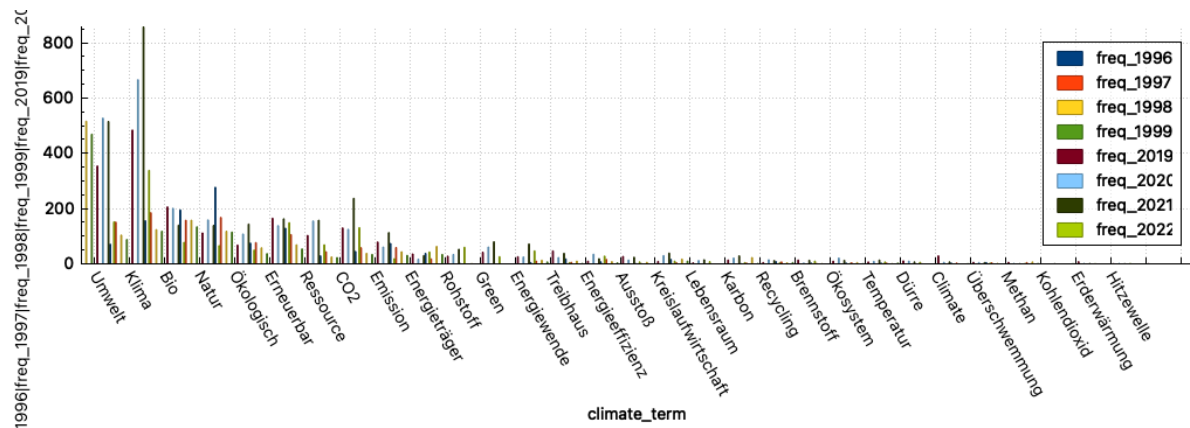
Analysis

To answer our research question of how the frequency of climate change terminology has changed in Austrian parliamentary debates over time, SQLite

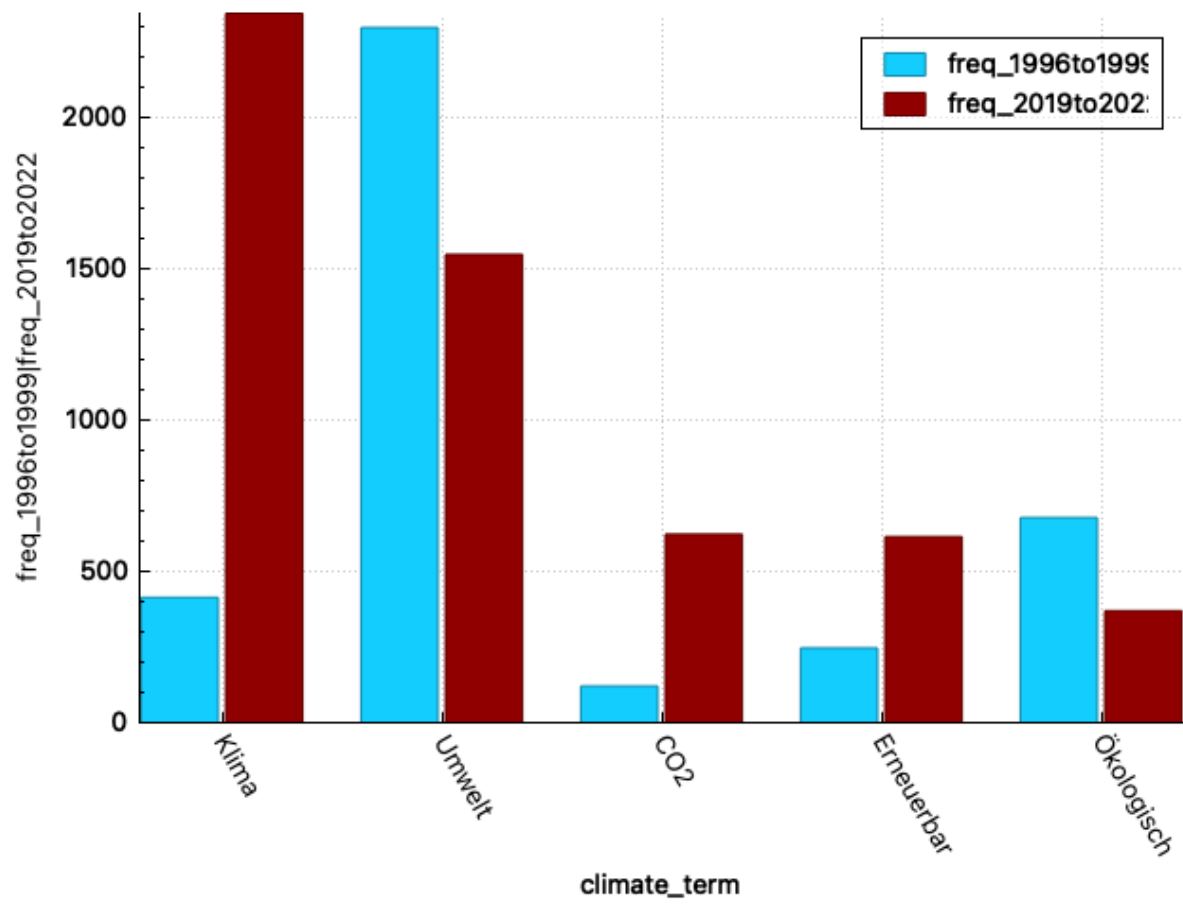
queries were used. To visualise the results, the Plot View in the DB Browser was used to generate bar charts.

