italian politicalscience

From Words to Facts: Wordfish, a Modern Technique to Estimate Policy Positions of Political Actors

by Andrea Ceron | Published in issue6 / Research

Estimating Policy Positions of Actors: The Bounds of Surveys and Roll-Call Analysis

Expert surveys and roll-call analysis are two of the most common approaches to measure party ideal points and party cohesion (Benoit and Laver 2006). In expert surveys some specialists are asked to identify parties positions on several policy dimensions defined "ex-ante" by the researcher. For the purpose of most research projects, this technique is much too costly and time consuming. Moreover, it has been shown that expert surveys are not very reliable in the following situations: a) when parties are close to each other on a particular issue; b) when this issue is less relevant for the party; c) when the party is divided on the issue at hand (Mair 2001). Finally, expert's judgment could be biased according to their preferences (Curini 2010).
Roll-Call Analysis consists of inferring the position of actors by observing their behaviour under the assumption that different preferences lead to divergent behaviour. This is not necessarily true: sometimes voting behaviour in Parliament could be strategic or could be affected by party discipline (Laver 1999). In these cases, the actual preferences of actors are hidden and the output of estimation could be biased. Moreover, not all votes are roll-call votes, thus they do not provide a complete picture of voting behaviour. The decision to call a roll-call vote may also be strategic and biased according to the issues or the group that call it (Carrubba et al. 2006). Finally, this technique ends up being unhelpful in estimating positions of actors not involved in parliamentary voting that nonetheless exert effect on policy agenda and policy-making (i.e. head of states, political movements, interests groups, trade unions, churches, mass-media, judges, bureaucracies).
Even in those cases where roll-call votes can be properly employed to produce a valid estimate of policy position there are reasons to believe that text analysis is able to add something more, providing more information about the actual position of actors. In fact roll-call analysis is strictly related to voting behaviour that provides actors with a limited set of options: "yes", "no" or "abstention". However there could be a variety of preferences behind a similar voting behaviour : actors may say "yes" or "no" for very different (and sometimes opposite) reasons. Text analysis helps scholars discover and highlight these true preferences.

Wordfish: Discovering the Advantages of Text Analysis

The analysis of policy documents to determine actors' positions has long been used. The first and most well-known attempt goes back to the hand-coding of party platforms made by the manifesto research group (CMP). Recently semi-automated and automated techniques of text analysis have been developed, providing

quicker (but still reliable) means to collect data on actors' ideal points (e.g. Wordscores).
One of the most modern and automated techniques is Wordfish (Slapin and Proksch 2008; Proksch e Slapin 2009a).
Wordfish is a package developed to run with the freely available (but very powerful) statistical software R. It allows one to compare documents by scaling them along a common dimension that catches the political meaning of these texts. The underlying idea is that the relative frequency with which each word appears in a document provides information about the policy position of that text. Wordfish considers each document as a vector of randomly distributed words, meaning that the probability of each word to occur in a document is assumed to be independent of the position of other words.
In particular it assumes that words in a document follow a Poisson distribution. Then it employs an expectation maximization algorithm to compare expected and actual values of words and actors' positions, estimating through Maximum Likelihood (ML) two parameters of interests: ω and β.
The first parameter, ω, represents the policy positions of actors (with mean 0 and standard deviation equal to 1). Actors' ideal points are aligned along a single latent dimension that must be interpreted according to the political meaning of the documents analyzed. When documents are related only to a single issue (e.g. environmental policy) this topic will define the policy continuum. To the contrary, when they encompass the entire spectrum of policy issues (like party manifestos do) the latent dimension extracted should be interpreted as a more general left-right scale. Note that the software also allows one to build multidimensional spaces by getting actor's placement along different issues. This can be done dividing each document in different sub-sections (e.g. economic, social, foreign policy, and so on) and running distinct analyses.
The second parameter estimated, β, is the word discrimination parameter. It corresponds to the words placement along the policy scale, telling us whether each word stands on the left or on the right of the scale. In addition, it expresses the discriminating power of each word. Words that appear only in a few documents will be located on the extremes of the scale; they retain a higher absolute value of β and are more helpful in discriminating among documents. To the contrary, words that appear with an higher frequency in all documents have a discriminating power close to zero.
β values are important because they allow diagnostics of the analysis to be carried. Scholars should compare the estimated policy position of words with their actual usage in political language to make sure that the software succeeded in understanding the substantial political meaning of these words.
Wordfish has some advantages with respect to other techniques of content analysis. The main one is the ability to produce time series estimates. The algorithm assumes words usage to be constant over time and considers documents related to the same actor at different points in time as independent with respect to each other. Hence the position of one actor at time t does not affect its position at time t+1. For this reason changes in actor's positions across time are not an artifact of the model but are due to a real modification in the content of texts analyzed. On the contrary, if an actor's position remains stable in time, this means that the frequencies of words used by him did not change too much.
Another prerogative of this software is the capacity to analyze all the words contained in each document assigning a substantial political meaning to them. For this reason Wordfish is suitable even when analyzing a position at one point in time as we will see below. Finally, Wordfish provides confidence intervals of the estimates through a parametric bootstrap.
Summing up, Wordfish is a fast way to estimate positions of actors. Furthermore these estimates are reliable when compared with other techniques commonly employed to estimate policy position. It has been proved that Wordfish outputs are highly correlated with expert surveys, hand coding (CMP) or other software of content analysis.
In the next section we will illustrate two practical examples.

