Lund University
Department of Political Science

STVM24
Spring 2022
Supervisor: Rikard Bengtsson

Language at the Heart of Lobbying Dynamics - A Quantitative Approach to Interest Group Research

Marcel Florian Loehr

Abstract

Within the issue area of lobbying dynamics in the European Union, the measurement of interest group success is heavily debated among scholars. This research examines the influence interest groups had in shaping the Artificial Intelligence Act via the Public Consultation process. In contrast to existing literature, a computer assisted quantitative content analysis was used to gain insights into lobbying processes, thus avoiding influencing the outcome of the study through interviews with the investigated political actors.

The research design puts specific emphasize on the direct exchange of information as the key resource that influences success, thus analyzing the exercise of power rather than the bases of power. The findings thus have twofold implications for the grander lobbying literature. Next to the common empirical findings the aim of this paper is to elaborate upon the methodological challenges of interest group research and why policy position and therefore success of interest groups ought to be analyzed via a quantitative content analysis rather than interviews of lobbyists or public officials.

From an empirical standpoint, the findings suggest that the smaller coalition of civil society interest groups achieved greater levels of success than anticipated, but more importantly, the applied method showed that analysis of interest groups can happen via usage of algorithms, thus avoiding bias.

Table of Contents

1 Introduction

Attempting to decipher the black box of lobbying initiatives in the European Union has been a subject of much academic research yet findings remain ambiguous and methodological challenges arise. While there is a general agreement that similarity of interest group preference with the policy outcome indicates successful lobbying (Bernhagen et al, 2014:203), determining policy preferences of interest group remains difficult. Thus, most literature on EU lobbying is as much a case study of a certain policy debate as it is a discussion of methodology. The research will follow this structure and while analyzing the chosen case, critically evaluate methodological differences, benefits, and detriments. While findings of the case study have significance for the empirical aspects of lobbying, the case is purely exemplary to display how language can be operationalized in interest group research.

This research is based on the premise that recent changes in EU policymaking, in specific the public consultation initiative and the concept of **Good Governance**, provides researchers with the opportunity to pinpoint policy preferences of interest groups based on language. Previous research focused on expert interviews or analysis of financial and personnel resources, meanwhile the public consultation initiatives provide insights into interest groups policy preferences that can be operationalized through a Content Analysis (CA).

Those recent events in EU policymaking instigated the research to analyze the public consultation initiative "Artificial Intelligence – ethical and legal requirements" posing following research question:

> *How can language be operationalized to evaluate lobbying attempts of interest groups in the EU public consultation initiative "Artificial Intelligence – ethical and legal requirements"?*

Case studies are the norm in interest group research, but the pivotal point of academic literature are the methodological challenges of researching lobbying dynamics. The research question is thus framed in a way that allows to highlight the linguistic aspects of lobbying. Empirical findings regarding interest groups dynamics in the specific case are merely a byproduct of the methodological debate.

To answer the research question the paper will discuss the specific characteristics of lobbying dynamics in the EU. It will continue by introducing the importance of language in politics and critically discuss previously used methodology and how lobby success can be measured.

Based on the existing literature the research hypothesizes civil society actors will largely focus on ethical issues whereas business interest, focusing on the economic policy dimension, has greater levels of success in their lobbying attempts. Furthermore it is expected that the EC sides with the larger lobby coalition.

The theoretical foundation of the research is resource dependence theory (RDT) and the assumption that lobbying ought to be seen as a relation based on persuasion rather than financial exchange. In contrast to previous research, the evaluation of interest group policy preferences happens solely through public available data and is not distorted by bias or misrepresentation of the investigated.

The research concludes by restating its proposition to move away from focusing on interviews as a mean to analyze lobbying dynamics and rather operationalize policy documents of interest groups through the use of language.

2 Literature Review

In its attempt to gain insights into the effect public consultation has on European policy drafting, the paper will elaborate on concepts, existing literature, and methodology in its literature review. This structure allows for a thorough discussion of events and lobbying dynamics while giving room to expose the gap in current academic literature that neglects public consultation as well as quantitative textual analysis.

Initially researcher wants to emphasize its premise, namely how changes in EU policymaking allow access to yet largely unexplored datasets. In a next step the research will give an overview of lobbying dynamics within the European Union with a focus on organizational structure and the European Commission. To finalize the literature review the paper will discuss methodological concerns of existing research.

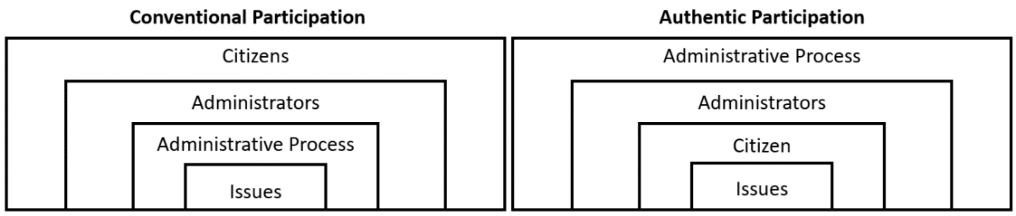
2.1 Public Consultation and Good Governance

Within recent European policy making history there are two developments of specific importance for the research. Expansion of the public consultation initiatives and the concept of **Good Governance**.

To understand EU lobbying processes the historical development of such is important. European level corporate lobbying was used by the EU to push the integration process, "greatest weight was given to those actors who were prepared to establish a "European Identity" through European alliances with rival firms and/or solidarity links with societal interest" (DG Internal Policies of the Union, 2007:10). Initially EU lobbying was mainly reserved to resource-rich business interest which led to questions of legitimacy and democratic values of the EU decision-making process, if "public policy is systematically biased in favor of some interests, while others are constantly losing, the democratic legitimacy of policy outcomes is greatly undermined" (Dahl, 1989:322-326).

The combination of lacking diversity of interest group participation and critique directed at decision-making processes culminated in "European Governance: A White Paper" (European Commission, 2001) , which introduced the concept of **Good Governance**. The white paper drew a line between **Governance** and **Good Governance** based on principles such as "openness, participation, accountability, effectiveness, and coherence" (European Commission, 2001:8). Furthermore, critique was voiced against the overreliance of EU institutions on business interest as a provider of technical expertise (European Commission, 2001:16), thus calling for greater inclusion of civil society and the need for political expertise (European Commission, 2001:13). Inclusion of a variety of interest groups became an "important asset on the EU's democratic balance sheet" (Coen, 2021:33). Therefore the EU public consultation process can be seen as a remedy for the overrepresentation of business interest by reducing access barriers to decision-makers.

The for the case study relevant public consultation has been largely neglected by scholars with exemptions such as Klüver (2009) and is situated in a vacuum between outside and inside lobbying strategies. It is a form of contact to private officials, but a one-way form of communication, thus also resembling outside lobbying strategies. In sum public consultation initiatives display a shift from conventional to authentic participation as illustrated below.



(Figure 1 King et al., 1998:320-321)

The expansion of public consultation initiatives does not only grant greater levels of access to decision-makers for members of the civil society but also gives researchers access to the contents of information exchange between interest groups and European Commission. This information is the very essence of the political sphere, to understand politics, one has to understand what political actors are saying and writing (Grimmer and Stewart, 2013:267).

2.2 Lobbying Strategies and the European Commission

Lobbying dynamics can be divided into two strategies, outside and inside lobbying. While outside lobbying refers to the use of public communication channels (De Bruycker and Beyers, 2018:57), inside lobbying is categorized as direct exchange between lobbyists and public officials (De Bruycker and Beyers, 2018:57-58). This concept of outside and inside lobbying is used by various scholars with slight changes of terminology such as voice (public policy strategies) and access (political bargaining on private venues) (Beyers, 2004:213).

In large scale surveys of public officials and lobbyists, Beyers concluded that issue areas that require specialized, and issue specific information are better suited for direct lobbying strategies, while diffuse interests with easy to grasp concepts are better suited for public discourse, thus outside lobbying is preferred (2004:234). There is not only a gap in research regarding the impact of public consultation on policy drafting, but the case of the AIA becomes of interest since it attempts to bridge between the rather complicated Information technology (IT) dimension and easier to grasp concepts such as data protection and human rights.

The focus of the research lies in determining the impact interest groups had on the legislative drafting process. While all EU institutions are target of lobbying initiatives, the EC, tasked with drafting policy, is the unit of analysis for the current research. Within the current literature there is general consensus regarding the resource dependence of public officials and lobbyists (Coen, 2007:334). The EC as the drafter of legislation needs reliable information and interest groups require favorable policy for their own survival. In comparison to other EU institutions, members of the Commission are not publicly elected thus depend less on public approval, which results to less influence of the public on the EC (Beyers, 2004:219). Furthermore, the proposal stage offers the most fertile opportunities for interest groups to influence legislation (Klüver, 2013:156), changing already drafted legislation requires greater intensity of lobbying efforts.

As mentioned, the EC demands external resources for its own functioning, those resources are provided by interest groups (Beyers, 2004:218). Findings regarding influence of interest groups on the Commission are ambiguous. The divide happens on what form of interest group manages to lobby more successful. Interviews with EU officials concluded that business interest gains privileged access to members of the Commission (Beyers, 2004:233-234, Dür and Mateo, 2012:970) while other research results indicate that "business actors are substantially less influential than is often perceived (Dür et al., 2015:960). Findings of the INTEREURO project conclude that "business often has to give way to consumers and environmental interests" (Dür et al., 2015:976).

It is important to note that some of the research uses access to decision makers as lobbying success while others focus on self-evaluation of lobbying processes by lobbyists to determine success. Interviews with lobbyists concluded that while NGOs do manage to get comparative good levels of access, there is no translation of access to lobbying success (Eising, 2007:331).

The researcher does not want to disregard the importance of the findings in academic literature but point out how differences in the definition of success, analyzed variables, and methodology can influence the results. Lobbying remains an area in which findings can rarely be generalized or taken for granted, and slight changes of contexts greatly influence results.

2.3 Resources and Success

Next to the lobbied issue area, the resources and form of organization, in specific civil society and business interests, are a large part of existing lobby literature. Especially when discussing the relation between resources and success findings remain ambiguous, which instigated the aim of the research to not focus on material resources but rather information exchange.

Measuring success has been the focal point of interest group research yet remains the most difficult aspect of exploring lobbying dynamics. While there is a broad consensus that influence aims at establishing "a causal relation between the preferences of an actor regarding an outcome and the outcome itself" (Nagel, 1975:29), determining an interest groups ideal point remains difficult. Before transparency initiatives of the EU, research claimed that measuring lobby success is almost impossible due to lack of publicly available evidence and public data bases (Beyers, 2014:175). A view that the researcher, in light of recent developments, disputes.

Lobbying efforts that do not result in preference attainment are not necessarily unsuccessful, if a lot of counteractive lobbying takes place, and success can be overstated if an interest groups preference is supported through other channels of influence (Dür, 2008B:1226). The importance in lobbying research thus lies in acknowledging that no single method is able to capture all variables that influence lobbying processes.