1. Term Frequency Analysis: Which terms dominate the climate discourse?

All climate terms (1996–1999 and 2019–2022):



Top 5 climate terms with the most significant change in frequency (1996–1999 compared to 2019–2022):



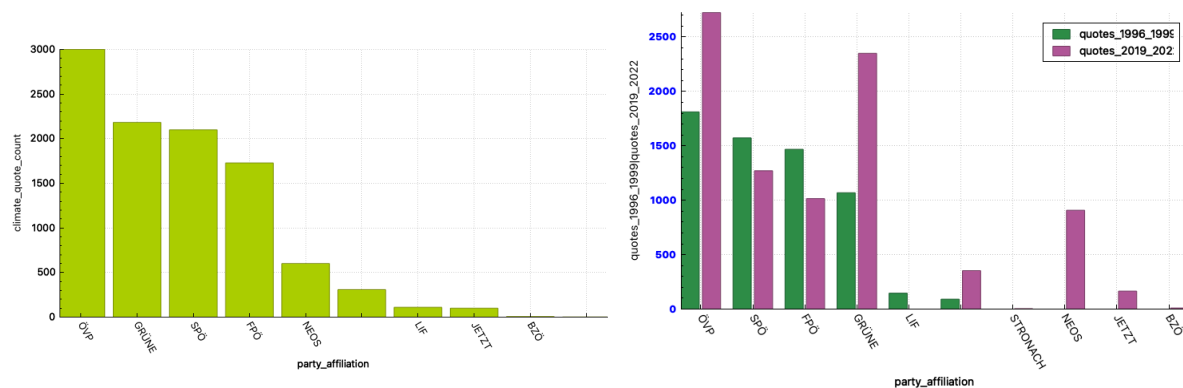
Frequency analysis shows that certain terms are particularly prominent in the political discourse of the Austrian parliament and according to the evaluation across all years, the following terms dominate:

- “Klima”: By far the most frequently used term, although despite extensive filtering, a certain amount of semantic noise remains possible.
- “Umwelt”: The term was dominant in the late 1990s, reflecting a more general environmental focus. Its relative decline in recent years suggests a shift toward more specific and technical climate terms like "Klima" and "CO2".
- “CO2”: These term is central in both technical and political contexts and reflects the increasing focus on emissions and climate targets.
- “Erneuerbar”: Their increased frequency in recent debates shows that the transformation of the energy sector has become a central topic of discourse.
- “Ökologisch”: This term shows no significant increase over time, rather has declined by a bit. It is often used in broader lifestyle or agriculture contexts and appears less frequently in today’s more policy-driven climate discourse.

Overall not only has the absolute number of climate-related terms increased, but there is also greater variance and specificity in the terms used. While the discussion in the late 1990s was still relatively broad, today it is more technical, differentiated, and politically charged.