Estimating Times Series or Single Point Estimates: Two Practical Examples

The first example concerns a time series estimation and it is drawn using data from Ceron (forthcoming). This analysis has been done by collecting data on Italian party factions. The author gathered approximately

140 motions presented by factions during Italian parties' congresses from 1946 to 2010. These documents have been analyzed to determine the ideal point of each faction along the left-right scale. Here we will focus on an illustrative example on the temporal evolution of the ideological position of the median faction within the Italian Socialist Party (PSI). Figure 1 plots this together with its confidence interval.

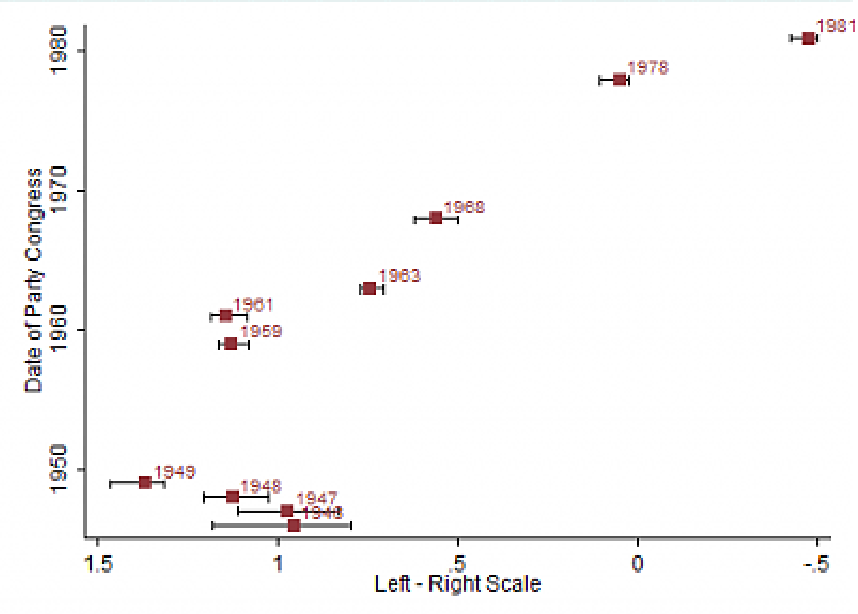

Figure 1. *The evolution of median faction within the PSI from 1946 to 1981 along the left-right scale. ML estimates with their 90% c.i.*

The evolution of PSI towards (more) moderate positions emerges clearly from the picture. In the '40s the party retained a maximalist platform and its median faction was located on the left of the scale. In 1947 two social-democratic factions left the party to create a reformist party, the PSLI, that later became the Italian Socialist Democratic Party (PSDI). After the fission the median faction within the PSI shifted towards the extreme left of the continuum. This is well established looking at 1949, when the party's left-wing won the congress gaining 51% of delegates votes. After that, under the leadership of Nenni in the '60s, the party started to moderate its position adopting reformist policy views. In the 1963 when the PSI was going to join the government coalition, the median faction made a major move adopting more moderate positions. This path continued in 1968 when the PSI merged with the PSDI. After 1978, when Craxi became the new party leader, the PSI median faction was located in the centre of the left-right scale. This shift from left to right is coherent with the findings published by Curini and Martelli (2009) who highlighted that this movement lead the PSI to gain a more powerful strategic position.

