Abstracts - Talk of Europe Creative Camp #2

Meertens Institute Amsterdam – 23-27 March 2015

Economic Crisis in Europe and EP Political Groups

Anastasia Deligiaouri - Technological Educational Institute of Western Macedonia, Greece

The European Parliament constitutes the only institution in Europe in which citizens have a voice, basically by electing their representatives. Therefore the EP, especially after the Lisbon Treaty, has upgraded duties and a crucial role in the formation of European politics. Analysis of EP proceedings during economic crisis is an important task as this can reveal crucial aspects of how political groups in EP have addressed significant social problems and which policies they propose.
For the purpose of analyzing the aforementioned issues we propose a political analysis of EP proceedings during the period of 2008-2012. Initially, we will search for the most frequently discussed topics in the plenaries and determine which topics are relevant to economic crisis. Then we will look – in these plenaries – for the keywords in the speeches of the MEP of big political groups in order to analyze how major issues of economic crisis are confronted. At the second stage a discourse analysis may be applied as well. The successful study of EP proceedings will provide us with useful insights regarding the ideological profile of each political group, and consequently, the grounds for the policies decided.

Higher Education in the EP: an Inventory and Social Network Analysis of Discussions

Julie M. Birkholz - Centre for Higher Education Governance Ghent, University Ghent, Belgium

Higher education is increasingly promoted as a policy solution to various other policy problems. The development of EU higher education policies, and linkages between higher education and other sectors in EU legislative processes are not necessarily apparent due to

complex governance arrangements between the different EU institutions and the lack of data that would allow a systematic analysis. The Talk of Europe dataset offers an avenue to systematically explore how higher education is proposed in non-higher education agenda topics considering the role of a number of factors: parliamentary committees, individual members of the EP, different political parties, coalitions, and nation-states in explaining the prominence of higher education. In order to use the dataset as a valid source for investigating these policy discussions and changes, it is essential to complement this data with two additional public data stores that document the activities within the EP – the documentation of the EP committees and the final budget decision documents of the EP. Through a mixed-methods analysis we seek to identify patterns in the relations of higher education and non-higher education agenda topics. This research is relevant not only to the field of higher education but also in understanding the role of EP in policymaking.

Mining the Relationship between UK Parliament and European Parliament Debate Data
Chaohai Ding - Web and Internet Science Group, University of Southampton, United Kingdom

The publicity surrounding the UK Parliamentary debates, such as the television channel BBC Parliament, has exerted tremendous influence on the transparency of politics in the UK. Political figures need to be responsible for what they have said in the debates as they are monitored by the media and public, as well as by the European Parliament. In order to investigate and mine the relationship between the UK Parliament and EU Parliament debates, we first want to use named entity recognition algorithms to automatically match the same MP in the Talk of Europe dataset, WhatTheySaid dataset, and DBpedia dataset, in order to interlink these datasets. After interlinking the same entities, we will use data mining algorithm to cluster the MPs based on the topics of debate, sentiments and their background information (such as education, university attended, major subjects). We also want to look into the relationship of the UK and EU Parliament debates after we interlink the two datasets together. We will mainly use topic extraction, summarization algorithm and sentimental analysis to find the similarities between the two debate datasets.

Kinds of Topics Discussed by Politicians in the European Parliament
Danguolė Kalinauskaitė - Vytautas Magnus University, Lithuania

This research project aims to enrich the proceedings of the European Parliament with topic information. It is intended to address the following research question: what kinds of topics are discussed by politicians in the EP? In order to classify the EP proceedings according to their topics, theme keywords seem to be valuable. The analysis of the debates will be carried out by fully automatically employing the corpus management and analysis system Sketch Engine (using one of its features – keyword extraction). This system enables gathering and studying of large text collections according to specific linguistically motivated queries. Moreover, it provides many open corpora that can be used for different kinds of text analysis. One of them – European Parliament Proceedings Parallel Corpus prepared by Philipp Koehn – will be used as a reference corpus in order to extract keywords from the EP proceedings in English. Corpus-derived keywords will provide some information concerning the contents of political debates in the EP. The intended outcome of this research is topic annotation, as this functionality of the dataset will help identify topics discussed by politicians in the EP and relate one debate to another on the basis of thematic criterion more accurately than it can be done now (only according to the titles of the debates).

DataBate: Debates Database

Alexander Asatiani - School of Journalism at Georgian Institute of Public Affairs, Georgia
Tamar Kalkhitashvili - The Center of Linguistic Research, Ilia State University, Georgia

Thousands of students doing their masters in journalism, political science and public relations at Georgian universities are not really aware of the issues related to Georgia that are being discussed in the EP. In order to discover the documents related to Georgia, we will analyze the dataset of all plenary debates held in the EP between July 1999 and January 2014. A database of Georgia-related EP documents – in which data will be structured according to the various categories – can serve as helpful teaching material for MA students of the abovementioned faculties: students can get familiar with this kind of environment and will be able to access needed information.
At the Journalism and Media Management School of GIPA (Georgian Institute of Public Affairs) we have an English department, where students from Azerbaijan and Armenia are learning together with Georgians. After the database is demonstrated to these students, we could collaborate with them in order to create a Transcaucasian database where EP documents related to the region will be gathered and categorized. The intended outcome of this project will be a web resource with a convenient search system, which can be used by any interested user, e.g. Georgian journalists who write about Georgia-EU relationships. The web page will contain sections with statistical data, infographics and data visualizations.