McKay's research based on interviewing representatives of interest groups concluded that "greater financial variables do not appear to help lobbyists' chances of achieving policy outcome" (2012:913) but that the intensity of lobbying efforts brings about greater success (2012:920). Those results are disputed by other research claiming a positive relationship between economic resources and success (Flöthe, 2019:175), due to the fact that the creation of knowledge is tied to financial resources of an organization. Coen supports this assumption declaring that there is a comparative advantage of business interest due to more resources (2007:335) and implies an "elite-trust-based" relationship between lobbyists and public officials who favor a long running relationship based on consistent information exchange (Coen, 2007:335).

Another variable explored in the literature are lobbying coalitions. Influencing decision-making processes is best understood as a collective endeavor of interest groups (Klüver, 2013:18). Gaining results of lobbying success thus ought to happen by looking at the grander picture of various interest groups seeking influence. Klüver concluded that the governing body is likely to side the larger lobby coalition (2013B:73). Notable when talking about lobbying coalition is that those coalitions do not need to be formally agreed upon but rather consist of interest groups with similar policy preferences (Baumgartner et al, 2009:6).

2.4 Methodology, and Bias

In the following the paper wants to briefly discuss the pitfalls of previously employed methodology. Assumptions were largely based on interviews and analysis of resources. Due to past barriers in accessing the concrete exchange of information between lobbyist and interest group, researchers had to resort to those variables to gain insights into lobbying dynamics, but due to recent changes in policy making research is now able to fill the gap in existing literature and perhaps explain the ambiguity in findings based on the "hidden" variable that is the direct exchange of information.

The two predominantly used methods were the "attributed influence" method (Dür, 2008B:565) and the "preference attainment" method (Dür, 2008B:566). Both methods have similar pitfalls, the former focuses on a self-assessment or peer-assessment of exercised influence for a specific policy proposal. This assessment can happen through interviews or surveys directed at lobbyists, public officials, or independent experts. Self-assessment and peer-assessment is heavily influences through bias and misrepresentation of lobbying processes (Dür, 2008B:568). Business interest attempts to downplay their influence to not undermine democratic processes in the EU and escape accusations of "buying" influence. Civil society actors are expected to complain about overrepresentation of business interest and lack of access to decision-makers. Public officials on the other hand aim at creating a picture of equal opportunities and influence of each participant in lobbying processes. The self-assessment is thus distorted by bias while peer-assessments largest detriment is the lack of information of other actors lobbying activities. The gravest inaccuracy in the attributed influence method is that it measures the perception of influence rather than actual influence.

The preference attainment method on the other hand attempts to determine an actors' ideal policy position through surveys and interviews, to then compare an actors' ideal policy position to the final policy outcome. Once again the harvested data is influenced through bias and misrepresentation. Furthermore, this method only allows to focus on recent lobbying initiatives since the time-lag dimension influences the interviewees opinion and knowledge of a policy agenda.

Both methods thus have similar pitfalls and are heavily influenced by bias, they are subjective in nature and drawing finite conclusion becomes impossible. Furthermore, it is common for response rates to be under 45% (Eising, 2009 or Dür and Mateo 2012) with various levels of access to different forms of interest groups (Dür and Mareo 2012 or Beyers 2004)

The Transparency Register of the EU gives detailed insights into lobbying processes and the ever-growing Public Consultation initiatives allow a more accurate definition of an actors' ideal point, the research goes so far as to claim that currently there is an "embarrassment of riches" (Benoit, 2020:461). The rapid expansion of publicly available data sets is supplemented with the growth and accessibility of computer based quantitative research, which allows a more cost and time efficient analysis than at any other point in history. Current and future research ought to use the hidden potential provided through transparency and technological development.

3 Theoretical Perspective

3.1 Change of Theoretical Perspective

Lobbying research is vast, and many different subjects have been covered, yet there are two factors that researchers struggle to operationalize, pinpointing the policy preference of interest groups and consequently measuring success. Traditionally lobbying happened behind closed doors and while data regarding "who lobbies whom and when" is publicly available through the Transparency Register of the European Union, the exact contents of those meetings remain unknown. It has been suggested that researching interest group influence is like "searching for a black cat in the coal at midnight" (Loomis, 1983:184) or that research produces case study after case study with no notable findings (Woll, 2007:58).

The researcher proposes that the shift of the EU to focus on transparency and participation must be accompanied by a shift in theoretical consideration of lobbying processes. In previous research, lobbying was mainly operationalized as an exchange of resources for favorable policy. In other words, researchers focused on analyzing the structure of Interest Groups (Dür et al. 2015 or Klüver, 2013), financial resources (McKay, 2012), personnel variables such as "flooding" governmental premises with lobbyists (Holman and Luneburg, 2012) or hiring revolving door lobbyists (Strickland, 2020). Rent-seeking is often used by scholars to analyze this kind of lobbying behavior yet using rent-seeking as a theory is accompanied by a negative connotation regarding lobbying. The theory claims that lobbying diverts resources from positive sum activities into zero and negative sum transactions, meaning that the cost of lobbying outweighs the potential benefit to society if resources were allocated elsewhere (Friedrich, 2013:289). Rent-seeking attaches a moral judgement to lobbying processes, thus engages in political philosophy rather than fulfilling the duty of political science to describe politics for what they are (Bitoni, 2017:21).

The great ambiguity in previous lobby research shows that variables such as resources, structure of IG, or personnel fail to bring about finite results of lobby success. The research believes that prior lobby research had to focus on those variables and used expert interviews simply since there was no other way to pinpoint policy preference of IGs and thus measure lobby success. The actual contents of information exchanged between Lobbyists and EU-officials remained imperceptible behind closed doors. While money can grant IGs access to legislators, influence begins when policy-drafters change their beliefs based on the information they are fed by interest groups (Hall and Deardorff, 2006:71), which is why the research proposes to put information rather than material resources at the heart of studying lobby dynamics. Lobbying should thus be seen as persuasion, a form of legislative subsidy (Hall and Deardorff, 2006:69).

3.2 Resource Dependence Theory

The public consultation initiative of the EU changed the knowledge researchers can acquire from lobbying dynamics. While previously data allowed to determine "Who lobbies Whom and When?", now researchers can know "Who says **What** to Whom?", thus an analysis of resources focused on information exchange rather material resource exchange can happen.

The paper wants to analyze the actual exchange of information thus employs RDT. Scholars established that institutions such as the EC depend on external resources for their own functioning (Bouwen, 2009:20), those external resources are information. Interest groups on the other hand have the main goal of survival, RDT proposes that organizations attempt to create an environment that is beneficial for their own survival, which is done by shaping government regulations (Hillman et al. 2009:1411). They achieve this by subjectively deciding which resources are important for organizational maintenance and engage in strategic behavior to obtain these resource (Braun, 2015:143). In the case of the public consultation process, those resources are information that potentially lead to favorable policy. Due to the fact that participation and thus access to decision-makers is granted without any membership cost, information is the only resource interest groups need to gather and therefore the only necessary variable for analysis.

In sum, neither the EC nor interest groups are fully self-sufficient and interaction between both needs to take place to attain their goals, namely policy drafting and survival. Yet according to RDT all organizations pursue a twofold aim, to reduce being subject to exercise of power and exercise power themselves (Pfeffer, 1987:26)

The interaction of public and private organizations can be conceptualized as a series of inter-organizational exchanges. The organizations involved in the exchange make an implicit or explicit cost benefit analysis on the basis of which they decide with whom to interact (Bouwen 2002:368). Those exchange relations are only durable when all sides receive benefits from the interaction. Benefits do not need to be equal (Bouwen 2002:368), but the result is a positive sum transaction. The public consultation process is a hand on example of resource exchange. Interest groups provide access goods in terms of expert knowledge and in exchange the EC drafts policy proposals influenced by this knowledge.

This information exchange between lobbyists and public officials happens through the use of language. Monroe and Schrodt define texts as the "most persuasive and certainly the most persistent artifact of political behavior (2008:351). The research will thus analyze documents submitted by IGs to gain reliable information regarding policy positions of IGs. The main challenge is thus to make sense of the language used by interest groups, through the transformation of texts into quantitative data. The "target concern is not what the text contains but what its contents reveal as data about the latent characteristics for which the data serves as an observable implication" (Benoit, 2020:466). In contrast to previous research focusing on studying the bases of power (e.g., financial and personnel resources) the paper will focus on the exercise of power (Lowery, 2013:3), in specific the persuasive aspect of knowledge that is transmitted through language.

4 Methodology

4.1 Research Design

The research design attempts to answer the methodological question as to how language can be operationalized to get insights into EU lobbying dynamics. In order to so, the research employs a variety of algorithms that all use words as signals that help interpreting the data. Uncovering those processes will happen through three steps.

First the different policy dimensions will be displayed through the use of Latent Semantic Scaling and Wordfish. Further, it attempts to uncover the policy position of interest groups with the usage of the Syuzhet dictionary and Wordfish. Finally it attempts show the actual impact interest groups had on the policy draft of the AIA by measuring the Jaccard similarity of submissions and the draft. The reasoning behind those three steps is to first understand the policy issue, then to categorize interest groups based on preference, and lastly measure influence.

4.2 Epistemology and Ontology

Due to its choice of theory, the research proposes that the information exchange between interest groups and public officials is the key variable when analyzing political bargaining in the EU. Language is thus crucial to determine ideological positions of political actors. Following this line of argumentation, the methodological choice is a Content Analysis (CA), which in its essence attempts to answer the question "Who says What to Whom?".

CA attempts to uncover reality as it exists and is considered a positivist, objective, and quantitative approach. CA aims at systematic analysis which is often carried out through the testing of hypotheses; thus the paper will return to the hypotheses established in the introductory part of the paper when discussing its findings.

Language used in the public consultation submissions can be operationalized through counting and coding which then allows the research to estimate policy position based on information exchange. At this point the public consultation process is the only form of lobbying that grants researchers insights into the information exchange between non-governmental political actors the European legislative drafting apparatus. The methodological framework the research attempts to create is based around the premise to be free of from bias of the investigator and the investigated, furthermore, it follows the purpose of generalization without access barriers such as low interview response rates.

4.3 Case Selection

The research chose the public consultation "Artificial Intelligence – ethical and legal requirements" as its case study. Important for the case selection is that the consultation process is closed, and a policy document has been drafted to allow a comparison of submissions with the final policy output. The research only aims to establish a relation between policy drafting and public consultation, thus the policy does not need to be adopted by the EU institutions but only drafted by the EC.

The consultation initiative aims at finding a balance between "tremendous opportunities" of AI and risks connected to fundamental rights, safety, and liability (European Commission, 2020). The consultation process thus represents an example of the well-discussed opposition between business interest and civil society.