This practical example therefore confirms that Wordfish is able to track changes in policy positions across time. As we discussed above these changes are not an artifact of the model but are due to a real modification in the political lexicon used by political actors.

The next example is intended to estimate party position at a single point in time and it has been sketched from current Italian political context. In December 2010, Berlusconi's cabinet faced a vote of no confidence and succeeded in keeping a majority both in the Senate and in the House (although with a narrow advantage, only three votes in the Lower Chamber). In the aftermath of this debate the government tried to increase its margin. The cabinet gathered together some MPs that were prone to support the majority and managed to form two new parliamentary groups loyal to the government. In January 2011 a group called Iniziativa Responsabile (IR) was formed in the House, while two months later, in March 2011, the Senate saw the birth of the group Coesione Nazionale (CN).

Estimating positions of these two new actors could be unclear using common techniques like roll-call analysis. Here we attempt to measure their ideal points through Wordfish. More precisely we want to assess positions of parliamentary groups by means of analyzing the speeches released during the debates about the vote of no confidence faced by Berlusconi's cabinet in December 2010. For each party, we collected all the speeches delivered by its members in both Houses under the assumption that these speeches allow us to estimate the position of the party as a whole. For the two new parliamentary groups we gathered the speeches released in December by the MPs that later decided to join these groups. Finally we include in the analysis the speeches held by the Prime Minister to evaluate the cabinet's position (GOV). Figure 2 shows the position of parties and the confidence interval of the estimates.

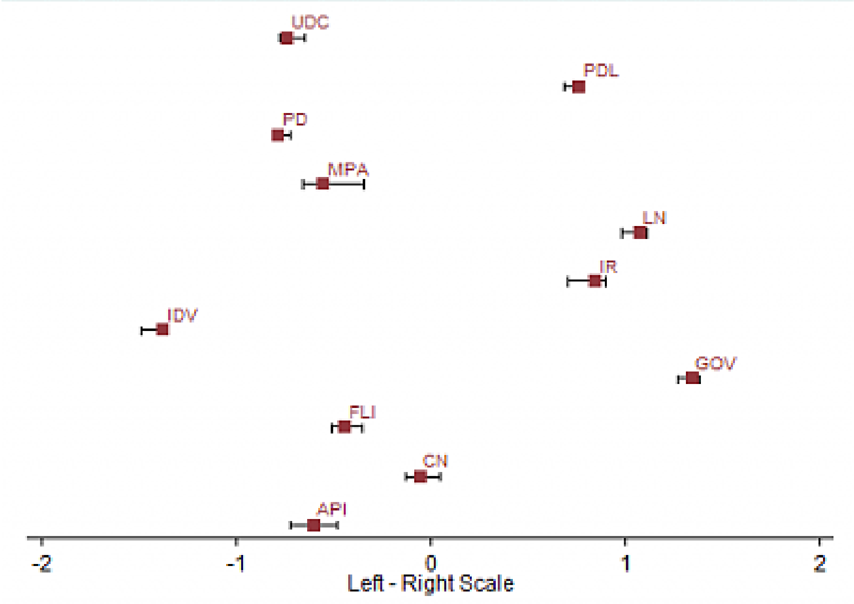

Figure 2. *Policy positions of parties' parliamentary groups along the left-right scale: ML estimates with 95% c.i. of the analysis of parliamentary speeches delivered during the vote of confidence in December 2010 (the y-axis is only used for graphical purposes)*

We observe that parties' positions are consistent with what we should expect having in mind the Italian party system. On the left wing of the scale we find Italia dei Valori (IDV) while the cabinet is on the opposite extreme. Partito Democratico (PD) , Unione dei Democratici Cristiani e di Centro(UDC), Alleanza per l'Italia (API) and Movimento per le Autonomie (MPA) are on the centre-left. We find Futuro e Libertà per l'Italia (FLI) in a more moderate position. One of the two new groups, Coesione Nazionale, is located exactly in the centre of the spectrum while the other, Iniziativa Responsabile, is more on the right side, closer to Popolo della Libertà (PDL) and Lega Nord (LN).

IR and CN are present in different Chambers. The distance between their positions could be due to differences among the two branches of Parliament. Moreover, the composition of the two groups is slightly different: IR is composed mainly by MPs that switched from MPA and UDC well before December 2010, while in CN there are many former FLI members that left the party only in 2011. Notwithstanding, the placement of CN is statistically different from those of opposition parties as we expected.