Linking the European and Hellenic Parliament through Debates

Iason Kostopoulos, Ioannis Moschonas and Georgios Smyrnaios - Computer Science Department, Institute of Computer Science, FORTH, Greece

Nowadays the question about the future of Europe as a monetary and political union has increasingly become an issue, especially with the rise of different opinions on several topics. Understanding how national parliamentary debates are related to the ones that take place in the EP will give us insight into how people view or respond to the issues raised at European level. The objective of our project is to use state of the art methods from the field of Information Retrieval in order to compare the relevance of debates in the Hellenic Parliament with those in the European Parliament with respect to the topics discussed. To achieve this purpose we will use tools for document preprocessing (e.g. stemmer, lemmatizer) and an indexing program to measure the relevance of debates. Our approach is to analyze the association between the main topics that are discussed in both parliaments and to determine whether the speeches of each MP are on topic. Our results can be used by many scientific agents in order to understand which topics the national parliaments are mostly concerned with. Furthermore, our research could be of use to any European citizen who wants to gain insight into the ways his national MPs are positioned in the European Parliament – as well as in relation to the position of the MPs party at a national level.

EU Country and/or MP Profiles

Alexander Tkachenko, Konstantin Tretyakov and Ilya Kuzovkin - University of Tartu, Estonia

We believe that one of the main kinds of information hidden within the otherwise unstructured pile of text of the EP proceedings is the implicit ‘profile’ projected by each country. This single statement immediately raises a number of interesting research

hypotheses along with constructive ways of answering them. In particular, we are interested in the following questions:
1). Is the initial statement true at all? The hypothesis could be formalized as a standard statistical question as follows: are there any features of the proceeding texts that are significantly discriminative among the representatives of various countries, apart from their source language?
2). If there is, indeed, a set of discriminative attributes, could those be conveniently interpreted and visualized? Depending on what features turn out to be relevant, possible visualizations may range from word clouds to scatter plots.
3). Are there any clear trends within the 'discriminative' feature sets over time? What parts of the 'proceedings profile' of a country are generally stable (and even independent of the MP) and which seem to change? Can we relate those changes to actual events (e.g. records in Wikipedia, stock indicators or currency rates)?

Hacking the Style: Who Talks the Same in EP?
Justina Mandravickaite - Baltic Institute of Advanced Technology, Lithuania
Žygimantas Medelis - TokenMill, UAB, Lithuania
Tomas Krilavičius - Vytautas Magnus University and Baltic Institute of Advanced technology, Lithuania

This project addresses the following topic and research questions:
1.) How is rhetoric of members of the EP similar or different to rhetoric of their factions (party groups) and to other parliamentary factions (party groups)?
2.) To whom are they similar in terms of rhetorical style (at individual as well as faction (party group) level?
To address these questions, stylometric analysis of speeches of EP members will be applied. Stylometry or computational stylistics is used for analyzing texts (it can also be applied to images, music, gene sequences, etc.) while focusing on their style and structure. Mostly shallow features (e.g., character count, average characters per word, letters, function words, punctuation, etc.) are used. They or their vectors are classified using supervised or unsupervised methods. The intended outcome is to have speeches of EP members visualized as a network according to their rhetorical similarity/dissimilarity to their parliamentary faction (party group) as well as other parliamentary factions (party groups). This will allow us to identify the EP members who are most similar to their own faction (in rhetorical sense), reveal rhetorical similarities/dissimilarities in EP at the level of factions (party groups) and to find 'outliers' – Parliament members who are similar to other factions (in rhetorical sense).

Agendas of EP Debates
Vytautas Mickevičius and Tomas Krilavičius - Vytautas Magnus University and Baltic Institute of Advanced Technology, Lithuania
Vaidas Morkevičius - Kaunas University of Technology, Lithuania

This project aims to develop a tool for automatic classification of EP debates into an existing thematic taxonomy of Comparative Agendas Project (www.comparativeagendas.info). We propose to enhance the existing Talk of Europe Database with thematic classification of debates into a taxonomy already extensively used for the classification of various political texts. Several different text processing (feature extraction) and classification techniques will be deployed. The main idea is to apply bootstrapping algorithm for the thematic classification, i.e. to take EP debates titles in Lithuanian, and apply classifier(s) trained on the

titles of debates of the Lithuanian parliament, then check the accuracy of the thematic classification for a random sample of classified EP debates titles (and correct it if needed), add them to the annotated corpora, retrain classifier(s), and repeat the procedure until the accuracy of classification is satisfactory. We are planning to experiment with different classifiers (SVM, Naive Bayes etc.). The intended outcome of our participation – a thematically annotated database of EP debates ('Talk of Europe') – would not only allow to systematically monitor topics debated in the EP, but could easily be linked with topics of debates in national parliaments as well. The resulting annotated corpora primarily based on analysis of Lithuanian texts will also be suitable for classification experiments in other languages.