While the paper aims to provide a methodology that can be applied to a great variety of public consultation initiatives, there are certain criteria that optimize computer assisted CA. Since the main assumption of the research is that language determines ideology, computer assisted CA works best when participants of questionnaires have great freedom of language and can put emphasis on certain topics. The case of the Artificial Intelligence Act gives no guidelines to participants regarding length, structure, or topics thus is a great fit for applied methodology. Close-ended, dichotomous, or Likert-scale questionnaires are not fitted for the chosen methodology. Furthermore, the accuracy of results provided by the algorithm increase parallel to the size of the dataset, the AIA consultation received a total of 133 individual submissions, which provides the research with a large enough dataset to gain reliably results.

While the research uses the AIA as a case study, the research design can be applied to every EU public consultation process that follows a similar structure and received enough feedback by non-governmental actors.

4.4 Data Selection

As previously stated, the initiative "Artificial Intelligence – ethical and legal requirements" received a total of 133 individual submissions, since the paper is especially concerned with lobbying processes in the EU, the paper will only include feedback submitted by organizations that are registered in the EU Transparency Register, thus excluding submissions of private citizens and organizations that did not follow EU guidelines for official participation in lobbying processes.

Furthermore, the research is limited by the language knowledge of the researcher itself as well as the algorithm, that can only analyze documents that are uniform in their language. While submitted documents in other languages than English could have been translated, certain words and intentions of the author would be lost, thus a potential misinterpretation based on translation could occur, which is why those documents were excluded from the analysis.

One document has been excluded based on its length, it included ~100.000 words, thus greatly exceeded the median number of words in other submissions and would have shifted the weight and frequency of certain words. A further document has been excluded for providing information that is AI related but did not give an opinion regarding ethical or legal requirements as asked by the EC public consultation.

After filtering the documents with above mentioned criteria, the research is left with a total of 92 individual submissions by interest groups.

Next to the interest group submissions, the policy proposal of the EC is used in proposed method. The comparison of submissions and policy proposal happens in the third part of the research design, in which the Jaccard similarity is determined.

4.5 Data Processing

The usage of computer assisted content analysis requires a preprocessing of data that is carried out in MS Word, MS Excel, and RStudio.

All documents are transformed into MS Word where the researcher spell checks and makes the language uniform. For the analysis all documents were transformed into American English. Furthermore, names, self-references, headers, footers, enumerations, and bullet-points were manually removed.

Next, all documents were put into an excel spreadsheet which allows the data to be loaded into RStudio were a more thorough processing of texts takes place. To carry out textual analysis with RStudio, all documents were tokenized, which refers to the process of splitting text into tokens, in this case a token is a word, which is the semantically meaningful component of textual analysis.

The next step consists of a normalization of the corpus of documents, in which the program removes stop-words, numbers, puts words into lower case, and stems the words. While reducing a word to its word-stem, the meaning of the word remains, and the size of vocabulary gets reduced to ensure greater comparability of documents.

After processing, the data is transformed into a Document-Term-Matrix that displays each token (or word) an interest group used in their submission. The removal of unnecessary information through MS Word and RStudio resulted in a total of 136053 words and 4292 individual words left for the analysis.

The policy draft of the EC undergoes the same document processing steps, but certain parts of the proposal have been cut for the Jaccard similarity measurement. Examples of cut parts are subheadings explaining the legal competences of the EU (e.g., Heading 2. Legal, Basis, Subsidiarity, and Proportionality) or elaborating on the context of the proposal (Heading 1), those subheadings do not give information regarding contents of the legislation and are thus not important when attempting to measure success of interest groups. For the measurement of Jaccard Similarity and consequently lobbying success only contents under subheading "Proposal for a regulation of the European Parliament and of the Council, laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts" were included.

4.6 Description of Applied Methods

Computer assisted CA is not a new method (Klüver, 2013 or Watanabe, 2021), yet especially in interest group research its value has been neglected and only few scholars make use of it. While current lobby research is divided by data selection, object of interest, or the variable to be analyzed, the aim remains the same, measuring influence and success. It is generally accepted that the policy preference of interest groups, compared to the final policy outcome, allows to measure lobby success (Bernhagen et al. 2014:207), yet accurately defining the policy preference of an interest group remains difficult. For this reason, the paper believes that the analysis of lobbying processes ought to start at the language and ideology interest groups display in their policy proposals.

Computer assisted CA can be carried out through various methods such as counting and dictionary approaches, unsupervised, and supervised machine learnings. The paper will thus employ a variety of methods to cross-validate and increase reliability of findings.

4.6.1 Supervised and Unsupervised Machine Learning

Supervised and unsupervised machine learning focuses on evaluating the entirety of the text and defining its policy position by comparison to other documents. The algorithm learns patterns and how to code a text. Supervised machine learning is guided by example texts, in the case of the AIA one that advocates mainly for fundamental human rights and risks AI poses for humans and one that advocates for the business side of AI, in specific economic progress and simplification of workflows.

Simplified a supervised machine learning algorithm works like following syllogism:

> If the algorithm is fed the information that “All basketball players are tall” and that “Dirk Nowitzki is a basketball player” it concludes that “Dirk Nowitzki is tall”

Unsupervised machine learning on the other hand does not require pre-defined coding rules or training data but the algorithm recognizes specific patterns in the given data sets and defines the policy positions of texts based on comparison. The research uses the program “Wordfish” as its tool for unsupervised machine learning (Slapin and Proksch, 2008). Wordfish uses the concept of word weights to differentiate between policy positions. A high word weights means a regular occurrence of words in some but not in other texts. The concept of word fixed effects on the other hand accounts for words that are used often in all documents. Word fixed effects thus do not indicate an ideological standpoint but are rather general terms used to explain the subject of policy. To contrast both concepts, high word weights indicate a clear ideological orientation and fixed word weight are descriptive of the policy proposal, which allows Wordfish to differentiate different policy positions.

The results of the computer assisted CA can then be compared to the drafted policy proposal by the EC. The research thus managed to define policy positions of interest groups based on the focus on language and can display how participation in the public consultation process helped shaping the drafted policy proposal.

4.6.2 Counting and dictionary approaches

Counting and dictionary approaches are similar to hand-coding and allow for sentiment expression of actors or analysis of specific words and their frequencies. The main difference to hand-coding lies in the accuracy and removal of bias. While the researcher still needs to read the public consultation documents and understand their meaning, the human part of machine learning lies in data selection, thus human judgement is skipped in the analysis of data, and then comes back to make sense of the results provided by the algorithm (Benoit, 2020:473). In specific the paper will employ Latent Semantic Scaling (Watanabe, 2021) and the sentiment dictionary Syuzhet (Jockers, 2020).

Latent Semantic Scaling refers to semi-supervised machine learning “that takes a small set of polarity words as “seed words” to assign polarity scores to other words in the corpus; it estimates semantic proximity of words employing word-embedding techniques” (Watanabe, 2020:82). Seed words are terms that are connected to the different policy dimension, in the case of the AIA the researched policy dimensions are ethical and economic issues. While the researcher choses the seed words, the selection does not happen arbitrarily but through the use of a dictionary embedded in the algorithm. The embedded dictionary "Data_Dictionary_Ideology" analyzes the corpus of text and suggests words on a left-right political ideology scale, which are then used to compute polarity scores of words within their local context. Employing Latent Semantic Scaling allows the researcher to get insights into the different policy dimensions and the divide between interest groups when lobbying the EC.

The Syuzhet dictionary is a sentiment dictionary developed in the Nebraska Literary Lab. It is based on 165.000 human coded sentences that assigned a sentiment to the words within the sentence. For example, words like "risk" receive a negative sentiment score whereas words like "benefit" receive a positive sentiment score. Applying the Syuzhet dictionary to the interest group documents will then result in a sentiment expression of each submission. The goal of the sentiment analysis via Syuzhet is thus to gain insights into the policy positions of interest groups based on the language they employ regarding artificial intelligence.

4.6.3 Jaccard Similarity

Jaccard Similarity measures the similarity between two texts in specific this measurement examines commonality of words (Mullen, 2020). It is an intersection of two documents divided by the union of those two documents that refer to the number of common words over a total number of words. Mathematically represented the Jaccard Similarity is calculated as in the following:

$$J(doc_1, doc_2) = \frac{doc_1 \cap doc_2}{doc_1 \cup doc_2}$$

Applying this equation to the current research, the Jaccard Similarity between the policy draft of the EC (doc_1) and the interest group submissions (doc_2) will be calculated. It is of specific interest to analyze the divide between economic and ethical issues of the AIA, thus interest groups will be clustered into types of organization and then compared to the policy draft. The various types of interest groups participating in the public consultation process are NGOs, Corporations, Trade Unions, Consumer Organizations, Research Institutions, Public Authority, and Other. For each form of organization a Jaccard Similarity will be calculated to draw conclusions regarding success based on organizational structure.

4.6.4 Keyword-in-Context

The researcher applied a variety of methods to gain insights into lobbying dynamics. A computer assisted CA is only fruitful if the researcher has read and understood analyzed texts. Different methods can and will lead to slight variations in results that need to be explained by the researcher. In order to do so the Keywords-in-context (KWIC, Benoit et al. 2018) program will be used to give greater insights into potential deviations of findings that allow for explanation of such. The KWIC program allows the researcher to filter the texts by specific terms and displays in which context they were used by interest groups. The KWIC program is thus supplementary to each of the previous discussed algorithms.

4.7 Potential limitations of the research design

Lobbying and interest group research remains a rather obscure process and gaining valid insights is tricky. Due to the various ways interest groups attempt to influence the policy drafting process, drawing finite inferences between public consultation and policy output is impossible. There are certain factors that certainly contribute to the policy drafting but cannot feasible be measured such as impact of private lobby meetings or public opinion regarding the issue area.

The research explained earlier that algorithms work best if the data set is extensive and if interest groups have great freedom in structuring their texts and are thus allowed to use linguistic finesse to express their policy position. The greatest strength of the algorithm is also the greatest limitation of the research design. Interest groups submissions are not coherent in the way they discuss artificial intelligence thus differences in structure, emphasize, or writing style can bring about varying results by the algorithm. The paper will further discuss those issues in its data presentation.

Nonetheless, the goal of interest groups remains the same, to have legislation drafted that is as similar as possible to their policy preference. While the exact impact of the individual submission towards the policy draft cannot be measured, a measurement of preference attainment can happen with the help of previously explained research design.

5 Analysis

The analysis part of the research will present and discuss the findings computer assisted CA can provide. The analysis will begin by presenting dictionary approaches to CA to give a brief overview of themes within the AIA and how interest groups attempt to influence the drafting process of the Commission. In a next step, the research will return to the pivotal point of lobbying research and present how computer assisted CA allows a determination of policy preference based on language. Lastly the research will discuss the actual impact interest groups had on the drafting of the AIA.