Apart from allowing us to determine policy positions of parties, Wordfish assesses the political meaning of words included in texts. This feature plays a twofold role. On one side it allows us to run a diagnostic of the results. In addition it helps us to detect the issues stressed by parties and to measure the linguistic differences among them. For instance we can observe how parties talk about one policy issue, what kind of words they employ to deal with it, what kind of topics are discussed mainly by one party or another.

Figure 3 plots word's β coefficient and their fixed effect (gray dots). The fixed effect assesses how often one word appears in all the documents while the β coefficient expresses word's political placement along the left-right scale. We plot some words as an example. Words that are located on the left are highlighted in red, right-wing words are written in blue while words that retain no political meaning are in black.

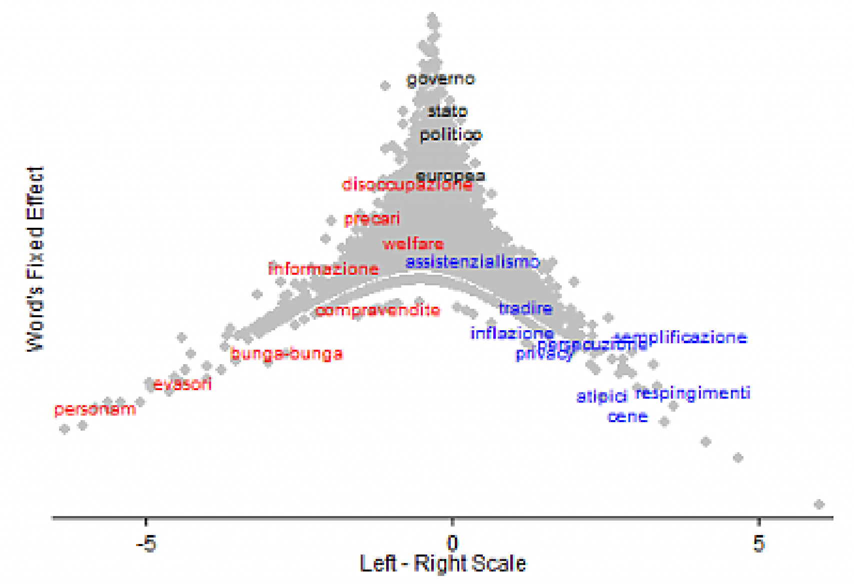

Figure 3. *Estimates of β coefficient and fixed effect of words. In red words that appear on the left side of the policy spectrum; in blue words located on the right*

Words like governo or stato (in black) are used by all parties, hence they are not useful to discriminate between them. For this reason they are placed in the centre of policy scale. Although some words refer to issues at stake in the debate, they might have no discriminating power. This happens because all parties talk about that topic using that word. This could be the case of the word europea or more in general of the word taxes. The beta coefficient of these words is close to zero, hence they are not directly informative. Nonetheless, these matters are not left aside from the analysis. For instance, when talking about taxes, parties will tell us that they are going to cut taxes or to increase them, and they may want to explain how and why. In this case what helps us to discriminate the left from the right is not the word taxes per se but the word "cut" or "increase" (Slapin and Proksch 2008). Of course parties hardly ever say they want to raise taxes. It is more likely that they all agree on cutting taxes "but…". In particular, they might want to cut taxes while preserving the welfare state or they may cut taxes precisely to end its abuse (assistenzialismo). In this context the words welfare and assistenzialismo will discriminate parties' position on the issue related to the dimension taxes/expenditure. We highlighted these two words in the graph and unsurprisingly they are located on the proper side of the scale: welfare on the left, assistenzialismo on the right. The same is true for all other words. This example shows that Wordfish has been able to catch the appropriate political placement of words on a classic left-right dimension. For example right-wing parties are usually concerned with inflation (inflazione) while left-wing parties pay more attention to unemployment (disoccupazione). Indeed inflazione retains a rightist meaning (and helps us to detect rightist parties) while disoccupazione is recognized as a left word. The same happens when looking to other issues. Deregulation is traditionally a right-wing issue and this is caught from the word semplificazione (the intention to simplify legislation). In addition the right cares about law and order and, as an example, supports the rejection of immigrants (respingimenti). On the contrary, the left stresses the topic of equality pointing the finger at tax evaders (evasori). It is worth noting that sometimes parties call the same object with different labels. For instance both left and right parties refer to "temporary workers" using different names: the left calls them precari while the right prefers the formal term atipici.