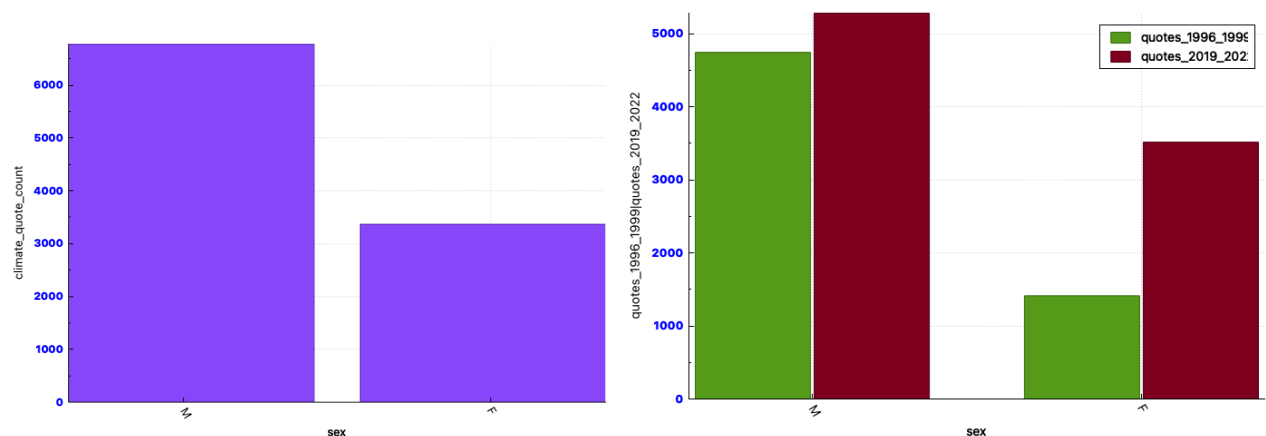
2. Speaker Analysis: Who contributes most to climate discourse? Is there a variation by party or sex?

Party: The diagrams in the report show clear differences in climate term frequency between political parties and indicate an increase in climate discourse for ÖVP, GRÜNE and NEOS in 2019–2022 compared to 1996–1999.



It is important to note that an increase in climate term frequencies may not reflect growing commitment in climate issues within a party. It can also be influenced by changes in parliamentary representation, with more members from certain parties holding seats and therefore contributing more to the total number of quotes.

Sex:



Overall it is clear that men are using climate terms more frequently than women, which corresponds to the general gender ratio in parliament. It is also apparent that

the usage of climate terms by women increased by more than the double amount in 2019–2022 compared to 1996–1999. This could be the result of the increase in women in the parliament in general.

Conclusions that can be drawn

Following this analysis several conclusions can be drawn:

Increased discussion: Climate change is no longer a marginal issue, but a central political topic, as evidenced by the significant increase in the frequency of the term between the periods examined.

Change in discourse: While the discourse in the 1990s was still overall rather general in nature, in recent years it is differentiated, strategic, and in some cases also ideologically charged.

Party political differences: The way in which climate is discussed is strongly party-bound. Progressive parties focus more on narratives of responsibility, justice, and transformation, while conservative voices argue in a more technocratic or business-oriented manner.

Gender gap in climate communication: Women are playing an increasingly important role in the climate discourse and are bringing more social and sustainable perspectives to the table, even though men have more speaking time overall.

Challenges:

- Some climate terms have multiple meanings; although we narrowed down the results with filters and exclusions, some noise may remain
- Fluctuations in total speaking time or number of debates per year may skew term frequency comparisons

- Sex/Gender dynamics: Conclusions should be contexted carefully, considering factors like party size or proportion of seats held by men/women

6. Possible Future Uses for the Database

Our database provides a foundation for further research into climate-related political discourse. One possible future use is to trace how political language around climate change has evolved over time through diachronic discourse analysis. This could include changes in terminology, rhetorical framing, and thematic focus. Discursive strategies in political climate communication, such as the use of emotional appeals, metaphors, or dominant narrative structures, could also be explored.

In addition, the database supports party-specific language analysis, allowing comparisons of how different political parties conceptualize and communicate climate-related issues, potentially reflecting ideological divergences or rhetorical strategies. It also offers the potential for speaker profiling by examining how individual politicians employ climate-related terminology across different time periods and contexts. Moreover, the database could be linked to external datasets, such as policy documents, legislative records, or public opinion data, in order to explore possible correlations between discourse and political action.

The database could be expanded in several aspects. For example, it could be extended to cover the full temporal range of the ParlaMint AT corpus, allowing for a more continuous and fine-grained analysis of climate discourse. Another direction would be to apply the same methodology to other ParlaMint corpora, enabling comparative analyses across different national parliamentary contexts. In addition, the climate term list could be extended to include further relevant concepts and

terminology, providing a more comprehensive view of climate-related language in political discourse.