5.1 Presentation of Data

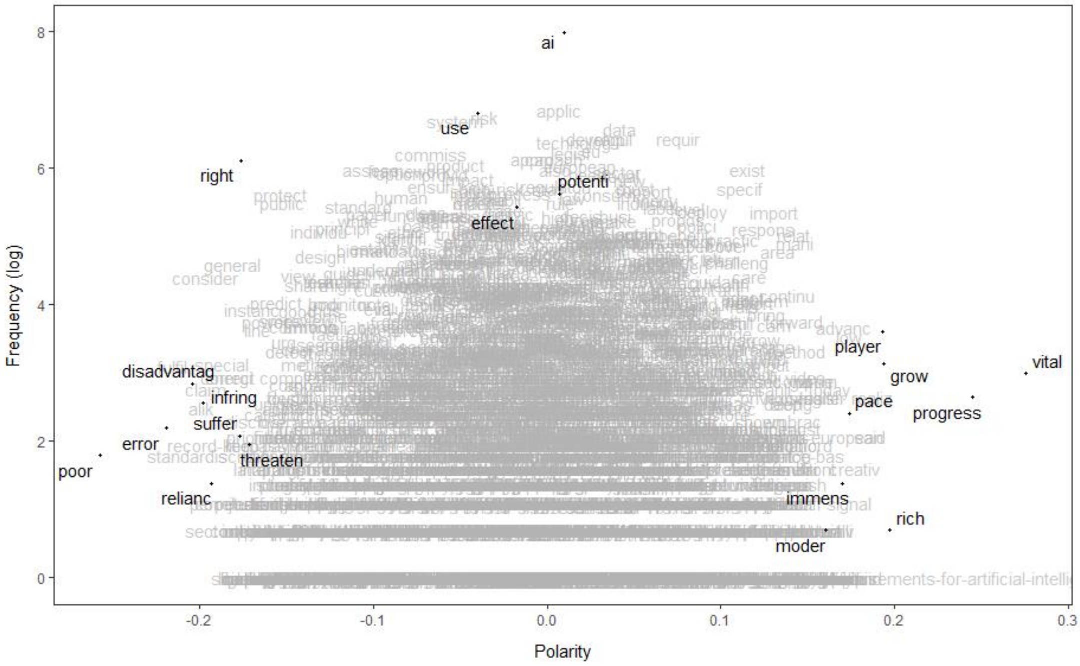
5.1.1 Ethical and Economic Issues

As previously mentioned, the Artificial Intelligence Act is not a single dimensional policy issue but rather an example where the EC is expected to perform a balancing act between accommodating ethical issues and economic opportunities. In a first step the paper will combine dictionary and semi-supervised machine learning approaches to display the divide between ethical and economic issues in the public consultation process. Latent Semantic Scaling (LSS) allows the determination of polarity scores within a data corpus, in other words and applied to the current case, display keywords used by interest groups in the chosen policy debate that either advocate for ethical issues or economic opportunities.

Figure 2 is the result of LSS based on sentiments given to words, negative polarity scores are associated with a negative sentiment regarding AI, thus highligting the ethical issues, while a high polarity score indicates to focus on the economic opportunities AI can provide.

The results show that the public consultation of the AIA experiences a great divide in policy preferences of interest groups. Words with a neutral polarity score and high frequency include **ai, use, effect, and potenti** (the words displayed are stemmed). This indicates that all interest groups recognize the potential and usability of artificial intelligence within the European market, yet they diviate in their opinion on the impact AI has on citizens and the economy.

The interest groups focusing on ethical issues are concerned that AI will threaten the already disadvantaged social groups and result in greater divide of rich and poor members of society, thus focusing on human rights issues that can either be created or strenghtened through the employment of AI. On the flipside of the coin there is a large representation of business interest that sees great improvements in workflow and considers AI vital for the competitiveness of the Union. Business Interest participating in the public consultation process believes that AI is a necessary tool to ensure future growth and result in economic progress.



(Figure 2 Latent Semantic Scaling)

This part of the analysis does not allow the research to draw conclusions regarding the impact interest groups had on lobbying the AIA but is necessary to show that the EC received submissions for both policy dimensions they requested in their call for public consultation. Thus far the research could only assume that interest groups were divided on ethical and economic issues but through LSS this multi-dimensionality can be utilized in further parts of the analysis to draw conclusions regarding success of certain interest groups when lobbying the EC.

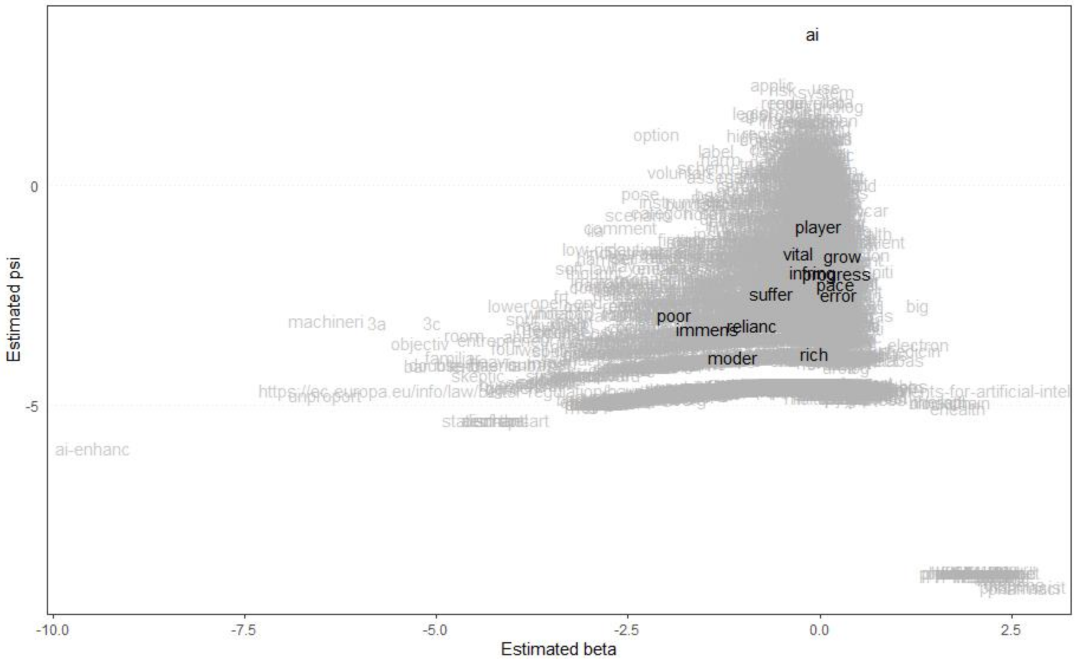
The research focuses on utilizing different programming techniques to estimates the importance of language in the public consultation process. It thus does not rely solely on one method to support its findings, but attempts cross-validate them through different algorithms. LSS displays policy dimensions based on a dictionary approach while Wordfish estimates policy dimensions based on ideology employed in interest group submissions.

Figure 3, stemming from the Wordfish algorithm, confirms the results of the LSS method. The algorithm recognizes that words such as **AI** are the most frequent used and do not give any indication about policy position, thus scoring a high word fixed effect and a neutral word weight. It further evaluates that words such as **poor, suffer, or reliance** receive a negative score which indicates that interest groups using those words see potential risks in the usage of artificial intelligence. It also estimates a positive beta towards words such as **grow, progress, or pace** which is used predominantly by interest groups that advocate for a more widespread and less regulated use of artificial intelligence.

Two unrelated computer algorithms thus produce similar results in the evaluation of the public consultation submissions which increases the reliability of the findings. Yet the results also highlight the pitfalls of relying only on algorithms for computer assisted content analysis. Certain terms such as **moder** (stemmed version of words such as modern) or **error** have been evaluated differently by both algorithms. For the word error, dictionary analysis calculates a negative polarity score thus associating it with interest groups advocating for more regulations and a more ethical approach, while the unsupervised program Wordfish gives it a positive word weight and associates it with interest groups advocating for the expanded usage of AI. This shows how a sole reliance on algorithms for CA can bring about deviating findings, thus a thorough reading of the submissions needs to happen by the researcher to evaluate how those variations occur.

The word **error** has been used by different interest groups in different contexts which results in an ambiguous evaluation of the algorithms. The Keyword-in-Context function has thus been used by the researcher to analyze certain words and their different meanings based on context. The NGO "Electronic Privacy Information Center" (IG #37) used the word error in the context of highlighting flaws in AI facial recognition techniques that discriminate against non-white people, while the company VDMA (IG #85) used the word **error** in a context of how AI is able to spot errors in human workflow that would not be recognized by human oversight. The usage of the same word in different context thus led both algorithms to draw different conclusions.

The assessment of multiple policy dimensions through the use of computer assisted CA shows that generally there is a high accuracy of findings and that different programs bring about similar results, but an algorithm alone cannot replace the duties of the researcher to ensure accuracy of findings. When carrying out computer assisted CA the researcher needs to be aware of deviations and explain those variances through own knowledge of the documents.



(Figure 3 Wordfish Analysis of Words)

5.1.2 Policy Preferences

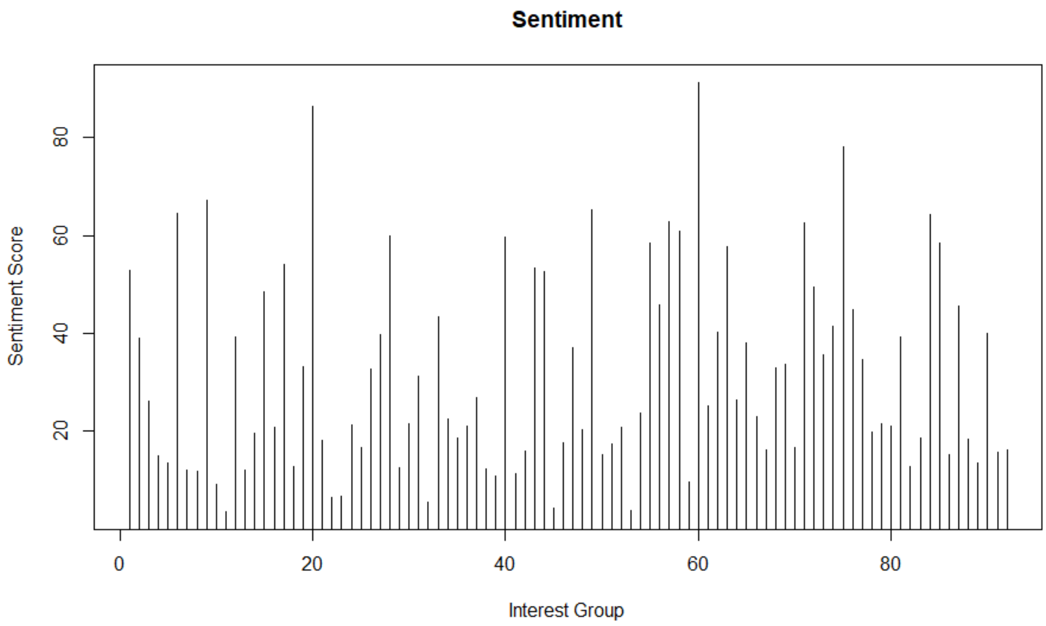
In a next step the research discusses how policy preference of an interest group can be determined via computer assisted CA. By putting language at the heart of its analysis, the paper can operationalize words to determine policy positions and consequentially lobbying success. The research will cross-validate its findings via two different methods and compare their significance.