Finding Crosslingual Similarity Measures with the Use of Distributed Semantics
Bartholomaeus Wloka - Faculty of Computer Science, University of Vienna, Austria
Vesna Lušicky - Centre for Translation Studies, University of Vienna, Austria

In this project, we will explore the support for multilingual and transcultural communication (translation, interpreting, multilingual terminologies) by applying the distributed semantics approach (Gyllensten 2015) to the corpora in four languages (English, German, Slovenian, Polish). This includes the collection and comparison of co-occurance statistics for terms, which will then be represented as vectors representing their distributional properties. The novel approach of Neighborhood Graphs will be utilized to find more similarities between terms. In this approach a topological model is used which takes into account the local structure of neighborhoods in semantic space. The result will be clusters of associated words and multi-word expressions. We will examine the topological structures of closely related expressions and the possibility of cross language connection between those clusters, and the possibility of creating concept frameworks, without previous knowledge and annotation. The comparison of clusters between different languages will result in crosslingual association clouds, which can be used as translation support in the future. The output will include a dataset describing the associations of selected within one language and between languages, and a toolchain to reproduce the results with other corpora in various languages.

Exploring Parliament Datasets on Cultural Heritage
Beatriz Barros - University of Malaga, Spain
Antonio Gallardo and Pedro Luengo - University of Sevilla, Spain

We propose to use the methodology of the SiSOB project (http://sisob.lcc.uma.es/) to extract data from parliament interventions using information retrieval and natural language technics. In some proofs, we have seen that the parliament dataset has been processed and stored in RDF. With SPARQL we have analyzed this dataset, specifically the speech part (e.g. ID, date, country of representation and text of the debate (in original language, English and the other languages)). With this dataset, the taxonomy of concepts in English, and the SiSOB extractor we are planning to annotate the EP debates. As methods, we propose to use SiSOB tools with a rich set of contextual information, such as a taxonomy of concepts and the set of dictionaries needed in order to work with the tool. After the extraction process, we will prepare an intermediate dataset to give answer to topics related to Cultural Policy. We will then create a webpage with a visual tool to have access to those debates using an analytics interface. The aim is to present the results visually with a link that gives access to the original debate (on a webpage on our server, available for running a demo during the campus).

Them and Us: Linking UK and EU Political Debate

Wim Peters and Adam Funk - Natural Language Processing Group, Department of Computer Science, University of Sheffield, United Kingdom

The investigation of the reflection of EU parliamentary topics in UK political debate will give interesting insights into the ways in which EU issues are received and understood in a national UK political context. In order to be able to investigate this, we will provide a humanities researcher with extendable automated assistance in domain knowledge acquisition and data search, resulting in the enrichment of the Talk of Europe data set with the following:

1). Domain modelling in the form of named entity recognition, term extraction and conceptual organization by means of freely available tools available from the General Architecture for Text Engineering (GATE; www.gate.ac.uk).

2). Links between terms derived from the EU proceedings to terms from UK parliamentary debates, written questions and answers, and written statements. These texts are available on the UK Parliament website (http://www.parliament.uk).

3). Linked open data resolution of named entities (persons, locations, organizations). The output will enable humanities researchers to focus on relevant material in both data sets, bootstrap her/his understanding of the domain, and perform further close reading of the selected texts in order to e.g. compare immigration issues and attitudes in politics on both sides of the channel.

Mapping Open Data as an Issue in European Parliament

Jonathan Gray - Department of Politics and International Relations at Royal Holloway, University of London, United Kingdom
Liliana Bounegru - Digital Methods Initiative, University of Amsterdam, the Netherlands

The concept of 'open data' has vaulted from being the rather rarefied preserve of a handful of information activists and technicians to possessing significant currency on the global political stage, featuring prominently in the speeches of Presidents, Prime Ministers, Mayors And Commissioners, as well as on the agendas of major international groups and organisations. This project seeks to address questions such as: who is talking about open data, what are they saying, and what are the different political visions and values underlying debates about open data? The project will use the Talk of Europe dataset to explore open data as a contested political concept that is continually reconfigured in response to shifting ideals, conceptions and practices of governance and democracy in different contexts. Building on previous research on this topic, it aspires to use metadata about speakers' countries and affiliations to develop a richer empirical picture of who is talking about open data and the visions, values and arguments articulated around it. The project uses a mixture of digital tools and methods developed at the University of Amsterdam's Digital Methods Initiative, tools developed at Science Po's médialab, as well as best of breed text mining, lexical analysis and network analysis tools. Results would feed into an interdisciplinary research project on open data and privacy, conducted as a collaboration between the Digital Methods Initiative and the Institute for Information Law at the University of Amsterdam, UC Berkeley and Open Knowledge, as well as other ongoing research projects on the topic of open data.