Finally, a potential area for future development concerns an aspect that had to be omitted in the current project due to limited time and resources: the contextual framing of climate-related terms. While this proved too complex for the current scope, it remains a promising area of future research.

7. Division of Responsibilities

All members of the project group were involved in key steps of the data collection and analysis process. Together, the team compiled the list of climate-related terms that served as the basis for all queries. Each member was responsible for two years of data from the ParlaMint AT corpus, reviewed whether structural adjustments or data cleaning were required for their assigned data, and conducted both the XQuery-based extraction and the frequency analysis in Sketch Engine. The results were manually reviewed by the group members to exclude false positives and ensure topic relevance. This approach ensured that all group members engaged with both the dataset and the methods introduced during the course.

In addition, each group member contributed to writing the Project Documentation Report. The team members also took on specific responsibilities: Sarah Wagner set up the corpus in Sketch Engine and ensured it was properly configured for frequency analysis. Leona Fischer created and refined the CQL queries, balancing the inclusion of compounds and inflected forms with the exclusion of false positives. Zlata Batmanova oversaw the documentation process and maintained the data management plan throughout the project. Maria Löschnauer developed guidelines to

standardize the XQuery extraction process across the group and merged the resulting datasets into a unified database.

References

- Austrian Academy of Sciences, Austrian Centre for Digital Humanities and Cultural Heritage. (n.d.). *ParlaMint*.
<https://www.oeaw.ac.at/acdh/research/dh-research-infrastructure/activities/research-infrastructures/parlamint>
- Austrian Panel on Climate Change (APCC). (2025). *Glossary*. In D. Huppmann, M. Keiler, K. Riahi, & H. Rieder, *Second Austrian Assessment Report on Climate Change (AAR2)*. Austrian Academy of Sciences Press. <https://aar2.ccca.ac.at/glossar>
- Bundesministerium für Innovation, Mobilität und Infrastruktur. (2022). *Klima-Glossar: Wie Treibhauseffekt, Jetstream & Kippelemente die Erde prägen (werden)*.
<https://infothek.bmimi.gv.at/klima-glossar-treibhauseffekt-kippelemente-jetstream-system-erde/>
- CLARIN (2024). *ParlaMint: Comparable Parliamentary Corpora*. GitHub.
<https://github.com/clarin-eric/ParlaMint>
- CLARIN.SI (2024). *Multilingual comparable corpora of parliamentary debates ParlaMint 4.1*. <https://www.clarin.si/repository/xmlui/handle/11356/1912>
- Müller, M. (2021). Schlüsselkonzepte im deutschen Mediendiskurs zum Klimawandel. Eine korpuslinguistische Verlaufsstudie der Diskursverschiebungen ab 2003. In M. Reisigl (Ed.), *Klima in der Krise: Kontroversen, Widersprüche und Herausforderungen in Diskursen über Klimawandel* (Osnabrücker Beiträge zur Sprachtheorie, Vol. 97, pp. 167–188). Universitätsverlag Rhein-Ruhr.
- Tereick, J. (2016). *Klimawandel im Diskurs: Multimodale Diskursanalyse crossmedialer Korpora* (Discourse Patterns, Vol. 13). Walter de Gruyter.
- Umweltbundesamt. (2012). *Glossar zum Ressourcenschutz*.
<https://www.umweltbundesamt.de/publikationen/glossar-ressourcenschutz>

Appendix

List of Climate Terms

- Ausstoß
- Bio
- Brennstoff
- Climate
- CO₂
- Dürre
- Emission
- Energieeffizienz
- Energieträger
- Energiewende
- Erneuerbar
- Erderwärmung
- Green
- Hitzewelle
- Karbon
- Kipppunkt
- Klima
- Kohlendioxid
- Kreislaufwirtschaft
- Lebensraum
- Methan
- Natur
- Naturkatastrophe
- Ökosystem
- Ökologisch
- Pariser Abkommen
- Recycling
- Ressource
- Rohstoff
- Temperatur
- Treibhaus
- Umwelt
- Überschwemmung