Figure 4 displays sentiment scores of interest group submissions based on pre-existing dictionaries. Those pre-existing dictionaries embedded as code in R, attach a sentiment to words used in the public consultation submission that will result in an overall sentiment of the submission. High sentiment scores indicate that interest groups use a great number of "positive" words (e.g., trust, voluntary, potential etc.) and advocate for the expansion of AI usage and fewer regulation. Low sentiment scores on the other hand are connected to AI skepticism with usage of words such as risk, harm, or ban. Those interest groups generally advocate for stricter regulation and highlight the risks attached to unsupervised deployment of AI systems.

To exemplify this, the interest group BusinessEurope (IG 20) scores very high on the sentiment scale and expresses are largely positive view on artificial intelligence thus receiving a relatively high sentiment score. The NGO FairTrial (IG 50) on the other hand is especially concerned with human rights issues, expressing a largely negative sentiment towards AI thus receiving a low sentiment score.

The results of the sentiment analysis show that there is a very mixed response regarding artificial intelligence in the submissions and interest groups are divided in their expressed sentiments.

The paper hypothesized that business interest will express a largely positive sentiment and civil society a largely negative sentiment regarding AI. The sentiment analysis shows that this division lobbying literature draws between business interest and civil society is not entirely accurate and a potentially outdated model of describing lobbying processes. With growing levels of corporate social responsibility, business interest can side with consumer interest rather than purely focusing on economic progress. For example IG 66 or 53 are corporations yet receive a low sentiment score. The findings thus refute the hypothesis that business interest is expected to only focus on economic aspects of AI usage and mentioned interest groups are exemplary for advocating a balancing act between economic prosperity and consumer interests.



(Figure 4 Syuzhet Sentiment Analysis)

The sentiment scores need to be taken with a grain of salt, the results displayed can only evaluate how positive or negative the submissions are written. The EU proposed various options on how to deal with the issue of AI, some interest groups might strongly respond and put emphasis on an option (e.g., strong regulation of AI) they refuse rather than advocating for the option they prefer. Sentiment analysis on its own can solely express the great divide in terms of policy positions and that the AIA is an example for a clash between interest groups with specific views on the regulation of artificial intelligence.

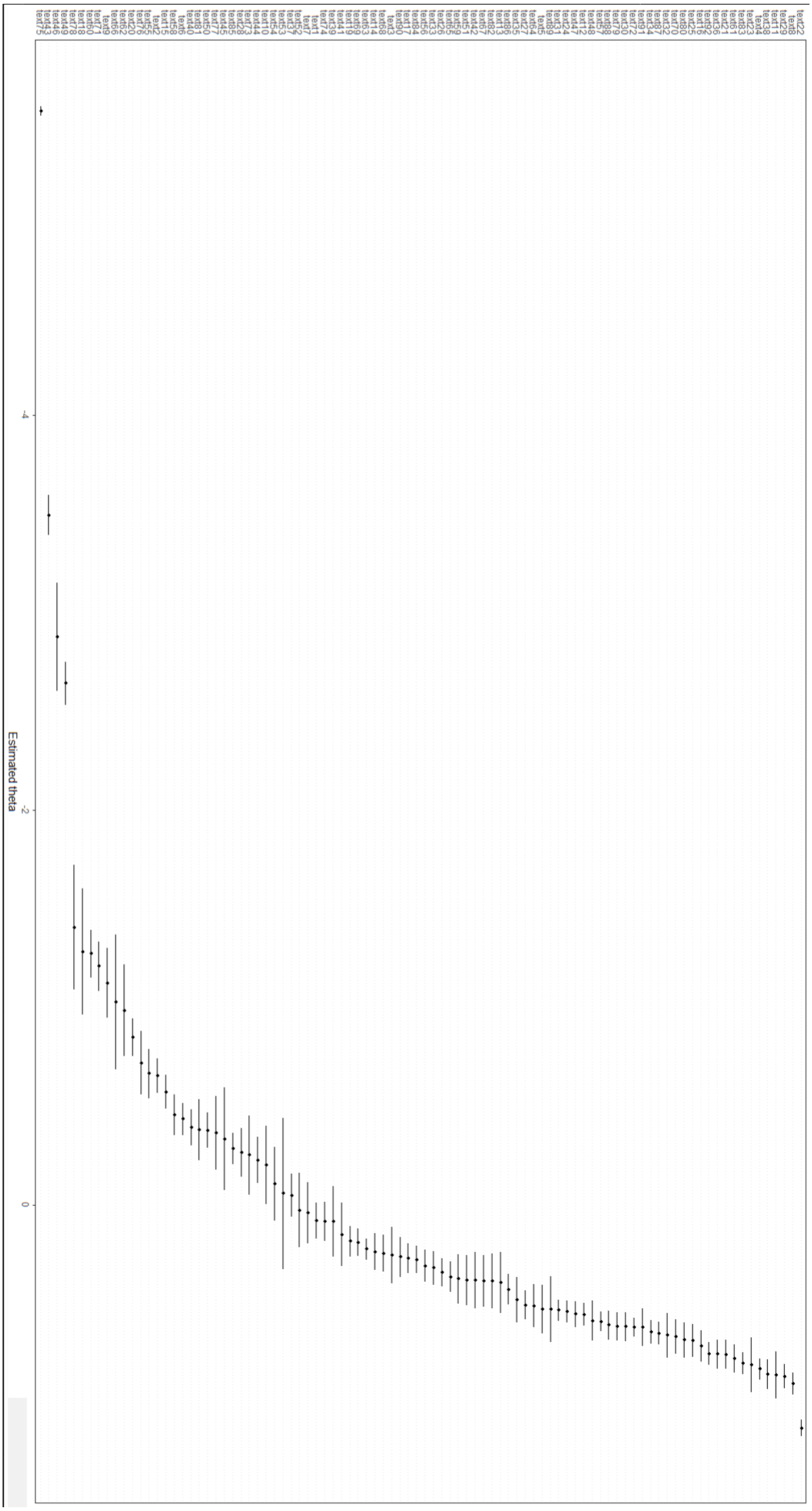
While dictionary-based approaches can estimate the sentiment given in analyzed texts, the program Wordfish estimates policy position through an unsupervised algorithm (Figure 5). The data confirms the researchers' expectation that NGOs, Public Authorities and Consumer Organizations that participated in the public consultation initiative have a skeptical view on artificial intelligence. 10 out of 14 submissions from NGOs, Public Authorities, and Consumer Organizations received a negative score by Wordfish. Overall a large part of interest groups was estimated with a positive score, which confirms the expectations of the research due to the large part of the submissions being made up by business interest.

The public consultation process did not give strict guidelines to interest groups regarding how to structure their submissions, which resulted not only in very different documents that need to be compared but also in deviation of research findings depending on which algorithm was used. As previously discussed, interest groups are either expected to focus on the ethical or economic side of artificial intelligence usage and had different ways of expressing their policy position.

It is important to note that a sentiment analysis, as well as the Wordfish algorithm are capable of allowing the researcher to draw conclusions regarding policy positions, yet this cannot happen without a thorough reading of the texts by the researcher itself.

Different use of language by interest groups can lead to different results of the sentiment or Wordfish analysis. For example, an interest group can oppose heavy regulation of AI, which would result in a low sentiment score, yet the Wordfish algorithm would categorize opposition of regulation with a positive theta score, thus knows that the opposition to heavy regulation means a pro-AI stance. Both programs thus allow the researcher to draw conclusion with the content of the submissions in mind but are not comparable to each other. The assumption that an interest group that received a low sentiment score should also receive a low Wordfish theta score can thus not be made. Rather the researcher needs to evaluate the findings based on its own understanding of the texts.

What the Wordfish results do indicate is that there is a large coalition of interest groups that advocate for expansion of AI usage and prefer fewer regulations that result in economic benefits. Opposing this large lobbying coalition is a minority of interest groups advocating for strict regulation of AI and protection of EU citizen rights. Based on the literature review it is thus expected that the EC ought to swing in the direction of the larger lobby coalition. The accuracy of this hypothesis will be tested via the Jaccard Similarity in the next section.



(Figure 5 Wordfish Policy Position)

5.1.3 Jaccard Similarity

The final data the paper wants to present is the Jaccard Similarity of public consultation submissions and the policy drafted by the EC. As previously established, an interest groups lobbying efforts are considered successful if the policy output aligns with the policy preference of an interest group. The Jaccard index determines the similarity between text documents by measuring common words of documents with the total words within the documents. This form of measurement is based on the researchers proposition that similarity of policy draft and interest group submission indicates success.

Using this method, the research focused on the different forms of interest groups (NGO, Business Interest, Trade Union, Public Authority, Consumer Organization, Research Institute, and Other), and how much of their submitted input was re-used in the policy draft. The Jaccard similarity measures documents on a scale from 0 to 1, 0 meaning texts do not share any common characteristics and 1 meaning the texts are identical.

Type of IG	Number of Submissions	Jaccard Similarity
NGO	10	0.3438045
Business Interest	67	0.2621985
Consumer Org.	3	0.3543506
Public Authority	1	0.252541
Research Institution	7	0.3422274
Trade Union	3	0.3583123
Other	1	0.1400844

(Table 1 Jaccard Similarity)

The results of the Jaccard similarity test show that documents submitted by NGOs have a greater resemblance to the draft of the EC than documents submitted by business corporations. Thus opposing the hypothesis that business interest has an advantage in lobbying the European legislative drafting process. NGOs, Consumer Organizations, Research Institutions, and Trade Unions received a similar score while Business interest, Public Authority, and Other scored considerably lower. In the case of Public Authority and Other, this can be explained through the significantly lower number of submissions. Business interest on the other hand had the largest number of submissions yet the documents submitted show a low similarity to the policy draft.

The results of the Jaccard similarity test are of special interest regarding the hypotheses the research established. First it refutes the hypothesis that business interest is expected to lobby more successful based on the resources they possess, which goes in line with the theoretical foundation of the paper that language ought to be the unit of analysis when determining interest group influence rather than material resources.

Furthermore, business interest makes up the majority of public consultation submissions yet has seemingly less influence than other forms of organizations that submitted less policy statements. Thus the Jaccard similarity test indicates that the size of the lobbying coalition does not affect preference attainment of interest groups.

It is important to note that textual similarity does not indicate attained policy preference per se. Interest groups discuss a variety of concerns in their submissions, it is thus possible that a low similarity score does not equal no influence, but a higher Jaccard similarity score indicates that a greater number of arguments brought forward by interest groups were inherited by the EC in their policy draft. It further does not consider other streams of influence such EC policy position prior to the public consultation or public opinion on the issue area. While those are important factors that influence the drafting process, they are neglectable when attempting to measure lobbying success. An interest group is considered successful if their policy position aligns with the policy draft, thus similarity indicates success unrelated to other variables.