So far we dealt with general concepts behind the definition of left and right. Now we can go in depth to see that Wordfish is able to understand the underlying topics that shape current political debate. Many speeches held during the parliamentary debate revolved around justice and the sex scandals that involved the Prime Minister. This is clear when looking at figure 3. The object of the scandals is called bunga-bunga, and the left adopts this term, while right-wing parties refers to refers to it with the word cene (dinner party). The left employs words like [ad] personam to criticize government's action on justice, while right-wing parties justify this legislation with the need to protect the Prime Minister from judges' persecution (persecuzione). For similar reasons the right calls for more privacy when dealing with the life of Prime Minister while the left uses more often the word informazione and wants to protect the freedom of press. Finally, Wordfish also catches an element related to a valence issue. During the vote of confidence in December 2010, opposition parties accused the government of paying MPs to switch sides, supporting the cabinet (later they created the parliamentary groups IR and CN). They spoke about compravendita. The opposing parties in government were upset because in the summer of 2010 a splinter group had left the PDL to form a new party (FLI) and this party changed sides in December voting against the cabinet. They were accused of betraying (tradire) the voters.

Conclusion

In this paper we presented a new method, Wordfish, that allows us to extract positions of political actors by analyzing words included in texts. This technique of automated quantitative text analysis is a fast and reliable way to produce an estimate of policy platforms that can be employed to estimate both time series data as well as positions at a single point in time. For all these reasons, is reasonable to expect that Wordfish, as well as similar programs, will play an increasing role within the political science community, including Italy, as valuable tools to better understand political dynamics.

References

- Benoit, K. and Laver, M. (2006). Policy in Modern Democracies. London: Routledge.
- Carrubba, C., Gabel, M., Murrah, L., Clough, R., Montgomery, E. and Schambach, R. (2006). "Off the Record: Unrecorded Legislative Votes Selection Bias and Roll-Call Vote Analysis", British Journal of Political Science 36(4), pp. 691-704.
- Ceron, A. (forthcoming). "Correnti e frazionismo nei partiti politici italiani (1946-2010): una analisi quantitativa delle mozioni congressuali", Rivista Italiana di Scienza Politica.
- Curini, L. (2010). "Experts' Political Preferences and Their Impact on Ideological Bias", Party Politics, 16(3), 2010, 299-321.
- Curini, L. (2011). "Government Survival the Italian Way: the Core and the Advantages of Policy Immobilism during the First Republic", European Journal of Political Research, 50, 2011, 110-142.
- Curini, L. and Martelli, P. (2009). I partiti nella prima Repubblica. Maggioranze e governi dalla Costituente a Tangentopoli, Carocci, Roma.
- Klüver, H. (2009), "Measuring Interest Group Influence Using Quantitative Text Analysis", in European Union Politics, 10, pp. 535-549.
- Laver, M. (1999). "Divided Parties, Divided Government", Legislative Studies Quarterly, Vol. 24, No. 1, pp. 5-29.
- Laver, M., Benoit, K. and Garry, J. (2003). "Extracting Policy Positions from Political Texts Using Words as Data", The American Political Science Review, Vol. 97, No. 2 (May, 2003), pp. 311-331.
- Mair, P. (2001) "Searching for the Position of Political Actors: A Review of Approaches and a Critical Evaluation of Expert Surveys" in Michael Laver (ed.), Estimating the Policy Position of Political Actors. London: Routledge, 2001, 10-30.
- Proksch, S. O., and Slapin, J. B. (2009a). WORDFISH: Scaling Software for Estimating Political Positions from Texts. Version 1.3 (22 January 2009), http://www.wordfish.org.
- Proksch, S. O. e J.B. Slapin (2009b), "How to Avoid Pitfalls in Statistical Analysis of Political Texts: The Case of Germany", in German Politics, 18, 3, pp.323-344.
- Proksch, S. O. e J. B. Slapin (2010), "Position Taking in European Parliament Speeches", in British Journal of Political Science, 40, pp. 587-611.
- Slapin, J. B. and Proksch, S.-O. (2008). "A Scaling Model for Estimating Time-Series Party Positions from Texts", American Journal of Political Science, Vol. 52, No. 3, (July 2008), pp. 705-722.