CQL Expressions

- [lc=".*ausstoß.*"]
- ([lc="bio"] | [lemma="biologisch"] | [lemma="Bioabfall"] | [lc="biodiversität.*"] | [lc="bioenergie.*"] | [lc="bioabfall.*"] | [lc="bioabfälle.*"] | [lc="biokraftstoff.*"] | [lc="biomasse.*"] | [lc="biogas.*"] | [lc="bioethanol.*"] | [lc="bioplastik.*"])
- [lc=".*brennstoff.*"]
- [lc=".*climate.*"]
- [lc=".*co(2|₂)².*"]
- [lc=".*karbon.*"]
- [lc=".*kohl(enstoff(dioxid)?|endioxid).*"]
- [lc=".*dürre.*" & lc!=".*dürrenmatt.*" & lc!=".*dürrenstein.*"]
- [lc=".*emission.*" & lc!=".*demission.*" & lc!=".*lärmemission.*" & lc!=".*schallemmission.*" & lc!=".*emissionsvolumen.*" & lc!=".*emissionsgeschäft.*"]
- [lc=".*energieeffizienz.*"] | [lc=".*energieeffizient.*"]
- [lc=".*energieträger.*"]
- [lc=".*energiewende.*"]
- [lc=".*erneuerbar.*"]
- ([lc=".*erwärmung.*" & lc!=".*schwimmbädererwärmung.*"] | [lc=".*erwärmen.*"] | [lc=".*erwärmt.*"])
- [lc=".*green.*" & lc!=".*greenwich.*" & lc!=".*greenham.*" & lc!=".*green-Potato-Jugend-Projekt.*" & lc!=".*evergreen.*" & lc!=".*walgreens.*"]
- [lc=".*hitzewelle.*"]
- [lc=".*kipppunkt.*"]
- ([lc!="viktor"] [lc=".*klima.*" & lc!=".*konjunkturklima.*" & lc!=".*arbeitsklima.*" & lc!=".*wirtschaftsklima.*" & lc!=".*konsensklima.*" & lc!=".*investitionsklima.*" & lc!=".*stimmungsklima.*" & lc!=".*gesprächsklima.*" & lc!=".*meinungsklima.*" & lc!=".*gründungsklima.*" & lc!=".*kulturklima.*" & lc!=".*kunstklima.*" & lc!=".*klimaanlage.*" & lc!=".*schulklima.*" & lc!=".*vollzugsklima.*" & lc!=".*betriebsklima.*" & lc!=".*vertrauenklima.*" & lc!=".*vranitzky-klima.*" & lc!=".*klimatisiert.*" & lc!=".*schulpartnerschaftsklima.*" & lc!=".*klimazone.*" & lc!=".*beratungsklima.*" & lc!=".*raumklima.*" & lc!=".*zusammenarbeitsklima.*" & lc!=".*koalitionsklima.*" & lc!=".*wahlkampfklima.*" & lc!=".*medienklima.*" & lc!=".*forschungsklima.*"])
- [lc=".*kreislaufwirtschaft.*"]
- [lc=".*lebensraum.*" | [lc=".*lebensräum.*"]]
- [lc=".*methan.*"]
- [lc=".*natur.*" & lc!=".*naturgemäß.*" & lc!=".*naturgegeben.*" & lc!=".*naturwissenschaft.*"]
- [lc=".*öko.*" & lc!=".*ökosystem.*" & lc!=".*öko-system.*" & lc!=".*ökonom.*"]
- [word="(Pariser|pariser)"] [word="(Klima|Klimaschutz)"]? [lemma="(Abkommen|Übereinkommen)"]
- [word="(Klima|Klimaschutz)"]? [lemma="(Abkommen|Übereinkommen)"] [word="von"] [word="Paris"]
- [lc=".*recycling.*"]

- [lc=".*ressource.*" & lc!=".*humanressource.*" & lc!=".*bildungsressource.*" & lc!=".*wissensressource.*" & lc!=".*geräteressource.*" & lc!=".*zeitressource.*" & lc!=".*sachressource.*" & lc!=".*arbeitskräfteressource.*" & lc!=".*hintergrundressource.*" & lc!=".*geldressource.*" & lc!=".*personalressource.*" & lc!=".*kammerressource.*" & lc!=".*finanzressource.*" & lc!=".*antwortressource.*" & lc!=".*natoressource.*" & lc!=".*verwaltungsressource.*" & lc!=".*sonderpädagogikressource.*" & lc!=".*aktivitätsressource.*"]
- [lc=".*rohstoff.*"]
- [lc=".*temperatur.*" & lc!=".*außentemperatur.*" & lc!=".*körpertemperatur.*" & lc!=".*kühlschranktemperatur.*"]
- [lc=".*treibhaus.*"]
- [lc=".*umwelt.*"]
- [lc=".*überschwemmung.*"]