5.2 Discussion

In the previous section the paper presented its data and explained what certain statistics are able to show and where their pitfalls are. The research thus critically engaged with its datasets and attempted to draw a fair picture of what the data can represent and what it cannot. Now the paper wants to turn back to established lobbying literature and theory to discuss methodological and empirical implications of computer assisted CA.

Returning to the research question, the paper proposed a method that allows an operationalization of the public consultation process based on language, thus avoided bias of any kind, which is the main criticism the research voiced against the current academic literature on lobbying.

The findings show that the statement of lacking publicly available data to carry out interest group research (Beyers, 2014:175) is dated, thus a methodological shift ought to happen that moves away from the reliance on interviews and surveys. The data provided through public consultation gives detailed description of policy positions that should be placed at the heart of interest group research. This shift in methodological considerations must be accompanied by a shift of theory on lobbying processes that views lobbying as an exchange of information rather than an exchange of resources (Hall and Deardorff, 2006:69).

While the researcher showed that with the help of computer assisted CA policy issues can be dissected into themes, policy preference analyzed and policy success evaluated, it barely scratched the surface of the seemingly endless possibilities of incorporating algorithms into political science research. The findings of computer assisted CA differ from findings based on interviews in the way they allow to explain the political arena. Previous findings often indicate how content interest groups were with the drafted policy (Dür, 2008B:566), rather than how similar drafted policy was with the initial policy position of interest groups.

From an empirical standpoint, the results were similar to the expectations of the researcher, certain arguments within the academic literature were confirmed while others were disputed. For example, the paper confirmed the statement that business interest often has to give way to societal interest (Dür et al., 2015:976) but disputed the claim that the decision maker tends to side with the larger lobby coalition (2013B:73). The explanation for this is twofold. Context matters in interest group research; a variety of policy issues is accompanied with a variety of influence streams towards the policy drafting process of the EC. While findings can be used to hypothesize about future research, each case of interest group research is unique in terms of actors, lobbying strategies, public opinion etc., thus a generalization of findings derived from a single case study cannot happen. The second explanation results from the very nature of previous research and the ambiguity within it. There is no consensus of methodology in interest group research which results in ambivalence of findings. Furthermore, those findings are not free from bias or misrepresentation of interest groups or public officials.

Usage of different algorithms led to isolated instances of varying results, while those instances were expected by the researcher, they are explained through the own understanding of the researchers reading of the texts. Thus, human judgement is skipped throughout large parts of the analysis but is necessary to explain singular deviations. Applying this form of methodology puts the researcher in a supplementary role to algorithms. Findings need to be interpreted and deviations explained, while the analytical part relies on the accuracy of the algorithm, it is thus free from any bias by either the researcher or the subjects of the research.

While the research can be seen as a case study of the AIA, it is mainly a discussion of lobbying concepts and methodology. It ought not to be seen as a total disregard for variables analyzed by previous researchers such as finances, personnel, or type of interest groups, but rather an expansion of existing literature due to changes in European policy making. With the expansion of available data in forms of texts, comes a responsibility of academic research to find a way of operationalizing those texts and draw conclusions based on the information exchanged.

While negating the bias of previous research, the results have similar pitfalls, public consultation is only one of many streams influencing policymakers, it thus does not take into account public opinion, private lobby meetings, or pre-existing policy opinion of the European Commission prior to the drafting process. Due to the scope of the research, those variables cannot be analyzed, but it is important to acknowledge that public consultation is not the only influential variable on the policy draft.

Thus far lobbying research was focused on analyzing one potential stream of influence on EU decision-makers, a potential for future research lies in a combination and comparative analysis of different streams towards the same policy proposal. An emphasize was put on the ever-growing focus on transparency and participation by the EU through **Good Governance** and public consultation, which resulted in a largely positive picture drawn by the researcher regarding those processes. The crucial point allowing for the potential of future research will be the continued expansion of transparency processes by the EU that not only allow insights into public consultation processes but also private lobby meetings.

Most importantly the research displayed a methodological approach that is replicable to all public consultation processes that are similar in structure to the AIA. Yet the research design barely scratched the surface of state-of-the-art algorithms that help in political analysis. The researchers own understanding of coding can thus be the greatest detriment or potential of computer assisted interest group research. The researcher does not have the pretension as to claim that the presented methodology is the only or most effective way to carry out lobbying research but solely aims to display the benefits of computer assisted CA when researching lobbying dynamics. Only previous explained expansions of EU transparency processes allow political science researchers to draw finite conclusions regarding lobbying success that is based on putting language at the heart of lobbying dynamics.

6 Conclusion

The researched pursued two goals, showing that language ought to be the main variable when analyzing lobbying dynamics, and that the use of computer algorithms helps establishing findings free from bias of the investigated. While voicing critique towards the multitude of case studies that focus on either expert interviews or financial capabilities of interest groups, the research proposed an approach that allows the analysis of policy proposal of interest groups based on language. Important to acknowledge is that the research focused on lobbying through public consultation rather than private meetings, it thus analyzes a different stream of influence that is not identical to previous research.

The operationalization of language as the variable to uncover lobbying processes, as well as determining success rates of interest groups, showed that in the specific case study there was a clear divide between economic and ethical interest. Furthermore computer assisted CA uncovered policy positions of interest groups based on language rather than opinion. This more accurate pinpointing of policy position could then be used to measure lobby success by comparing public consultation submission with the policy draft of the European Commission. The researcher utilized various techniques of computer assisted CA and tailored them to create findings of lobbying dynamics.

The actual findings confirmed some of the existing literature such as the power of civil society interest groups in issue areas that have an ethical dimension. Other research findings were disputed, for example the size of lobby coalition did not bring about greater lobby success. While the research was in agreeance with some and disagreeances with other existing literature, the crucial point is how data was gathered and operationalized. The results were not influenced by bias or misrepresentation of interest groups or decision-makers in question.

The researcher believes that future lobby research ought to rely more on the possibilities algorithms provide and that further transparency developments in the European Union will result in greater access to data regarding lobbying dynamics that ought to be utilized by researchers.

7 Bibliography

Baumgartner, Frank and Berry, Jeffrey and Hojnacki, Marie and Kimball, David and Leech, Beth (2009) Lobbying and Policy Change Who Wins, Who Loses, and Why, The University of Chicago Press, Chicago

Benoit, Kenneth (2020) "Text as Data: An Overview" in (eds.) Curini, Luigi and Franzese, Robert, The SAGA Handbook of Research Methods in Political Science and International Relations, SAGA Publications Ltd, Los Angeles

Benot, Kenneth and Watanabe, Kohei and Wang, Haiyan and Nulty, Paul and Obeng, Adam and Müller, Stefan and Matsuo, Akitaka and Lowe, Willian (2018) "Quanteda: An R package for the quantitative analysis of textual data" Journal of Open Source Software Vol. 3(30): 774

Bernhagen, Patrick and Dür, Andreas and Marshall, David (2014) "Measuring lobbying success spatially" Interest Groups & Advocacy Vol. 3(2): 202-218

Beyers, Jan (2004) "Voice and Access" European Union Politics Vol. 5(2): 211-240

Beyers, Jan (2014) "Let's Talk! On the practice and method of interviewing policy experts" Interest Groups & Advocacy Vol. 3(2): 174-187

Bouwen, Pieter (2002) "Corporate Lobbying in the European Union: the logic of access" Journal of European Public Policy Vol. 9(3): 365-390

Bouwen, Pieter (2009) The European Commission in Coen, David and Richardson, Jeremy (ed.) Lobbying the European union: Institutions, actors and issues (Oxford, Oxford university press), pp.19-38

Braun, Celesta (2015) "Lobbying as a Leveraged Act: On Resource Dependencies and Lobby Presence" in (eds.) Lowery, David and Halpin, Darren and Gray, Virginia, The Organization Ecology of Interest Communities, Palgrave Macmillan, Houndmille

Bitoni, Alberto (2017) "The Role of Lobbying in modern democracy: A theoretical framework" in (eds.) Bitoni, Alberto and Harris, Phil, Lobbying in Europe, Public Affairs and the Lobbying Industry in 28 EU Countries, Palgrave Macmillan: London

Coen, David and Katsaitis, Alexandra and Vennoni, Matia (2021) Business Lobbying in the European Union, Oxford Scholarship Online, Oxford

Dahl, Robert Alan (1957) "The Concept of Power." Behavioral Science, Vol. 2(3), pp.201–215

De Bruycker, Iskander and Beyers, Jan (2018) "Lobbying Strategies and Success: Inside and Outside Lobbying in European Union Legislative Politics" European Political Science Review Vol. 11(1): 57-74

DG Internal Policies of the Union (2007) Lobbying in the European Union

Dür, Andreas (2008) "Interest Groups in the European Union: How Powerful are They?" West European Politics Vol. 31(6): 1212-1230

Dür, Andreas (2008B) "Measuring Interest Group Influence in the EU – A Note on Methodology" European Union Politics Vol. 9(4): 559-576

Dür, Andreas and Bernhagen, Patrick and Marshall, David (2015) "Interest Group Success in the European Union: When (and Why) Does Business Lose?" Comparative Political Studies Vol. 48(8): 951-983

Dür, Andreas and Mateo, Gemma (2012) "Who lobbies the European Union? National interest groups in a multilevel polity" Journal of European Public Policy Vol. 19(7): 969-987

Eising, Rainer (2007) "Institutional Context, Organizational Resources and Strategic Choices: Explaining Interest Group Access in the European Union." European Union Politics 8(3): 329–362.

Eising, R. (2009) The Political Economy of State –Business Relations in Europe: Modes of Interest Intermediation, Varieties of Capitalism, and the Access to EU Policy-making, London: Routledge.

