1) Recreating the Political Compass

The Political Compass is a popular website that aims to overcome the traditional unidimensional division between left and right, capturing political orientation on a two-dimensional chart, with a horizontal axis relative to the economy and a vertical axis corresponding to society. Using Quotebank quotation-specific dataset, we could try to recreate a similar model, aggregating all the quotations for each politician (or non-politician with enough “political” quotes) and representing their alignment with a single point. This kind of research has an interest because it allows citizens to compare their positioning with those of the candidates, and vote in a more informed way. Furthermore, the alignments can be aggregated partisan-wise to study not only the position of a party but also its diversity and plurality, measured by the scatter of corresponding points, and its evolution with time, emphasizing polarization trends.

A possible approach would be to identify a set of topics, inspired by the 62 original questions from politicalcompass.org, and extract only the quotes that address them, assigning then to each of these a score based on entity sentiment analysis and averaging on topics for each person. After this process, we could use PCA to reduce the high-dimensional vectors of features to just 2 dimensions, allowing simple representation and comparison with the traditional Compass.

On the same line, a more modest idea could be the prediction of US politicians’ party affiliation, based on a classification model fed with all the quotes from party leaders, with the same research purposes.

2) Investigating biases in news media

News media often propose themselves as impartial narrators of events, but it is a well-known fact that different sources of information tend to be more or less subtly aligned with different political positions. It is thus crucial to identify systematic biases to reduce manipulation and push towards neutral and fair information. Using Quotebank article-centric version, we could analyze for each newspaper or news website the space dedicated to different political parties and selected issues, alongside the corresponding opinion that emerges from the quotes, using topic extraction and entity sentiment analysis techniques. Normalizing the results on all the considered quotes, one would expect unbiased sources to lie around the mean, and biased ones to exhibit a significant shift.

3) Understanding political communication

Politicians often meticulously shape their language to convey a specific message about themselves, in an effort to persuade a certain audience. Authoritarian leaders make extensive use of words relative to the concepts of “strength” and “order”; those presenting themselves as innovators insist more on “evolution” and “novelty”; those aiming to project a human and personal dimension exploit the lexis of “family” and “emotions”. Using Quotebank quotation-specific dataset, we could study the paradigms of communication by investigating for each individual the type of vocabulary used and the feelings it arouses, the readability score (syntactic complexity) and other linguistical metrics. In addition, of particular interest would be the analysis of differences in personal branding based on gender, ethnicity or political affiliation.