European Commission (2001) European Governance: A White Paper [online available at: https://ec.europa.eu/commission/presscorner/detail/en/DOC_01_10] last accessed 17. May 2022

European Commission (2020) Artificial intelligence – ethical and legal requirements [online available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements_en] last accessed 10. May 2022

Flöthe, Linda (2019) "The cost of interest representation - a resource perspective of informational lobbying" European Political Science Review Vol. 11: 161-178

Friedrich, Silke (2013) "Policy Persistence and rent extraction" Public Choice Vol. 157(1/2): 287-304

Grimmer, Justing and Stewart, Brandon (2013) "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts" Political Analysis Vol. 12: 267-297

Hall, Richard and Deardorff, Alan (2006) "Lobbying as Legislative Subsidy" The American Political Science Review Vol. 100(1): 69-84

Hillman, Amy and Withers, Michael and Collins, Brian (2009) "Resource Dependence Theory: A Review" Journal of Management Vol. 35(6): 1404-1427

Holman, Craig and Luneburg, William (2012) Lobbying and transparency: A comparative analysis of regulatory reform. Interest Groups & Advocacy, Vol. 1(1): 75-104

Jockers, Matthew (2020) "Extract Sentiment and Sentiment-Derived Plot Arcs from Text" [online available at: https://cran.r-project.org/web/packages/syuzhet/syuzhet.pdf] last accessed 10. May 2022

King, Cheryl Simrell and Feltey, Kathryn and O'Neill Susel, Bridget (1998) "The Question of Participation: Toward Authentic Public Participation in Public Administration" Public Administration Review Vol. 58(4): 317-326

Klüver, Heike. (2009) "Measuring Interest Group Influence Using Quantitative Text Analysis." European Union Politics Vol. 10(4): 535–549

Klüver, Heike (2013) Lobbying in the European Union Interest Groups, Lobbying Coalitions, and Policy Change, Oxford University Press, Oxford

Klüver, Heike (2013B) "Lobbying as a collective enterprise: winners and losers of policy formulation in the European Union" Journal of European Public Policy Vol. 20(1): 59-76

Loomis, Burdett (1983) "A New Era: Groups and the Grassroots", in Allan J. Cigler and Burdett A. Loomis (eds) Interest Group Politics, pp. 169–90. Washington, DC: CQ Press.

Lowery, David (2013) "Lobbying influence: Meaning, measurement and missing " Interest Groups and Advocacy Vol. 2(1): 1-26

McKay, Amy (2012) "Buying Policy? The Effects of Lobbyists' Resources in Their Policy Success" Political Research Quarterly Vol. 65(4): 908-923

Monroe, Burt and Schrodt, Philip (2008) "Introduction to the Special Issue: The Statistical Analysis of Political Texts" Political Analysis Vol. 16: 351-355

Mullen, Lincoln (2020) "Detect Text Reuse and Document Similarity" [online available at: https://cran.r-project.org/web/packages/textreuse/textreuse.pdf] last accessed 10. May 2022

Nagel, Jack H. (1975) The Descriptive Analysis of Power. New Haven, CT: Yale University Press.

Pfeffer, Jeffrey (1987) "A resource dependence perspective on interorganizational relations" in Mizruchi, Mark and Schwartz Michael (Eds.), Intercorporate relations: The structural analysis of business: 22-55. Cambridge, UK: Cambridge University Press

Slapin, Jonathan and Sven-Oliver Proksch. (2008) "A Scaling Model for Estimating Time-Series Party Positions from Texts." American Journal of Political Science 52(3): 705-772

Strickland, James (2020) “The Declining Value of Revolving-Door Lobbyists: Evidence from the United States” American Journal of Political Science Vol. 64(1): 67-81

Watanabe, Kohei (2021) “Latent Semantic Scaling: A Semisupervised Text Analysis Technique for New Domains and Languages” Communication Methods and Measures Vol. 15(2): 81-102

Woll, Cornelia (2007) “Leading the Dance? Power and Political Resources of Business Lobbyists” Journal of Public Policy Vol. 27(1): 57-78

8 Appendix

8.1 List of Interest Groups

Number	Interest Group	Type of Organization
1	*FERMA*	*Trade Union*
2	*ACCESSNOW*	*NGO*
3	*ACEA*	*Business Association*
4	*ADIGTAL*	*Business Association*
5	*AGORIA*	*Business Association*
6	*ALLAI*	*NGO*
7	*American Chamber of Commerce*	*Business Association*
8	*AMETIC*	*Business Association*
9	*AmsterdamAITechnologyForPeople*	*Research Institution*
10	*ANEC*	*Consumer Org*
11	*APPLIA*	*Business Association*
12	*AFME*	*Business Association*
13	*ATOS SE*	*Business Association*
14	*EDHEC Business School*	*Research Institution*
15	*BEUC*	*Consumer Org*
16	*BITKOM*	*Business Association*
17	*BSA*	*Business Association*
18	*BUNDESÄRZTEKAMMER*	*NGO*
19	*BDI*	*Business Association*
20	*BUSINESSEUROPE*	*Business Association*
21	*CEA*	*Research Institution*
22	*CECE*	*Business Association*
23	*CEMA*	*Trade Union*
24	*Center for Data Innovation*	*NGO*
25	*Center for Democracy & Technology*	*NGO*
26	*CLEPA*	*Business Association*
27	*CLIFFORD Chance LLP*	*Business Association*
28	*COCIR*	*Business Association*
29	*CCIA*	*Business Association*
30	*Confederation of Industries CR*	*Business Association*
31	*Consumer Technology Association*	*Business Association*
32	*DIKU*	*Research Institution*
33	*Developers Alliance*	*Business Association*
34	*DIGITALEUROPE*	*Business Association*
35	*EDIMA*	*Business Association*
36	*EGMF*	*Business Association*
37	*Electronic Private Information Center*	*NGO*
38	*ENBW*	*Business Association*

39	*ENEL SPA*	*Business Association*
40	*ETNO*	*Business Association*
41	*EUNITED*	*Business Association*
42	*EURALARM*	*Business Association*
43	*EUROCITIES*	*Public Authority*
44	*EUROCOMMERCE*	*Business Association*
45	*EACA*	*Research Institution*
46	*European Association of Urology*	*NGO*
47	*European Banking Federation*	*Business Association*
48	*European Tech Alliance*	*Business Association*
49	*European Trade Union Confederation*	*Trade Union*
50	*FAIR TRIALS*	*NGO*
51	*Federation Francais de l'Assurance*	*Business Association*
52	*FEDMA*	*Business Association*
53	*FORTUM OY*	*Business Association*
54	*F-SECURE Corporation*	*Business Association*
55	*FUJITSU*	*Business Association*
56	*GDV*	*Business Association*
57	*GOOGLE*	*Business Association*
58	*HUAWEI*	*Business Association*
59	*HUTCHISON Europe*	*Business Association*
60	*IBEC*	*Business Association*
61	*IBM*	*Business Association*
62	*IKEA*	*Business Association*
63	*ITI*	*Business Association*
64	*INSURANCEEUROPE*	*Business Association*
65	*Interactive Software Federation of Europe*	*Business Association*
66	*STM*	*Business Association*
67	*Japanese Business Council*	*Business Association*
68	*KEIDANREN*	*Business Association*
69	*LIDERLAB*	*Research Institution*
70	*MASTERCARD*	*Business Association*
71	*MEDTECH*	*Business Association*
72	*MICROSOFT*	*Business Association*
73	*NL AIC*	*Other*
74	*ORGALIM*	*Business Association*
75	*PGEU*	*NGO*
76	*PHILIPS*	*Business Association*
77	*RELX*	*Business Association*
78	*SCIENCEEUROPE*	*Research Institution*
79	*Software Alliance for IT*	*Business Association*
80	*SOFTWAREAG*	*Business Association*
81	*Tech in France*	*Business Association*
82	*TECHUK*	*Business Association*
83	*THEGOODLOBBY*	*NGO*
84	*TWILIO*	*Business Association*

85	*VDMA*	*Business Association*
86	*VZBV*	*Consumer Org*
87	*VISA*	*Business Association*
88	*VODAFONE*	*Business Association*
89	*WKO*	*Business Association*
90	*WORKDAY*	*Business Association*
91	*ZVEI*	*Business Association*
92	*ZPP*	*Business Association*

8.2 RStudio Code

```
# To prepare for the analysis, following packages need to be loaded into the
R workspace

install.packages("ggplot2")
require(ggplot2)
install.packages("lsa")
require(lsa)
install.packages("quanteda")
require(quanteda)
install.packages("quanteda.textmodels")
require(quanteda.textmodels)
install.packages("quanteda.textplots")
require(quanteda.textplots)
install.packages("LSX")
require(LSX)
install.packages("LSX")
require(LSX)
install.packages("quanteda.textstats")
require(quanteda.textstats)
install.packages("plotly")
library(plotly)
install.packages("wordcloud")
require(wordcloud)
install.packages("syuzhet")
require(syuzhet)
install.packages("topicmodels")
library(topicmodels)
install.packages("tidyverse")
require(tidyverse)
install.packages ("tm")
```

```
require(tm)
install.packages("readtext")
require(readtext)
install.packages("stringr")
require(stringr)
install.packages("writexl")
library("writexl")
install.packages("plotly")
require(plotly)
install.packages("topicmodels")
library(topicmodels)
install.packages("readtext")
require(readtext)

# 1. All Public Consultation Documents were downloaded and pre-processed in Microsoft word
# as described in the Methodology section. They were then transposed into an Excel spreadsheet
# that can now be loaded into RStudio with following function:

dat_PC <- read.table(file = "clipboard",
                     sep = "\t", header=TRUE)
print(dat_PC)

# 2. In the next step the Data loaded into R needs to be transformed into a corpus for further processing:

corp_PC <- corpus(dat_PC, text_field = "Text")

summary(corp_PC)

# 3. Now further processing of the Public Consultation documents needs to happen, the documents are
# tokenized and unnecessary information such as stop words, enumerations etc. are removed.
# Furthermore the words are stemmed and transformed into lower case

toks_TEXT <- tokens(corp_PC, remove_punct = TRUE)
toks_TEXT <- tokens(toks_TEXT, remove_punct = TRUE, remove_numbers = TRUE)
toks_TEXT <- tokens_select(toks_TEXT, stopwords('english'),selection='remove')
toks_TEXT <- tokens_wordstem(toks_TEXT)
```

```
toks_TEXT <- tokens_tolower(toks_TEXT)

# 4. For future analysis a Document-Feature-Matrix (DFM) is created

dfm(toks_TEXT)
print(dfm)
dfm_PC <- dfm(toks_TEXT)

# 5. The DFM can now be exported as an excel document


df <- data.frame(dfm_PC)

write_xlsx(df,"C:\\Users\\loehr_gz9ch6v\\THESIS/dataframe.xlsx")


# 6. In this step the research uses self-created dictionary to filter the words used in the
# submissions to the public consultation process according to the policy dimension they discuss.
# The self-created dictionary puts emphasis on ethical and business issues

install.packages("LSX")
require(LSX)

topfeatures(dfm_PC, 100)
seed <- as.seedwords(data_dictionary_sentiment)
print(seed)
seed <- as.seedwords(data_dictionary_ideology)

tmod_lss <- textmodel_lss(dfm_PC, seeds = seed,
                          k = 92, cache = TRUE)
head(coef(tmod_lss), 40)

tail(coef(tmod_lss), 40)

textplot_terms(tmod_lss, dict)

dict <- dictionary(list(Neutral = c("ai", "use", "potenti", "effect"), economy = c("advantage", "rich", "vital", "progress", "grow", "player", "moder", "pace", "immens"),
                     HR = c("right", "poor", "suffer", "error", "disadvantag", "relianc", "infring", "threaten")))
```

```
topfeatures(dfm_PC, 4292)
seed <- as.seedwords(data_dictionary_sentiment)
print(seed)
seed <- as.seedwords(dict)

tmod_lss <- textmodel_lss(dfm_PC, seeds = seed,
                    k = 92, cache = TRUE)
head(coef(tmod_lss), 40)
tail(coef(tmod_lss), 40)

textplot_terms(tmod_lss, dict)


kw_TEXT <- kwic(toks_TEXT, pattern =  'error*') # need to define keywords
that i want to use in my analysis
head(kw_TEXT, 300)


# In the next step the research will run a similar dictionary analysis based
on pre-existing dictionaries.
# The pre-existing dictionaries connect a sentiment to each word which
allows a sentiment analysis of
# each submission in the public consultation

get_sentiment_dictionary(dictionary = "syuzhet", language = "english")

syuzhet <- get_sentiment(corp_PC, method="syuzhet")
bing <- get_sentiment(corp_PC, method="bing")
afinn <- get_sentiment(corp_PC, method="afinn")
nrc <- get_sentiment(corp_PC, method="nrc")

recent_corpus <- corpus_subset(corp_PC)
ndoc(recent_corpus)
text_corpus <- as.character(recent_corpus)[1:ndoc(recent_corpus)]
str(text_corpus)
syuzhet_vector <- get_sentiment(text_corpus, method="syuzhet")
syuzhet_vector #scores text by sentiment

sentiments <- data.frame(syuzhet, bing, afinn, nrc)
print(sentiments)

plot(
```

```
  syuzhet_vector,
  type="h",
  main= "Sentiment",
  xlab = "Interest Group",
  ylab= "Sentiment Score"
)

# In this step the computer program defines themes and topics within the different submissions of the
# public consultation process. The themes are created by contects the words are used in

texts = corpus_reshape(corp_PC, to = "paragraphs")
PC_dtm <- dfm(texts, stem = TRUE, remove_punct = TRUE, remove = stopwords("english"))

PC_dtm <- dfm_trim(PC_dtm, min_count = 10)
PC_dtm <- convert(PC_dtm, to = "topicmodels")
set.seed(1)
lda_model <- topicmodels::LDA(PC_dtm, method = "Gibbs", k = 20)
terms(lda_model, 20)

# In the following the unsupervised machine learning program "Wordfish" is applied. It scales documents
# according to their policy dimension on a left-right scale


tmod_wf <- textmodel_wordfish(dfm_PC)
summary(tmod_wf)
tmod_wf <- textmodel_wordfish(dfm_PC, dir = c(60, 10))
summary(tmod_wf)

textplot_scale1d(tmod_wf)

textplot_scale1d(tmod_wf, margin = "features", highlighted = c("ai", "suffer", "poor", "relianc", "error", "error", "infring", "rich", "vital", "progress", "immens", "moder", "grow", "pace", "player"))

textplot_wordcloud(dfm_PC)
topfeatures(dfm_PC)

features_dfm <- textstat_frequency(dfm_PC, n = 100)
summary(features_dfm)
```

```
results$documents<-textmodel_wordfish(dfm_PC,dir=c(2,5))
summary(results$documents)

textmodel_wordfish(
  dfm_PC,
  dir = c(2, 20),
  priors = c(Inf, Inf, 3, 1),
  tol = c(1e-06, 1e-08),
  dispersion = c("poisson", "quasipoisson"),
  dispersion_level = c("feature", "overall"),
  dispersion_floor = 0,
  sparse = FALSE,
  abs_err = FALSE,
  svd_sparse = TRUE,
  residual_floor = 0.5
)

tmod1 <- textmodel_wordfish(dfm_PC, dir = c(2,20))
summary(tmod1, n = 10)
coef(tmod1)
predict(tmod1)
predict(tmod1, se.fit = TRUE)
predict(tmod1, interval = "confidence")

tmod2 <- textmodel_wordfish(dfm_PC, dir = c(2,20))

tmod3 <- textmodel_wordfish(dfm_PC, dir = c(2,20),
                            dispersion = "quasipoisson", dispersion_floor = 0)

tmod4 <- textmodel_wordfish(dfm_PC, dir = c(2,20),
                            dispersion = "quasipoisson", dispersion_floor = .5)

plot(tmod3$phi, tmod4$phi, xlab = "Min underdispersion = 0", ylab = "Min
underdispersion = .5",
     xlim = c(0, 1.0), ylim = c(0, 1.0))

plot(tmod3$phi, tmod4$phi, xlab = "Min underdispersion = 0", ylab = "Min
underdispersion = .5",
     xlim = c(0, 1.0), ylim = c(0, 1.0), type = "n")


# Textstat similarity
```

```
dfmat <- dfm(corpus_subset(corp_PC),
           remove_punct = TRUE, remove = stopwords("english"))

dfmat <- dfm_PC



(tstat1 <- textstat_simil(dfmat, method = "cosine", margin = "features"))



# In this part of the research similarity between texts is analysed. The Code
uses
# previously created tokens of 2 by the researcher chosen documents and
compares the
# similarity of such. The texts are analysed on a scale from 0 to 1, 0 meaning
that there is
# no similarity between texts and 1 meaning that the texts are identical.

install.packages("textreuse")
require(textreuse)

toks_TEXT <- tokens(corp_PC, remove_punct = TRUE)
toks_TEXT <- tokens(toks_TEXT, remove_punct = TRUE,
remove_numbers = TRUE)
toks_TEXT <- tokens_select(toks_TEXT,
stopwords('english'),selection='remove')
toks_TEXT <- tokens_wordstem(toks_TEXT)
toks_TEXT <- tokens_tolower(toks_TEXT)



a <- toks_TEXT[["text8"]]
b <- toks_TEXT[["text75"]]



jaccard_similarity(a, b)
ratio_of_matches(a, b)



path_data <- system.file("C:/Users/loehr_gz9ch6v/THESIS/Thesis
Reading/AIA short.docx", package = "readtext")



dat_AIA <- read.table(file = "clipboard",
```

```
                  sep = "\t", header=TRUE)
print(dat_AIA)

corp_AIA <- corpus(dat_AIA, text_field = "Text")

summary(corp_AIA)
toks_AIA <- tokens(corp_AIA)



toks_AIA <- tokens(corp_AIA, remove_punct = TRUE)
toks_AIA <- tokens(toks_AIA,  remove_punct = TRUE, remove_numbers =
TRUE)
toks_AIA <- tokens_select(toks_AIA,
stopwords('english'),selection='remove')
toks_AIA <- tokens_wordstem(toks_AIA)
toks_AIA <- tokens_tolower(toks_AIA)

# combine all parts of the AIA corpora

allAIA <- c(toks_AIA[["text1"]], toks_AIA[["text2"]], toks_AIA[["text3"]])

print(allAIA)

c <- allAIA



# combine all different types of organisation that participated in the public
consultation to see
# if any form of org lobbied more or less successfull

# NGO

allNGO <- c(toks_TEXT[["text2"]], toks_TEXT[["text6"]],
toks_TEXT[["text18"]], toks_TEXT[["text24"]], toks_TEXT[["text25"]],
toks_TEXT[["text37"]]
         , toks_TEXT[["text46"]], toks_TEXT[["text50"]],
toks_TEXT[["text75"]], toks_TEXT[["text83"]])

print(allNGO)

jaccard_similarity(allAIA, allNGO)

# RESULT: 0.3438045
```

```
# TRADE UNION

allTU <- c(toks_TEXT[["text1"]], toks_TEXT[["text23"]],
toks_TEXT[["text49"]])
print(allTU)

jaccard_similarity(allAIA, allTU)

# RESULT: 0.3583123


# RESEARCH OR ACADEMIC INSTITUTION

allRCI <- c(toks_TEXT[["text9"]], toks_TEXT[["text14"]],
toks_TEXT[["text21"]], toks_TEXT[["text32"]], toks_TEXT[["text45"]],
toks_TEXT[["text69"]]
        , toks_TEXT[["text78"]])

print(allRCI)

jaccard_similarity(allRCI, allAIA)

# RESULT: 0.3422274

# Public Authority

allPA <- c(toks_TEXT[["text43"]])
print(allPA)

jaccard_similarity(allPA, allAIA)


# Result: 0.252541

# Other
allOTHER <- c(toks_TEXT[["text73"]])
jaccard_similarity(allOTHER, allAIA)

# RESULT: 0.1400844
```

```
# Consumer ORG

allCO <- c(toks_TEXT[["text10"]], toks_TEXT[["text15"]],
toks_TEXT[["text86"]])

jaccard_similarity(allCO, allAIA)
#RESULT: 0.3543506

# BUSINESS

allBUSINESS <- c(toks_TEXT[["text3"]], toks_TEXT[["text4"]],
toks_TEXT[["text5"]], toks_TEXT[["text7"]],
        toks_TEXT[["text8"]], toks_TEXT[["text11"]], toks_TEXT[["text12"]],
toks_TEXT[["text13"]],
        toks_TEXT[["text16"]], toks_TEXT[["text17"]], toks_TEXT[["text19"]],
toks_TEXT[["text20"]],
        toks_TEXT[["text22"]], toks_TEXT[["text26"]],
        toks_TEXT[["text27"]], toks_TEXT[["text28"]], toks_TEXT[["text29"]],
toks_TEXT[["text30"]],
        toks_TEXT[["text31"]], toks_TEXT[["text33"]], toks_TEXT[["text34"]],
toks_TEXT[["text35"]],
        toks_TEXT[["text36"]], toks_TEXT[["text38"]],
        toks_TEXT[["text39"]], toks_TEXT[["text40"]], toks_TEXT[["text41"]],
toks_TEXT[["text42"]],
        toks_TEXT[["text44"]], toks_TEXT[["text47"]], toks_TEXT[["text48"]],
toks_TEXT[["text51"]],
        toks_TEXT[["text52"]], toks_TEXT[["text53"]],
        toks_TEXT[["text54"]], toks_TEXT[["text55"]], toks_TEXT[["text56"]],
toks_TEXT[["text57"]],
        toks_TEXT[["text58"]], toks_TEXT[["text59"]], toks_TEXT[["text60"]],
toks_TEXT[["text61"]],
        toks_TEXT[["text62"]], toks_TEXT[["text63"]],
        toks_TEXT[["text64"]], toks_TEXT[["text65"]], toks_TEXT[["text66"]],
toks_TEXT[["text67"]],
        toks_TEXT[["text68"]], toks_TEXT[["text70"]], toks_TEXT[["text71"]],
toks_TEXT[["text72"]],
        toks_TEXT[["text74"]], toks_TEXT[["text76"]],
        toks_TEXT[["text77"]], toks_TEXT[["text79"]], toks_TEXT[["text80"]],
toks_TEXT[["text81"]],
        toks_TEXT[["text82"]], toks_TEXT[["text84"]], toks_TEXT[["text85"]],
toks_TEXT[["text87"]],
        toks_TEXT[["text88"]], toks_TEXT[["text89"]], toks_TEXT[["text90"]],
toks_TEXT[["text91"]],
```

```
          toks_TEXT[["text92"]])
print(allBUSINESS)

jaccard_similarity(allBUSINESS, allAIA)

# RESULT: 0.2621985
```